International Monetary Fund

IMF staff papers

Volume 53 Number 1
2006

IMF staff papers

The objective of *IMF Staff Papers* is to publish high-quality research produced by IMF staff and invited guests on a variety of topics of interest to a broad audience including academics and policymakers in the member countries of the Fund. The papers selected for publication in the journal are subject to an extensive review process using both internal and external referees. *IMF Staff Papers* also welcomes outside comments, criticisms, and interesting replications of published work. The views presented in published papers are those of the authors and should not be attributed to or reported as reflecting the position of the IMF, its Executive Board, or any other organization mentioned herein.

Subscription: US$99.00 a volume or the approximate equivalent in the currencies of most countries. Three issues constitute a volume. Single copies may be purchased at $30.00. Individual academic rate to full-time professors and students of universities and colleges: $58.00 a volume. Subscriptions and orders should be sent to:

International Monetary Fund
Publication Services
700 19th Street, N.W.
Washington, D.C. 20431, U.S.A.

Telephone: (202) 623-7430
Fax: (202) 623-7201
E-mail: publications@imf.org
Internet: *http://www.imf.org*

International Monetary Fund

Volume 53 Number 1

IMF staff papers

2006

EDITOR'S NOTE

The Editor invites from contributors outside the IMF brief comments (not more than 1,000 words) on published articles in *IMF Staff Papers*. These comments should be addressed to the Editor, who will forward them to the author of the original article for reply. Both the comments and the reply will be considered for publication.

The data underlying articles published in *IMF Staff Papers* (where available) may be obtained from the journal's website (*http://www.imf.org/staffpapers*). Readers are invited to use these data to expand on the material in the articles, and the journal will consider publishing such work.

ISBN 1-58906-509-3
International Standard Serial Number: ISSN 1020-7635

This serial publication is catalogued as follows:

International Monetary Fund
IMF staff papers — International Monetary Fund. v. 1– Feb. 1950–
[Washington] International Monetary Fund.

v. tables, diagrs. 26 cm.

Three no. a year, 1950–1977; four no. a year, 1978–

Indexes:
Vols. 1–27, 1950–80, 1 v.
ISSN 1020-7635 = IMF staff papers — International Monetary Fund.
1. Foreign exchange—Periodicals. 2. Commerce—Periodicals.
3. Currency question—Periodicals.

HG3810.15 332.082 53-35483

International Monetary Fund

Volume 53 Number 1

Contents

2006

This page intentionally left blank

IMF Staff Papers
Vol. 53, No. 1

Russia and the WTO: The "Gravity" of Outsider Status

BOGDAN LISSOVOLIK AND YAROSLAV LISSOVOLIK*

With China's accession to the WTO in 2001, Russia is by far that organization's most prominent nonmember. This paper applies the gravity model to gauge whether this "outsider" status has been affecting Russia's export structure. On the basis of cross-section and panel regressions for 1995–2002, we find that Russian exports to WTO members have fallen short of the model's predictions. The paper discusses possible explanations of this result, including Russia's exclusion from various WTO procedures, although own-export restrictions could have a similar effect. The model points to Russia's further trade reorientation toward WTO members after a putative accession. Our results also prompt some ideas that may clarify the recent empirical controversy over the WTO's overall role in promoting trade.
[JEL F14, F18]

The role of the World Trade Organization (WTO) in international trade has recently come under increased scrutiny. In a provocative study, Rose (2002) concluded—on the basis of a gravity model—that the effect of the WTO on trade was insignificant. This result was disputed by several authors.[1] In particular, Subramanian and Wei (2003) employed an augmented specification of the gravity model to estimate the relationship between trade and WTO membership. These

*Bogdan Lissovolik is a Senior Economist in the European Department of the IMF. Yaroslav Lissovolik is a Chief Economist with United Financial Group in Moscow and was an Advisor to Russia's Executive Director at the IMF when this paper was written. The authors would like to thank Hans Peter Lankes, Nienke Oomes, David Owen, Andrew Rose, Emil Stavrev, Patrizia Tumbarello, Bert Van Selm, an anonymous referee, and participants in the seminar at the IMF's European Department for helpful comments and suggestions. Tom Walter provided valuable editorial input, and Arlene Tayas excellent logistical support.

[1]See Rose (2005).

authors argued that the organization strongly promotes trade, although with considerable asymmetries across sectors and groups of countries. In particular, WTO liberalization appears to be more useful to members than to nonmembers, although the latter also could benefit somewhat from the spillover effect of expanding global trade. Rose has countered, however, that these asymmetries are not central to the WTO's overall role, and thus the debate continues.

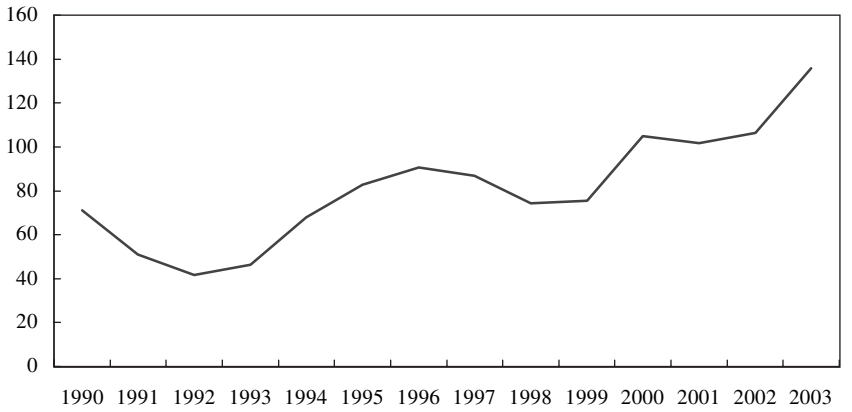
This paper uses a gravity model to evaluate the impact of the WTO on the trade of Russia—currently the largest nonmember. There are several reasons for the single-country focus. First, it partly gauges the insights of the "multicountry" gravity models, since these should, to a certain extent, apply to large countries with reasonably diversified trade flows. Second, it assesses the WTO-related issues from a national perspective, which is, in practice, a key focus of most policy decisions. Third, Russia's case is particularly important because of its size and the fact that its accession negotiations provide a reference point for some other nonmember countries (notably in the Commonwealth of Independent States (CIS)) involved in a similar process. Finally, Russia's WTO accession has been a very contentious topic, both in terms of its domestic political economy and with respect to entry conditions demanded by some WTO members.

Our main conclusion is that the impact of Russia's accession to the WTO on the structure and possibly the level of trade may be quite significant and could be much higher than current estimates. As a first step, we find cross-section and panel data evidence that, on average (after adjusting for the customary gravity model and country-specific effects), Russia's exports to the WTO countries underperformed its exports to other countries in 1995–2002. These results are somewhat surprising in light of both the general trade-promoting view of the WTO and the overwhelming casual evidence of Russia's trade reorientation toward the more advanced developed and developing countries, all of which are or have recently become WTO members.

As a second step, we explore the possible reasons for this empirical result, which include (1) Russia's nonparticipation in WTO procedures, (2) Russia's own export restrictions, (3) possible model specification and data caveats, (4) structural path dependence in Russia's exports, and (5) Russia's comparative advantage in bargaining with WTO nonmembers. The first two factors come out as more plausible and consistent with the stylized facts: It appears that, with WTO accession (which will relieve trade restrictions), Russia would further reorient its trade in line with the gravity model. If so, the magnitude of the trade-related benefits from joining the WTO appears quantitatively large in the long term, although the precise parameter estimates must be interpreted with caution, given the partial nature of the model and certain data problems.

Finally, we discuss how our results might help shed some light on the current empirical controversy over the WTO's overall role in influencing trade. We conclude that, while the WTO exhibits some trade-promoting features, the asymmetry between members and nonmembers also may constrain trade and may obscure econometric evidence of the WTO's role in multilateral models. The reason is that a large country like Russia may be a local "center of gravity" for some non-WTO members, and the customary controls of the gravity model do not eliminate this effect fully in multilateral WTO studies.

Figure 1. Russia's Exports in 1990–2003
(*US$ billion*)



Sources: Central Bank of Russia, and IMF staff estimates.

I. Russia's Evolving Integration into the World Trading System

Trade Developments

After proclaiming independence in late 1991, Russia embarked on a difficult transition from central planning to a market economy. The key dimension of this transition has concerned the opening up and integration of Russia's economy into the world economy, with the primary focus on trade's role in unleashing incentives in line with the fundamental principle of comparative advantage. This process seemed critical, as much of external trade of the socialist bloc had not been based on market principles but reflected priorities imposed by political and ideological factors. Thus, export and import flows had been heavily concentrated in the former socialist economies and involved various forms of explicit or implicit subsidization. The sudden change in the policy course was exacerbated by the severe economic dislocation of the early 1990s, with further significant effects on the extent, direction, and time profile of adjustment in the external trade sector.

Russia experienced large shifts in levels and structure of external trade during the transition period. In terms of levels, the shock of Soviet disintegration caused a trade implosion in 1991–92 (see Figure 1); this was followed by a steady rebound in 1993–2003, except for a brief slippage in 1997–98.

The significant structural changes in Russia's trade occurred mostly in the early 1990s. The main geographical trend was a reorientation away from the former socialist economies toward more advanced market economies (see Table 1). Trade with the former socialist economies contracted very abruptly, as the value of Russia's exports to those countries decreased by more than half in 1991 alone.[2]

[2]See Granville (1995). However, pre-1994 data on Russia's trade are highly imperfect for a variety of reasons, including very weak compilation capacity and distorted valuation.

Table 1. Geographical Structure of Russia's Exports in 1990, 1995, and 2002
(In percent of total exports)

	1990	1995	2002
Commonwealth of Independent States	64.0	18.5	14.6
European Union	. . .	33.6	34.9
Other Former COMECON[1]	15.5	11.0	10.7
United States	. . .	6.6	6.1
China	. . .	4.4	6.3

Sources: IMF, *Direction of Trade Statistics;* Daviddi and Espa (1996) for the 1990 data.
[1]Refers to Bulgaria, Cuba, former Czechoslovakia, former East Germany, Hungary, Poland, Romania, Mongolia, and Vietnam.

Since the mid-1990s, however, changes in the geographical composition have not been very large. In the end, Russia's exports remained geographically diversified, at least compared with those of the CIS countries. Thus, the share of Russia's three main export markets in total exports was 23 percent in 1995, compared with between half and two-thirds for other CIS countries (see Elborgh-Woytek, 2003). In 2001, this figure stood at 22 percent for Russia, compared with a 31–65 percent range for other CIS countries.

The commodity composition of trade also has changed significantly compared with that of socialist times, particularly on the export side, with a reduction in machinery exports accompanying a steady expansion in shipments of energy and, to a lesser extent, semiprocessed goods (metals and chemicals). For example, the share of machinery in Russia's exports to non-CIS countries fell from 18 percent in 1990 to 7 percent in 1993 (Daviddi and Espa, 1996). To some extent, these changes reflected the expected market-based pressure to downsize low or negative value-added activities. While there were further changes in the commodity composition

Table 2. Sectoral Structure of Russia's Exports in 1995 and 2002
(In percent, current U.S. dollar value terms)

	1995	2002
Agricultural products	3.3	2.6
Minerals	42.0	55.2
Chemical industry products	9.9	7.0
Wood, pulp, and paper	5.6	4.6
Textiles	1.5	0.8
Metals and precious stones	26.1	18.6
Machinery and transport equipment	9.9	9.5
Other	1.7	1.7

Source: Goskomstat of Russia.

of exports between 1995 and 2002 (see Table 2), they were much less pronounced than in the early 1990s.

These trade-related developments appear to have had a sizable impact on the key macroeconomic and structural dimensions of Russia's transition. On the macroeconomy, the trade contraction greatly exacerbated the early output decline and "disorganization" (see Blanchard and Kremer, 1997), while the subsequent start of trade recovery preceded the stabilization of output (see Figure 2), and that of the exchange rate, and prices in the mid-1990s. The 1998 currency crisis was in no small measure triggered by adverse shocks to Russia's exports, while the post-1998 recovery relied on the reversal of those shocks in conjunction with the substantial expansionary impact of the real depreciation of the ruble.[3] Regarding structural issues, trade has not only subjected enterprises to competitive pressures and world price signals but also helped keep in check barter and other forms of a noncash economy, which had plagued Russia's transformation for most of the 1990s.

Despite its largely beneficial effect in terms of market-based adjustment, the role of foreign trade in Russia's economy has been constrained by domestic and external policy decisions. Domestically, the decisions reflected discretionary government interference, with more than occasional administrative actions at federal and local levels against the principles of free trade. A case in point is various export restrictions or bans imposed by local governments to protect the supply of essential products during the 1990s. Another example is federal government decisions to tax or restrict exports of energy products more heavily starting in 1999. Whatever the merit of some of these steps, they caused concern over policy reversals in the absence of a comprehensive strategy and framework for such decisions. Externally, the key problem concerned the substantial remaining restrictions on Russia's exports by industrialized and some developing countries, which have mostly affected semiprocessed products such as metals and chemicals. These policies may explain the evidence that the bulk of trade-related structural change occurred in the early 1990s but slowed substantially thereafter.

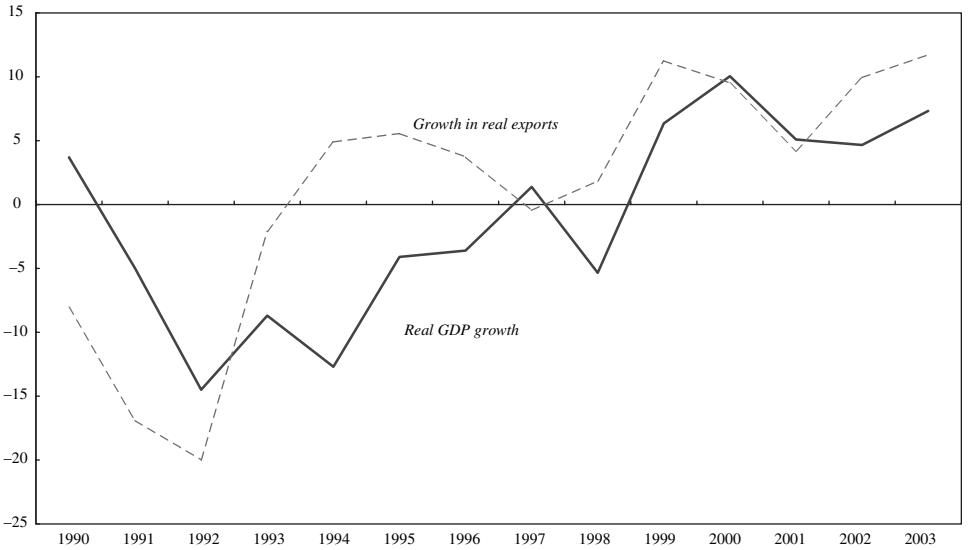
The WTO Entry Debate

Russia's accession to the WTO has emerged as a key step for further market-oriented reform (see Lissovolik and Liventsev, 2002).[4] On the one hand, this accession could harmonize Russia's domestic legislation and practices with those of its major trading partners. On the other hand, it would remove the main remaining obstacles to Russia's exports to WTO members, amplifying the substantial gains from trade that Russia has already been able to generate. The unused potential for further trade reorientation is illustrated by the fact that Russia's exports to the WTO accounted for "only" about 80 percent of its total exports in 2002 (Figure 3),

[3]See Owen and Robinson (2003) for an overview of macroeconomic and structural developments in Russia.

[4]A short chronology of Russia's ongoing WTO accession process is provided in Appendix Table A.1.

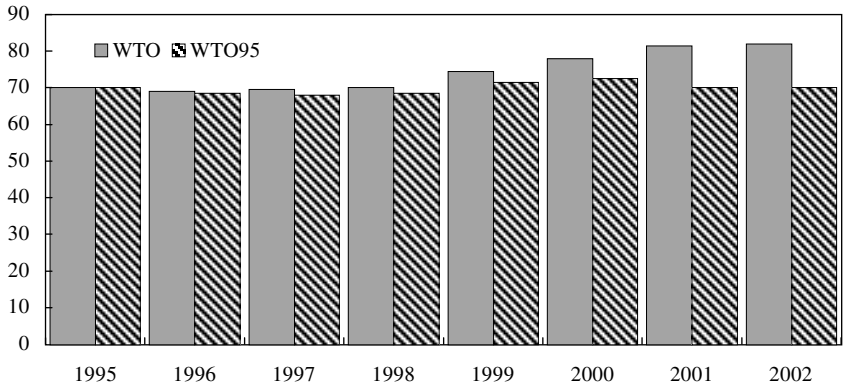
Figure 2. Growth in Real GDP and Real Exports in 1990–2003
(*In percent*)



Sources: Goskomstat, and IMF staff.

compared with the 95 percent share of WTO members in world trade. Furthermore, the share of Russia's exports to WTO countries that became members of the organization in 1995 hardly changed in the period 1995–2002, with the growth in the share of exports directed to all WTO countries largely accounted for by the increase in WTO membership. However, despite these arguments and the procla-

Figure 3. Share of Russia's Exports to WTO Countries in Its Total Exports
(*In percent*)



Source: IMF, *Direction of Trade Statistics*.

Note: the WTO series shows the share of Russia's exports to all WTO members in each year, while the WTO-95 series gives the same indicator only for those countries that were WTO members by the end of 1995.

mation of WTO entry as Russia's key policy priority, the debate on the benefits of Russia's membership has become more ambiguous lately, causing substantial delays in the already protracted accession process.

The uncertainty over Russia's gains from WTO membership is based on several factors. First, the past few years have seen less optimism over the benefits and prospects of free trade, both around the world (following the well-known events in Seattle and Cancun) and within Russia. Second, domestic opponents of Russia's WTO entry have argued that additional gains from membership would be limited, since the country already enjoys most favored nation (MFN) status with many WTO members, and some advanced countries have accorded preferential treatment to Russia under the Generalized System of Preferences (GSP). Third, there have been concerns over asymmetric treatment, as Russia's concessions (in the form of lower tariffs) on imports and a possible short-term output contraction would far outweigh any benefits for its exports. Fourth, some of the conditionality demanded during the accession process—particularly on the liberalization of domestic energy prices—has been widely unpopular because of the perceived economic and social consequences inside Russia. Finally, there has been a determined opposition to WTO entry from some sectoral lobbies, as well as from the ideological opponents of "economic liberalism."

The advocates of WTO accession have not been short on counterarguments. They point to gains from trade liberalization as largely accruing to WTO members, benefiting nonmembers much less. Importantly, some of the effects of WTO membership may not be explicitly embedded in the lower tariff levels but concern intangibles (such as access to the dispute settlement body of the WTO) that are important for establishing a level playing field with WTO members. Thus, Russia's outsider status not only deprives it of automatic MFN treatment but also of the option of defending its interests through the institutional framework of the WTO. Another counterargument is that the possible short-term losses from import competition may be manageable in light of longer-term efficiency gains. The latter would likewise result from the level playing field facilitated by appropriate pro-market structural reforms and by constraints on vested interests.

While WTO accession has proved to be a bone of contention, there is a relative dearth of quantitative evidence to help resolve or anchor this debate. Russia's accession to the WTO has been the subject of a number of studies, but they have had fairly disparate frameworks and the results have varied. The more comprehensive studies have generally focused on the output implications of accession. For example, the Russian Academy of Sciences and the National Investment Council (RAS, 2002) used an input-output model to explore the consequences of accession-related reductions in import duties for sectoral and regional output dynamics. The overall output effect was estimated at 1 percent of GDP. Another study by Tarr, Jensen, and Rutherford (2002) employed a computable general equilibrium (CGE) model, which, in addition to changes in import duties, also estimated the effects of foreign investment and concluded that the latter was likely to account for up to 70 percent of Russia's benefits from WTO accession. On this basis, GDP gains were estimated to range from 3 percent in the medium term to about 25 percent in the longer term.

A major underlying problem with these estimates has been the lack of a reasonably comprehensive projection of the effect of accession on Russia's exports, which may be of primary importance both for the derivation of the output effects and in its own right. Alekseev, Tourdyeva, and Yudaeva (2003) briefly explore this issue in a CGE model, with the assumption that Russia's export competitiveness would be enhanced with lower import duties. Their study yields a broad-based but very modest estimate of a 0.9 percent export expansion. However, the calculation disregards the likelihood that Russia's exports may already be constrained as a consequence of nonmembership. Berglöf and others (2003) put this shortfall at $3 billion, or about 3 percent of Russia's total exports, based on an assessment of actual restrictions on Russian exports across sectors. Still, one may argue that this effect could be very different in the longer term, as the economy adjusts to the improved market access. We know of no studies that explore such a general equilibrium effect within a model of Russia's external trade, and we intend to fill that gap in what follows.[5]

II. Econometric Specification and Data

Model

The gravity model has been one of the notable successes in empirical economics (see Anderson and van Wincoop, 2003). Its applications have by now become standard for the evaluation of various issues in trade, migration, investment, currency unions, and so on. In particular, it has served as both an alternative and a complement to the CGE models (Greenaway and Milner, 2002). It has recently been used to assess the effect of the WTO on multilateral trade patterns, notably by Rose (2002) and Subramanian and Wei (2003), who evaluated the sign and magnitude of the "WTO dummy" after controlling for the customary gravity effects. We will employ a Russia-centered gravity model to evaluate Russia's trade determinants and patterns, with particular attention to the role of the WTO.[6]

The standard gravity model specification relates total trade turnover or exports/imports as a dependent variable to distance and GDP as independent variables. Most such models also include additional independent variables, such as population, GDP per capita, and dummy variables denoting borders, islands, and former colonies, as well as membership in free trade areas (FTAs), currency unions, and so on. Our basic model is given by

$$\begin{aligned} \mathrm{Ln}(X_{it}) &= \ln(D_i) + \ln(Y_{rt} * Y_{it}) + \ln(y_{rt} * y_{it}) \\ &\quad + USSR_i + COMECON_i + EAEC_i + BORDER_i + WTO_{it} + \mu_{it}, \end{aligned} \tag{1}$$

[5]Dean and Eremenko (2003) used gravity model simulations for Ukraine, concluding that there would be no measurable improvement of its market access in the event of WTO accession. However, their model did not explicitly include a WTO-related variable.

[6]The gravity model has already been used for Russia and the former Soviet Union in several studies, notably by Gros and Steinherr (1995) and van Selm (1997). That research pointed to the high predictive power of the model with respect to the intra-Soviet trade flows.

where *X* stands for Russia's exports to country *i* at time *t*; *D* is the distance between Russia and country *i*; the subsequent two terms denote the products of Russia's and the partner's GDPs and GDPs per capita, respectively; *USSR* and *COMECON* are dummy variables denoting the partner country's former status as a Soviet republic or a former member of the Council of Mutual Economic Assistance; *EAEC* stands for the countries forming the Eurasian Economic Community; and *BORDER* denotes the existence of a common border with Russia. The inclusion of these variables allows us to control for as many natural causes of trade as possible and search for effects of the WTO dummy variable (reflecting the partner country's membership in the WTO) on the residual. To check the robustness of the basic results, more independent variables common to the literature will be added to the basic regression; in particular, those reflecting trade regimes and restrictions (i.e., tariffs and the GSP), as well as those pertaining to country and sectoral asymmetries, identified by Subramanian and Wei (2003) as important for assessing the general impact of the WTO on trade.

We define distances between Russia and its trading partners as the "great circle" between Moscow and the respective capitals of the partner countries. Given Russia's size, there may be a problem with this definition; for example, Morocco is deemed closer to Russia than China is, even though Russia shares an extensive border with the latter. Gros and Steinherr (1995) resolved this problem by splitting Russia's economic space into several macroregions, whose distances to Russia's trading partners were estimated separately. In what follows, we confine ourselves to adding a dummy variable for bordering countries, which in part addresses the distortion of distance measurement. We believe that our capital-distance-corrected-for-border approach may be much more reasonable than the alternative (smallest distance between the borders), given that much of Russia's economic potential is heavily concentrated in the European part of the country, of which Moscow is the approximate center.

Regarding global trading arrangements, our analysis focuses narrowly on the role of the WTO proper, leaving the effects of its predecessor—the General Agreement on Tariffs and Trade (GATT)—outside the scope of this study. This is largely because the WTO was created only three years after Russia's independence and about the time when Russia's trade flows were beginning to be guided mostly by market forces. There are also important differences between the WTO and the GATT that may rationalize their separate treatment in empirical studies (although much of the existing empirical literature has ignored them): (1) the WTO is an intergovernmental organization (while the GATT was essentially a set of rules) with an institutional framework and a greater role of multilateral agreements, (2) the WTO dispute settlement system appears to be significantly more efficient than the system under the GATT, and (3) the WTO has a much wider coverage of trade issues.

Our specification of the gravity model differs in several ways from that of most models employed in the related literature, reflecting the particular aims and constraints of our exercise. First, we confine the data set to the post-Soviet period of 1995–2002, thereby skipping the statistical chaos in the Russian export series of the early 1990s. However, this comes at the cost of significantly reducing the number of periods for time-series analysis. Second, the model investigates only

Russia's pairwise trade with other countries, as opposed to the multilateral setup of most gravity models. This country-centered specification of the gravity model is not unusual (see Hufbauer and Oegg, 2003) and allows us to focus on idiosyncratic patterns of Russia's foreign trade through a more precise modeling of the country-specific parameters. However, we should be cautious about generalizing some of our conclusions on the effects of WTO membership, since Russia's specific characteristics may impart a bias or complicate the interpretation of the WTO variable.

Data

While existing data sets on the gravity model (most notably that of Andrew Rose, posted on his website—see Rose, 2005) served as a guide, the bulk of the data have been compiled from various sources. The data set spans a period of eight years from 1995 to 2002 and encompasses 171 countries, most of which are WTO members, including the recent entrants (see Table 3). In order to account for data imperfections in the course of the sensitivity analysis, five countries were excluded for reasons of likely measurement error and incomplete observations.[7] The exact definitions of the variables are presented in Appendix Table A.2. The data on exports are derived from the IMF's *Direction of Trade Statistics,* with the occasional gaps covered by Russia's customs statistics. The sectoral dummy variables (for the metals and oil sectors) were derived with the help of data from the Customs Committee of the Russian Federation. The information on world population was obtained from the United Nations and from the U.S. Census Bureau. GDP figures across countries are taken from the IMF's World Economic Outlook (WEO) database. The data on distances from world capitals to Moscow were derived from computer software that is readily available on the website of the U.S. Department of Agriculture. All regressions were performed using Stata 8.0.

As recognized by Rose (2002), the data on the GSP are imperfect, although the United Nations Conference on Trade and Development (UNCTAD) has been making an effort to compile and update the data on the beneficiaries of GSP regularly. One such database posted on the official UNCTAD website[8] contains a list of countries that granted Russia GSP status as of June 1, 2001. Of the 19 countries that have accorded GSP status to Russia, 15 are members of the European Union; the 4 remaining countries are Canada, the Czech Republic, the Slovak Republic, and the United States.

Data on Russia's exports in 1995–2002 exhibit significant volatility across time, with exports to some countries in Africa and Latin America emerging only toward the second half of the period. Owing to the conversion of exports into natural logarithms, the observations denoting zero exports to trading partners were treated as missing by Stata, thereby significantly reducing the overall number of observations in the panel data set. To check for this problem, we also transform the

[7]Including the estimates for these countries and territories (Bermuda, Cuba, Iraq, Democratic People's Republic of Korea, and Serbia and Montenegro) does not alter the statistical significance of the results or their qualitative nature.

[8]See http://www.unctad.org/Templates/Page.asp?intItemID=1418&lang=1.

Table 3. Annual Increase in WTO Membership, 1995–2002

Members as of January 1, 1995	76
1995	112*
1996	16
1997	4 (Democratic Republic of the Congo, Republic of Congo, Mongolia, Panama)
1998	1 (Kyrgyz Republic)
1999	2 (Estonia, Latvia)
2000	5 (Albania, Croatia, Georgia, Jordan, Oman)
2001	3 (Moldova, China, Lithuania)
2002	1 (Taiwan Province of China)
Total WTO members as of end-2002	144

Source: WTO.
*Includes countries that became members of the WTO as of January 1, 1995.

underlying data in various ways: by adding 1 to the number under the logarithm, by assigning negligible values to the missing observations, or by averaging observations across periods. As reported below, these changes did not have a significant effect on the basic regression results.

III. Econometric Results

To shed light on the factors behind Russia's export performance, we run several sets of gravity model regressions. We also explore various modifications to the underlying model to check the robustness of our results.

Cross-Section Regressions

Table 4 presents the main cross-section results, which are tabulated as sequential independent regressions with each of the year-specific WTO dummies for the 1995–2002 period. The outcomes appear reassuring. The gravity model seems to fit Russia's export data quite well, as can be seen from the high *R*-squared, averaging some 70 percent across the regressions for different years. Distance and output coefficients are all significant, with signs and magnitude similar to those of other applied gravity models. In particular, the coefficient on distance is generally slightly lower than −1, while the sum of coefficients on output and output per capita is close to unity. The former Soviet Union and COMECON dummy variables are also highly positive and significant. The dummies for common border countries and the EAEC also are of the right sign but are generally not significant, at 5–10 percent levels.

Interestingly, the WTO coefficient is always negative, although in most cases (with the exception of 1996) it is statistically insignificant at the 5 percent level. At the same time, the coefficient is close to being significant at the 10 percent level for most years, while truly negligible only for 1998 data. However, 1998 was the year of severe macroeconomic crisis in Russia and thus could possibly reflect

Table 4. Cross-Country Ordinary Least Squares (OLS) Regressions, 1995–2002
(Regressand: Russia's exports; standard errors below)

									Annual Means	
	1995	1996	1997	1998	1999	2000	2001	2002	WTO-95	WTO-96
WTO	–0.55	–1.19	–0.45	0.00	–0.55	–0.55	–0.54	–0.54	–0.76	–0.98
	0.38	0.36	0.36	0.28	0.34	0.31	0.32	0.43	0.32	0.34
Log distance	–0.98	–1.22	–0.96	–1.01	–0.96	–1.35	–1.35	–1.63	–1.38	–1.34
	0.20	0.20	0.21	0.18	0.24	0.21	0.22	0.29	0.24	0.24
Log product	0.72	0.75	0.65	0.64	0.84	0.88	0.91	1.02	1.00	0.99
real GDP	0.10	0.08	0.08	0.07	0.08	0.07	0.07	0.09	0.08	0.07
Log product										
real per	0.11	0.38	0.26	0.18	0.06	0.08	0.07	–0.01	0.14	0.15
capita GDP	0.11	0.10	0.10	0.08	0.10	0.09	0.09	0.12	0.10	0.10
Common	0.11	0.23	0.40	0.39	0.59	0.44	0.56	0.30	0.23	0.18
border	0.45	0.45	0.47	0.40	0.58	0.51	0.53	0.69	0.60	0.60
USSR	2.01	1.82	2.17	1.91	1.93	1.51	1.77	1.80	2.13	2.00
	0.62	0.59	0.60	0.49	0.70	0.62	0.63	0.82	0.72	0.72
COMECON[1]	2.14	2.66	2.62	2.00	2.36	2.22	2.18	1.98	2.22	2.30
	0.50	0.49	0.52	0.43	0.62	0.54	0.56	0.73	0.61	0.60
EAEC	0.89	0.95	0.73	0.89	0.96	1.20	1.06	0.94	1.10	1.09
	0.76	0.76	0.80	0.69	1.00	0.88	0.91	1.20	1.04	1.03
Constant	5.93	7.06	5.65	6.42	4.70	7.74	7.04	8.51	6.18	6.10
	1.89	1.89	1.94	1.61	2.21	1.93	2.01	2.67	2.29	2.27
R-squared	0.62	0.75	0.70	0.76	0.69	0.76	0.76	0.69	0.76	0.77
Number of										
observations	92	103	117	123	153	153	146	155	168	168
F-test	16.97	35.23	31.39	45	40.19	57.6	53.5	40.97	63.86	65.12

Source: Authors' calculations.
[1]For COMECON see footnote to Table 1.

some specific factors. The persistently negative sign on the WTO dummy basically means that, after controlling for the gravity factors, Russia systematically tended to export more to non-WTO countries than to WTO countries.

To even out year-specific idiosyncrasies, we run a cross-section regression on the anual means of the model's variables (last two columns of Table 4). These "core" regressions include the dummy variables reflecting WTO membership at end-1995 and at end-1996, the last year when changes in membership took place on a significant scale. The basic "gravity" coefficients on distance and total GDP continue to be highly significant and appropriate in sign and magnitude, as are the dummies on the former Soviet Union and COMECON. The WTO dummy becomes statistically significant and numerically larger, whereas those that denote the customs union and the Russia-bordering countries do not, possibly because their effect is already captured by the strongly significant variable reflecting the role of the former Soviet Union.

The results of the regressions on annual means hold for exports as the dependent variable in the regression but also for the overall trade turnover. The WTO

coefficient remains statistically significant and negative in all those regressions. The absolute value of the distance coefficient is smaller than with exports as the dependent variable, whereas the opposite is true for the GDP coefficient. Additionally, the results were robust to different gravity model specifications (Linnemann, 1966; Bergstrand, 1985; and Wang and Winters, 1992), some of which included single-country population and GDP variables separately (instead of per capita GDP or a product of country-pair GDPs). In all such specifications, the statistical significance of the negative WTO coefficient remained largely intact.

Pooled Regressions

Despite their widespread use in the academic literature, cross-section gravity regressions have a number of limitations compared with full-fledged panel data regressions. First, they entail a loss in the number of observations that could be used in the regression, thereby possibly affecting the robustness of the results. Second, they disregard time variation in the data and thus may result in inconsistent estimates (Matyas, 1997). Third, they do not fully answer some policy questions of interest; for example, regarding (1) any effects of the model on the direction of change in trade flows or (2) a trade effect of a Russian partner country's joining the WTO.

As a further step, ordinary least squares (OLS) pooled regressions were performed on all of the observations of the data set for 1995–2002. Compared with the cross-section regressions, the WTO dummy becomes time varying, so instead of the period-specific WTO dummies, we construct a single WTO dummy variable. The results (see first column of Table 5) do not appear to differ much from those of cross-section regressions, with the WTO coefficient negative and statistically significant. Also, the customs union and border dummies become statistically significant at the 5 percent level, which is to be expected with the large increase in the number of observations.

This pooled regression was tested through several robustness checks, none of which mattered for the basic results, including the WTO variable. To account for the possible correlation of country observations over time, we also used robust standard errors (column 2 of Table 5). We also experimented with the inclusion or exclusion of alternative country-specific observations (columns 3–6 and 8), as well as with the quadratic gravity term (column 7).

In sum, the pooled regressions confirm the results of cross-section data but do not appear to add much new qualitative insight. This may indicate that the main driving force for the results is heterogeneity across countries. However, while pooled regressions appear to reinforce cross-sectional results (including on the role of the WTO), this method imposes identical coefficients across countries and thus may induce misspecification. A more sophisticated framework would be helpful to check the robustness of the results, including by disentangling the time-invariant and country-specific effects.

Time, Fixed, and Other Effects

Table 6 presents panel data regressions with time effects. The results remain very similar to those for pooled regressions and withstand a number of robustness

Table 5. Pooled OLS Regressions, 1995–2002
(Regressand: Russia's exports; standard errors below)

	Simple	Robust Standard Errors	All Countries*	Over $0.5 Million	Excluding Poor Data Countries**	Weighted by Real GDP	Quadratic Gravity Term	Excluding Outliers of over 2 Standard Deviations (sd)
WTO	–0.52	–0.52	–0.32	–0.41	–0.40	–0.44	–0.52	–0.41
	0.12	0.13	0.09	0.10	0.13	0.12	0.12	0.10
Log distance	–1.22	–1.22	–1.15	–1.15	–1.29	–1.30	–0.21	–1.16
	0.08	0.08	0.06	0.07	0.08	0.07	1.67	0.07
Log product real GDP	0.82	0.82	0.72	0.70	0.85	0.86	0.38	0.72
	0.03	0.03	0.02	0.02	0.03	0.03	0.14	0.02
Log product real per capita GDP	0.11	0.11	0.15	0.14	0.06	0.06	0.07	0.13
	0.04	0.04	0.03	0.03	0.04	0.04	0.06	0.03
Common border	0.37	0.37	0.53	0.44	0.27	0.31	0.44	0.43
	0.19	0.14	0.16	0.15	0.20	0.18	0.20	0.16
USSR	1.86	1.86	1.96	1.60	1.89	1.83	2.36	1.66
	0.23	0.15	0.19	0.19	0.23	0.23	0.20	0.19
COMECON	2.27	2.27	2.35	2.07	2.20	2.08	1.91	2.11
	0.20	0.13	0.17	0.16	0.21	0.19	0.25	0.17
EAEC	0.97	0.97	0.96	0.94	0.98	0.95	0.89	0.95
	0.32	0.12	0.28	0.26	0.32	0.34	0.32	0.27
Constant	6.78	6.78	6.74	6.78	7.13	7.11	4.51	7.19
	0.73	0.66	0.60	0.73	0.76	0.67	7.07	0.62
R-squared	0.72	0.72	0.78	0.74	0.73	0.73	0.72	0.75
Number of observations	1,042	1,042	1,366	992	1,006	1,042	1,042	1,007
F-test	330.45	268.01	603.99	351.39	340.18	348.90	244.02	370.99

Source: Authors' calculations.
*Logarithmic transformation on the dependent variable: log(x + 1).
**Excluding Democratic People's Republic of Korea, Serbia and Montenegro, Iraq, Bermuda, and Cuba.

Table 6. Regressions with Time Effects, 1995–2002
(Regressand: Russia's exports; standard errors below. Coefficients for distance, GDP, and GDP per capita not reported.)

	Simple	Simple	All Countries*	Over $0.5 Million	Excluding Poor Data Countries**	Quadratic Gravity Term	Excluding Outliers of over 2 sd
WTO	–0.54	–0.51	–0.44	–0.47	–0.41	–0.52	–0.47
	0.12	0.12	0.09	0.10	0.13	0.12	0.11
USSR	1.92	1.91	2.02	1.67	1.95	1.98	1.73
	0.23	0.23	0.18	0.19	0.23	0.25	0.19
COMECON	2.30	2.10	2.40	2.13	2.23	2.39	2.16
	0.20	0.21	0.16	0.16	0.21	0.20	0.17
EAEC	0.97	0.98	0.95	0.95	0.98	0.89	0.95
	0.32	0.32	0.26	0.26	0.32	0.32	0.27
GSP		0.64					
		0.23					
Developed economy		–0.84					
		0.23					
Constant	6.55	6.24	6.23	7.06	6.92	2.83	6.90
	0.73	0.76	0.58	0.61	0.76	7.05	0.62
R-squared	0.72	0.72	0.78	0.74	0.73	0.72	0.75
Number of observations	1,042	1,042	1,366	992	1,006	1,042	1,007
F-test	310.83	252.85	665.4	343.65	319.25	229.29	360.13

Source: Authors' calculations.
*Logarithmic transformation on the dependent variable: log(x+1).
**Excluding Democratic People's Republic of Korea, Serbia and Montenegro, Iraq, Bermuda, and Cuba.

checks, with the WTO coefficient continuing to be significantly negative throughout. In addition to the robustness checks shown in Table 5, we test (in the second column of Table 6) the significance of GSP preferences and of the asymmetry between Russia's exports to the developed and developing economies (through a developed-country dummy). Whereas the WTO coefficient is not affected by these additions, the latter modifications enter with statistically significant coefficients. In particular, the sign on the developed-country dummy is negative, indicating that for some reason Russia's exports to developed countries were "limited" compared with those to other countries, all other things being equal. The sign on the GSP dummy is, as expected, positive.

Table 7 presents augmented robustness checks to the time effects regressions, which do not affect the negative sign of the WTO variable. These include a number of additional controls common in the gravity model literature, such as islands, landlocked areas, the size of the country, and import duties. Additionally, country and sectoral asymmetries are tested, including by dummies for nations that import oil

Table 7. Robustness Checks, 1995–2002
(Time effects, unless otherwise noted. Regressand: Russia's exports; standard errors below.)

	1	2	3	4	5	6	7*
WTO	–0.57	–0.41	–0.41	–0.43	–0.40	–0.35	0.20
	0.12	0.14	0.14	0.13	0.13	0.16	0.19
USSR	1.97	1.98	2.00	1.17	1.31	0.94	
	0.22	0.22	0.22	0.26	0.24	0.34	
COMECON	2.39	2.37	2.25	1.96	2.21	1.70	
	0.20	0.20	0.21	0.21	0.21	0.28	
EAEC	1.29	1.30	1.26	1.11	0.96	1.27	
	0.32	0.32	0.32	0.31	0.32	0.38	
Landlocked	–0.28	–0.26	–0.25	–0.29	–0.39	–0.46	
	0.14	0.14	0.14	0.14	0.14	0.16	
Island	–0.10	–0.10	0.01	–0.16	0.05	–0.36	
	0.17	0.17	0.18	0.18	0.17	0.22	
Log area	–0.25	–0.26	–0.24	–0.20	–0.20	–0.21	
	0.04	0.04	0.04	0.04	0.04	0.05	
GSP			0.55	0.16	0.84	0.12	
			0.24	0.24	0.23	0.29	
Developed economy			–0.43	–0.50	–0.86	–0.64	
			0.24	0.24	0.25	0.29	
Oil				1.03		1.18	
				0.16		0.21	
Steel					1.06		
					0.16		
Import duties						–0.07	
						0.11	
GATT-94		–0.32	–0.33	–0.21	–0.32	–0.14	
		0.13	0.13	0.13	0.13	0.15	
Constant	8.13	7.89	7.42	7.09	7.45	7.76	31.70
	0.77	0.77	0.80	0.79	0.78	1.03	6.41
R-squared: overall	0.73	0.73	0.74	0.75	0.75	0.75	0.92
Number of observations	1,042	1,042	1,042	1,042	1,042	795	1,042
F-test	242.39	223.83	192.92	189.79	190.56	145.00	58.19

Source: Authors' calculations.
Notes: Coefficients for distance, GDP, and GDP per capita not reported.
*Fixed-effects regression with country dummies included.

and steel from Russia. Accounting for these specificities may be important, given the somewhat skewed structure of Russia's exports of those products and because of pronounced trade idiosyncrasies in those sectors (i.e., large trade restrictions for steel). Incidentally, both the oil and steel dummies have large, positive, and significant coefficients. Finally, a variable, GATT-94, has been included to measure the impact for those countries that were the founders of the WTO, since they could be seen as a core group of economies with a long-term commitment to multilateral trade liberalization. Interestingly, the coefficient on this term is significantly negative, and its inclusion does not have much impact on the coefficient of the broader

WTO variable.[9] In addition, we have run several regressions by adding a lagged dependent variable, which have also yielded a negative WTO coefficient, although more weakly significant (at 5–20 percent, depending on the exact specification).

We also attempted to infer from time variation in the data, although this exercise was somewhat limited by modeling and data problems. The last column of Table 7 contains estimates of a country fixed-effects regression[10] with country-specific dummies (with standard adjustments to avoid the overidentification problem). In such regressions, all time-invariant variables (distance and most regional and country dummies) are effectively subsumed into the constant term, thereby exploiting only the variation among four variables: total and per capita GDP, WTO, and exports. The regression essentially addresses a "within" question: What does joining the WTO do to a country's imports from Russia? The positive (albeit not significant at the 10 percent level) coefficient on the WTO dummy weakly suggests, in line with intuition, that the countries that joined the WTO over the sample period actually favored Russian exports at the margin. However, this result has to be taken with caution, because (1) the time period of analysis is short and the number of observations fairly small; and (2) important "control" variables (regional dummies) and the distance variable drop out from the analysis, which may modify the relationship within this "aborted" gravity model.

IV. Discussion of Results

The main result is that within a well-fitting gravity model for Russia, the WTO coefficient is negative and statistically significant in the baseline cross-section and most panel data regressions. On this basis, the structure of Russia's trade remains in some sense "suboptimal" or "different" compared with the benchmark offered by the gravity model, in that it trades "too little" with the WTO members and/or "too much" with non-WTO members. This may not seem surprising, given that a good portion (but by no means all) of non-WTO members are former socialist economies, with significant historical and systemic ties with Russia. However, these results *are* somewhat surprising, since we control for these specificities through various regional dummy variables, which are themselves intuitive, with large coefficients, and highly significant.

This basic result applies to the 1995–2002 period on average but does not yet indicate how this bias evolved over time. First, judging from Table 4, the WTO coefficient in independent cross-section regressions was negative, albeit volatile, in 1996–98 and stabilizing within a remarkably narrow range around –0.55 in 1999–2002. This indicates that the extent of anti-WTO trade bias appeared to be

[9]A "between-effects" panel estimation yielded an even larger absolute estimate of the WTO coefficient, in the range of minus 0.7 to minus 0.9, with the coefficient being robustly statistically significant. Otherwise, results were similar to those of the time effects regression.

[10]A random-effects regression seems less appropriate, given that its underlying assumption—that unobserved individual heterogeneity is uncorrelated with the included variables—seems somewhat heroic in the case of our model (for example, heterogeneity with respect to the GDP variable). In any case, the results from the random-effects regression showed a negative coefficient on the WTO variable, albeit small and insignificant.

roughly constant over the sample period. Second, the fixed-effects regression with country-specific dummies actually suggests that Russia's trade with the WTO was encouraged at the margin, since some 30 countries that joined the WTO after 1995 tended to favor Russian exports, all other things being equal (Table 3). This may be a consequence of the general trade-liberalizing measures that these countries undertook in the process of WTO accession and thus may be consistent with the overall WTO positive spillover effect found by Subramanian and Wei (2003). Still, our regressions show only a limited "within" effect during 1995–2002 and thus do not reject the possibility of a "trap" or some persistence of an anti-WTO equilibrium.

In sum, the highlighted bias concerns the trade structure in terms of levels, while the evidence on the direction of change is largely inconclusive. Numerically, the value of the WTO coefficient in most regressions suggests that, in the long run, Russia exports on average at least one-third less [exp(–0.4) – 1] to members of the WTO than to all of its trading partners. These are very large estimates, implying potentially huge trade costs of nonmembership in the WTO, subject to a number of qualifications, which will be investigated below.

Factors Behind the Regression Results

Why does the structure of Russia's trade differ from that suggested by the gravity model? Clearly, various factors could account for Russia's particular trading pattern, ranging from statistical and historical reasons to more substantive economic and policy issues. We briefly review key arguments below; more details can be found in Lissovolik and Lissovolik (2004).

First, Russia's exports may be constrained by *restrictions imposed by its WTO trading partners,* either because these do not enjoy the full benefits from the trade liberalization rounds, or owing to barriers—formal or informal—levied by individual WTO members. WTO member countries may impose these barriers on nonmembers more readily than on members, because the former cannot retaliate, because they do not belong to the WTO's dispute settlement bodies, or for other reasons.

To gauge the influence of trade barriers on our regression results, we have included a separate GSP term in the regressions (Table 7); it carries the expected positive sign, but has only a marginal impact on the significance of the WTO sign. In addition, we have controlled for the level of explicit import duties (Table 7); the sign is, as expected, negative, but the economic and statistical significance of the coefficient is marginal. These weak results are not surprising, given that dummies are only rough proxies for policies, but this may also suggest a role for "implicit" and "microlevel" restrictions, such as antidumping actions, or the probability of recourse to them.

Regarding the sensitivity of our results to these de facto restrictions (including implicit trade barriers), we introduced sectoral dummies within the regressions (Table 7), focusing on two benchmark export commodities that are starkly different from the point of view of trade restrictions: steel and oil. As expected, both dummies appear highly significant and have positive coefficients. A steel importer dummy, which controls for a sector in which trade restrictions (explicit and implicit)

are prevalent, reduces the size of the coefficient on the WTO variable; while an oil importer dummy—with no perceptible trade restrictions on the part of importers—increases the absolute size of the negative coefficient.[11] These results suggest a positive link between the anti-WTO bias and restrictions on Russia's exports. However, the evidence is only suggestive, since the changes in the WTO coefficient are fairly small.

Second, the export pattern may have been influenced by Russia's *domestic export restrictions.* While most of Russia's export duties were gradually dropped by mid-1996, some duties were introduced or reintroduced in 1999–2000. There is at least some evidence pointing to the role of these domestic export restrictions in generating the gravity model's negative WTO coefficient. In the cross-section results of Table 4, the significance of the WTO variable becomes much lower in 1997 and essentially breaks down in 1998, when domestic export restrictions were minimal.[12]

Third, the negative sign on the WTO variable may be connected to Russia's excessively good *bargaining position* vis-à-vis some non-WTO members, which may be labeled as "relatively small and dependent economies." However, the coefficient on the real GDP in the regressions is positive and—at just below unity—consistent with those from other gravity model applications. This, all other things being equal, suggests that there is nothing special in Russia's trade with smaller countries.

Fourth, despite its good fit, the gravity model is susceptible to *specification problems.* These may include (1) insufficient control for multilateral resistance (Anderson and van Wincoop, 2003); (2) endogeneity bias (countries with substantial trade with Russia are less interested in WTO entry); (3) simultaneity bias (political preferences or other factors pushing some countries toward Russia and away from the WTO at the same time); and (4) omitted variables that tilt Russia's trade toward particular countries (including insufficient quality of Russia's manufacturing products). We explored these issues to some extent in Lissovolik and Lissovolik (2004) and found no clear evidence of these problems, either on economic or statistical grounds. For example, the GATT-94 coefficient that controls for the established WTO members (and is not affected by endogeneity) is significantly negative in all specifications of the gravity model.

We also have checked the sensitivity of our results to country groupings more generally. Thus, industrial countries—all of which were WTO members during the sample period—appear to be particularly nonreceptive to Russia's exports. As we can see from Table 7, a dummy for those countries is significant in most of our benchmark regressions, and always with a negative sign. If we split the WTO dummy into developed- and developing-country dummies, the developed-country dummy has a larger absolute value of almost −1.1, compared with the developing country WTO dummy of –0.55, with both being significant. Regarding the sensi-

[11]Oil, with the role of pipeline infrastructure in generating export inertia, may not be a perfect example, but a fair portion of Russia's oil exports (by sea) can be redirected at the margin.

[12]At the same time, there could be an alternative explanation based on our first hypothesis (WTO countries' restrictions on Russia's exports). Russia's declining export performance in 1997–98 may have been caused by the concurrent global crisis, with some preset import restrictions becoming less binding because of the contraction in world trade.

tivity of the regression results to individual country observations, we compared leverage against normalized squared residuals for cross-section regressions on annual means that were reported in the last two columns of Table 4. The leverage measure indicates the extent to which an observation is influential for the regression results, while greater residuals denote outliers. Observations high in leverage (Belarus, Kyrgyz Republic, etc.) are not the ones that are large outliers (Gabon, Botswana, and the Democratic Republic of the Congo). Excluding the most important outliers or leveraged observations does not alter the results with respect to the sign and the significance of the WTO dummy (Table 8).

This list of reasons is by no means exhaustive. Still, the model's outcomes and manipulations lend some credence to the link between export restrictions, both from outside and within Russia, and the "anti-WTO tilt" in Russia's trade. In any case, the combination of the "overall" and "within" results reinforces a lack of broader evidence that *established* WTO members have been as receptive as others to Russia's exports.

Implications for Russia's WTO Accession

Assuming that the gravity model results can proxy Russia's export structure following its WTO entry, one may project Russia's putative trade developments in a "postentry" world. The numerical coefficients in our regressions imply that, in the long run, Russia's exports to WTO members could expand by a very large amount—

Table 8. OLS Cross-Country Regression on Annual Means, 1995–2002
(Coefficients on constant and border not reported)

	Without Outliers*		Without High-Leverage Observations**	
Number of observations	160		164	
F (8, 151)	101.90		70.41	
Prob >F	0.00		0.00	
Adjusted *R*-squared	0.84		0.75	
Exports (dep. var.)	Coefficient	Standard error	Coefficient	Standard error
WTO96	–1.16	0.28	–0.97	0.35
EAEC	1.10	0.80	—	—
COMECON	2.20	0.47	2.28	0.61
USSR	1.82	0.56	1.95	0.73
GDP per capita	0.20	0.08	0.15	0.10
Distance	–1.30	0.19	–1.35	0.25
GDP	0.98	0.06	0.99	0.07

Source: Authors' calculations.

*On the basis of the "leverage vs. squared residuals" plot, outliers were identified as including St. Vincent and the Grenadines, Bahrain, Cape Verde, Botswana, Gabon, Democratic Republic of Congo, Oman, and Liberia.

**On the basis of the "leverage vs. squared residuals" plot, high-leverage observations were identified as including Belarus, Kazakhstan, Kyrgyz Republic, and Tajikistan.

according to most regressions, by around 50 percent.[13] A major issue is the extent to which this correction of trade diversion would occur through export expansion as opposed to reorientation. In the latter case, there would be a substitution effect, and a smaller increase in Russia's exports to the WTO countries would suffice to align its export structure. In all likelihood, both the expansion and substitution effects would be present, but a precise configuration could not be modeled without a more detailed structure. Generically, though, there has to be some export expansion. In particular, if the economic reasons for our underlying results are trade restrictions of any type, it is unlikely that the trade-off between exportables and nontraded goods would be unaffected after the restrictions on exports have been relaxed.

In any case, our estimates indicate an asymptotic upper bound of the long-run effect. In the short run, the WTO-related export expansion may be further limited, given the evidence that, in some key export-oriented sectors (such as metals), Russia's capacity utilization rate has approached fairly high levels. In this situation, a further export expansion hinges on substantial investment and could be delayed.

Another caveat with respect to future projections is our narrow interpretation of the WTO membership dummy as the main causal channel in the model. The WTO variable may be proxying other factors than membership; for example, Russia's recognition as a market economy for antidumping purposes. Russia obtained this status with most major countries by late 2002, without becoming a WTO member. Thus, the magnitude of the WTO dummy should be interpreted not so much as a precise elasticity but more as an indication of the role of this broad channel. Note that with our data set ending in 2002, the market status argument could not be checked directly.

Implications for WTO-Related Multilateral Studies

In what way can the negative WTO result for Russia be generalized? The apparent persistence of an anti-WTO bias in Russia's trade structure may call into question the WTO's trade-promoting role and relevance, including the spillover effects emphasized by Subramanian and Wei (2003). But this reasoning leads to an uneasy puzzle, since WTO membership in general and for particular countries is considered important, not only in academic and public circles but also in practical negotiations, decisions, and outcomes.[14] We believe that our results point to why the WTO may be relevant. While our analysis cannot prove or disprove the insights of Rose (2004), it suggests that the situation may be more complex and that focusing on averages may be insufficient for understanding the WTO's exact role and the associated policy issues.

In particular, we would argue that three asymmetries (in addition to those emphasized by Subramanian and Wei, 2003) may be important in generalizing

[13]Our baseline assumption is that, in the long run, the gravity model coefficient on the WTO would approximate zero, which is roughly consistent with the multilateral results of Rose (2004).

[14]The recent debates on the WTO's role in job outsourcing in industrialized countries and China's export expansion are cases in point.

insights from our results. First, the asymmetry between members and nonmembers should be studied more carefully. Our negative WTO result for Russia's trade pattern may crucially depend on Russia being a nonmember country. It is thus possible that a strong anti-WTO bias for some country combinations is (largely) offset by pro-WTO results for other combinations, even if one controls for the usual gravity factors.

The second type of asymmetry is between the policy response to larger and smaller countries, as it may be quite logical that larger countries like China or Russia may be generating more concern over their export potential and, hence, may face more restrictions from third countries. This may explain why our Russia-centered results are more significant than Rose's (2004) multilateral average for nonmembers.

Third, there are asymmetries based on political and systemic alliances that affect WTO members and nonmembers differently and at the same time influence the direction of trade. As an example, a stark division of the world into (pro-)socialist (largely non-WTO) and capitalist (largely WTO) camps for much of the postwar period may substantially bias the results of the long-term models and may have contributed to the emergence of localized centers of gravity. These issues may not be fully accounted for in gravity models, and a country-centered model seems better suited than multilateral gravity models to include essential policy-cum-political dummies. For this purpose, the role of country-specific fixed-effects results becomes more important, and the results are more positive for the WTO's trade-promoting role both in the multilateral and the Russia-centered framework.[15]

Interestingly, our evidence on the specific asymmetry between developed and developing countries emphasized by Subramanian and Wei (2003) differs from that of the authors. While they linked the lack of trade openness within the WTO to developing countries, Russia's exports to the industrialized WTO members appeared to be disadvantaged to the same or even greater extent than those to developing WTO countries. This in part reflects our "within" result that Russian exports seemed to benefit (albeit slightly) from trade liberalization of new WTO members, all of which were either developing or transition countries. Thus, our results imply that WTO membership may be particularly important for access to developed country markets by current nonmembers.[16]

More generally, our results point to the possibility of a domino effect in the expansion of WTO membership that is akin to the similar effect estimated in the operation of regional trade arrangements (RTAs) (see Sapir, 2001). The driving force behind the operation of the domino effect is the costs of nonmembership, which are countered by applying for membership in an RTA. As in any RTA,

[15]Still, the evidence for a positive WTO impact in fixed-effects regressions is fairly weak, partly because the coefficient captures a shorter-run "within" effect, as opposed to the longer-run "between" effect of the OLS regressions.

[16]This conclusion would run counter to Rose's (2004) point that developed WTO country members indiscriminately extend trading preferences to nonmembers; instead, it would support a study by UNCTAD (2004) arguing that the world's poorest countries are receiving only limited benefits from preferential trade schemes designed to help them, because of gaps in coverage and restrictive rules of origin.

incomplete WTO membership creates a wedge between members and nonmembers that may turn out to be costly for the latter.

In sum, we are more comfortable with this complex view, which also goes some way toward explaining the WTO "irrelevance puzzle." Thus, the WTO appears relevant and may have inner trade-promoting features that appear to average out in the simple gravity comparisons. At the same time, we would not want to be too categorical on this point, since Russia's specific circumstances as a nonmember and the short period of analysis weaken comparability with the long-term multilateral studies.

V. Concluding Remarks

We show that Russia's export structure, after controlling for gravity effects and several country- and region-specific factors, was tilted significantly away from WTO members in 1995–2002. The evidence of trade reorientation toward the predicted pattern during the same period is quite weak, as it affects only new WTO members and does not apply to Russia's exports to advanced economies. We reason that these results may be at least partially related to the external and possibly domestic constraints on Russia's exports, which, in turn, depend on Russia's continued nonmembership in the WTO. This would suggest that Russia's WTO accession may redirect and/or expand Russia's exports substantially, although possibly only in the very long term. While it cannot be fully ruled out that Russia's historical, cultural, or other idiosyncratic factors may bias our results, various methods and controls for most country- and region-specific observable effects did not eliminate the significance of the negative influence of the WTO variable.

Our findings are related to two strands of WTO-related research: global and Russia-specific. Globally, this is an interesting case study within a general debate between Rose (2002, 2004) and other authors on the extent of the WTO-induced gains from trade. Our results suggest that the WTO impact is quite relevant but likely combines trade promotion among members with some implicit trade frictions between members and nonmembers, while nonmembers may have additional reasons to trade with each other that cannot be captured by the gravity model. At the same time, our analysis does not formally prove or disprove the multilateral model insights, and further rigorous testing is required.

With respect to Russia-specific issues, our results suggest tangible long-term trade gains for Russia from its WTO accession, not least *because* of the highlighted friction between members and nonmembers. Although the time profile of Russia's gains is uncertain, waiting to join the WTO is likely not an attractive option, given the weak pace of the underlying reorientation to date. While generalizations may be misleading, China's early experience seems to (albeit casually) confirm the WTO membership's strong effect on exports.[17] It also should be cautioned that the empirical model is a simplification, in that it treats WTO accession

[17]In 2002–03, according to the IMF *Direction of Trade Statistics,* China's annual growth in exports and net foreign direct investment was 27 percent and 14 percent, respectively, compared with corresponding annual averages of 17 percent and 7 percent over the decade prior to its WTO entry.

as a binary choice (accession or nonaccession), while the reality is more complicated, as the effect of Russia's WTO accession obviously depends on the outcome of the negotiations.

Still, at the very least, we have shown that (1) there is an underlying bias in Russia's trade structure that may need to be addressed, and (2) entering the WTO appears to be the most logical way to address it. Regarding current trade policy, Russia should, through WTO membership, benefit from a stronger focus on ensuring uniformity and a level playing field across regional trading patterns. The sheer size and diversity of the country favor trade integration via WTO membership rather than through regional arrangements. Also, Russia's entry would make the WTO nearly universal, thereby possibly giving a multilateral boost to global trade.

Table A.1. Chronology of Russia's WTO Accession Process

Stages of Accession	
Request for accession	1993
Memorandum on the foreign trade regime to be submitted to the working party	March 1994, May 2001, November 2001
Questions and replies: Submission of the memorandum followed by questions from the members of the working party and answers from Russia	June 1995, June 1996
Working party multilateral negotiations to determine the general conditions of accession, including commitments to observe WTO rules and the length of transitional periods necessary to ensure their implementation	20 meetings of the working party in 1995–2003
Working party bilateral negotiations. Each accession working party makes decisions by consensus; thus, Russia must reach an agreement with each member of the working party.	Agreements on market access concluded with several countries, including with the European Union in May 2004.
The accession package: The results of the negotiations are reflected in three documents that form the accession package: —Report of the working party that contains a summary of proceedings and conditions of entry. —Protocol of accession. This document provides all the general terms and conditions of membership, including the areas of the administration of trade regime, nondiscrimination, the use of nontariff measures, and so on. —Schedules of market access commitments in goods and services that reflect the agreements reached with the members of the working party.	Draft working party reports (March 2002, November 2002, May 2003)

Source: Authors' elaboration based on www.wto.ru.

Table A.2. Definitions of Variables

Variable	Definition
Log exports	Natural logarithm of Russia's annual exports in 1995–2002 deflated by U.S. CPI
Log turnover	Natural logarithm of Russia's annual trade turnover in 1995–2002 deflated by U.S. CPI
Log product GDP	Natural logarithm of the product of Russia's GDP and that of Russia's trade partners in 1995–2002 deflated by U.S. CPI
Log product GDP per capita	Natural logarithm of the product of Russia's GDP per capita and that of Russia's trade partners in 1995–2002 deflated by U.S. CPI
Log distance	Natural logarithm of the distance from Moscow to the capitals of Russia's trade partners
COMECON	Dummy variable (1=former COMECON member (East Germany excluded), 0=not a former member of COMECON)
WTO	Dummy variable (1=member of the WTO, 0=not a member of the WTO)
USSR	Dummy variable (1=former member of the USSR, 0=not a former member of the USSR)
EAEC	Dummy variable (1=member of EAEC, 0=not a member of EAEC)
Developed	Dummy variable (1=developed country, 0=developing country)
GSP	Dummy variable (1=trading partner accords GSP treatment to Russia, 0=trading partner does not accord GSP treatment to Russia)
Island	Dummy variable (1=island country, 0=non-island country)
Landlocked	Dummy variable (1=landlocked country, 0=non-landlocked country)
Log product area	Natural logarithm of the product of Russia's area (in sq. km) and that of Russia's trade partners
Oil	Dummy variable (1=importer of oil from Russia in 1995, 0=all other countries)
Steel	Dummy variable (1=importer of steel from Russia, 0=all other countries)
GATT-94	Dummy variable (1=member of the WTO as of January 1, 1995; 0=all other countries)
Import duties	Natural logarithm of the average import tariff in Russia's trading partners

Source: Authors' calculations.

REFERENCES

Alekseev A., N. Tourdyeva, and K. Yudaeva, 2003, "Estimation of Russia's Trade Policy Options with the Help of the Computable General Equilibrium Model," CEFIR Working Paper No. 42 (Moscow: Center for Economic and Financial Research). Available via the Internet: www.cefir.org.

Anderson, J., and E. van Wincoop, 2003, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review,* Vol. 93 (March), pp. 170–92.

Berglöf, E., and others, 2003, *The New Political Economy of Russia* (Cambridge, Massachusetts: MIT Press).

Bergstrand, J.H., 1985, "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Review of Economics and Statistics,* Vol. 67 (3), pp. 474–81.

Blanchard, O., and M. Kremer, 1997, "Disorganization," *Quarterly Journal of Economics,* Vol. 112 (November), pp. 1091–126.

Daviddi, R., and E. Espa, 1996, "The Liberalization of the Foreign Trade of the Russian Federation and the Accession to the WTO," in *Russia and the World Economy,* ed. by G. Mureddu and M. T. Salvemini (Rome: La Sapienza).

Dean, J., and I. Eremenko, 2003, "Will WTO Membership Really Improve Market Access for Ukrainian Exports?" paper presented at the 37th annual meeting of the Canadian Economics Association, Ottawa, May.

Elborgh-Woytek, K., 2003, "Of Openness and Distance: Trade Developments in the Commonwealth of Independent States, 1993–2002," IMF Working Paper 03/207 (Washington: International Monetary Fund).

Granville, B., 1995, *The Success of Russian Economic Reforms* (London: Royal Institute of International Affairs).

Greenaway, D., and C. Milner, 2002, "Regionalism and Gravity," Leverhulme Centre Research Paper No. 2002/20 (Nottingham, England: University of Nottingham, Leverhulme Centre).

Gros, D., and A. Steinherr, 1995, *Winds of Change: Economic Transition in Central and Eastern Europe* (London; New York: Longman).

Hufbauer, G.C., and B. Oegg, 2003, "The Impact of Economic Sanctions on U.S. Trade: Andrew Rose's Gravity Model," International Economic Policy Briefs No. PB03-4 (Washington: Institute for International Economics).

International Monetary Fund, *Direction of Trade Statistics,* 1995–2003.

Linnemann, H., 1966, *An Econometric Study of International Trade Flows* (Amsterdam: North-Holland).

Lissovolik, B., and Y. Lissovolik, 2004, "Russia and the WTO: The 'Gravity' of Outsider Status," IMF Working Paper 04/159 (Washington: International Monetary Fund).

Lissovolik, Y., and N. Liventsev, 2002, *Actualnyye Problemy Prisoyedineniya Rossii k VTO* (Moscow: Ekonomika).

Matyas, L., 1997, "Proper Econometric Specification of the Gravity Model," *The World Economy,* Vol. 20 (May), pp. 363–68.

Owen, D., and D. Robinson, eds., 2003, *Russia Rebounds* (Washington: International Monetary Fund).

Rose, A.K., 2002, "Do We Really Know That the WTO Increases Trade?" NBER Working Paper No. 9273 (Cambridge, Massachusetts: National Bureau of Economic Research).

———, 2004, "Do We Really Know That the WTO Increases Trade?" *American Economic Review,* Vol. 94 (March), pp. 98–114.

———, 2005. Available via the Internet: http://faculty.haas.berkeley.edu/arose/.

Russian Academy of Sciences and the National Investment Council (RAS), 2002, *The Implications of Russia's Accession to the WTO for the National Economy* (Moscow: RAS).

Sapir, A., 2001, "Domino Effects in Western European Regional Trade, 1960–92," *European Journal of Political Economy,* Vol. 17 (June), pp. 377–88.

Subramanian, A., and S.J. Wei, 2003, "The WTO Promotes Trade, Strongly but Unevenly," IMF Working Paper 03/185 (Washington: International Monetary Fund).

Tarr, D., J. Jensen, and T. Rutherford, 2002, "Economy-Wide and Sector Effects of Russia's Accession to the WTO," paper prepared for the New Economic School Tenth Anniversary Conference, "The State of Economics and of Transition," Moscow, December.

UNCTAD, 2004, "Trade Preferences for LDCs: An Early Assessment of Benefits and Possible Improvements" (UNCTAD/ITCD/TSB/2003/8). Available via the Internet: http://www.unctad.org/Templates/webflyer.asp?docid=4293&intItemID=1397&lang=1&mode=highlights.

van Selm, B., 1997, *The Economics of Soviet Break-Up* (London: Routledge).

Wang, Z.K., and L.A. Winters, 1992, "The Trading Potential of Eastern Europe," *Journal of Economic Integration,* Vol. 7 (Autumn), pp. 113–36.

IMF Staff Papers
Vol. 53, No. 1

Rent Seeking

SHANKHA CHAKRABORTY AND ERA DABLA-NORRIS*

This paper examines the relationship between rent seeking and economic performance when governments cannot enforce property rights. With imperfect credit markets and a fixed cost to rent seeking, only wealthy agents choose to engage in it, as it allows them to protect their wealth from expropriation. Hence, the level of rent seeking and economic performance are determined by the initial distribution of income and wealth. When individuals also differ in their productivity, not all wealthy agents become rent seekers, and the social costs of rent seeking are typically lower. In both cases, multiple equilibria with different levels of rent seeking and production are possible. [JEL D23, D31, D72, O11]

Lack of development, stagnant growth, insecure property rights, income inequality, and widespread rent seeking[1] are common features of many economies. One consequence of such rent seeking is that productive resources are diverted toward appropriative activities, resulting in a misallocation of resources in the economy. Rent-seeking activities, such as corruption and tax farming, can reduce growth by lowering overall incentives and opportunities for production and investment. Cross-country studies find that countries in which corruption and rent seeking are rampant suffer from lower capital accumulation, productivity, and growth (Mauro, 1995; Keefer and Knack, 1997; and Hall and Jones, 1999). Extensive rent seeking and insecure property rights are also associated with substantial inequal-

*Shankha Chakraborty is an Assistant Professor of Economics at the University of Oregon; Era Dabla-Norris is an Economist in the IMF's Middle Eastern and Central Asia Division. The authors are grateful to Andrew Feltenstein and Jean-Francois Ruhashyankiko for their valuable comments.

[1]Rent seeking refers to largely unproductive, expropriating activities that bring positive returns to the individual but not to society (Krueger, 1974).

ity in income and wealth.[2] Olson (1965) argued that corruption and rent seeking are more likely in unequal societies. Easterly (2002) and Keefer and Knack (2002) show that the adverse effect of inequality on growth operates through weak institutions and lax enforcement of property rights.

While the empirical evidence suggests links among rent seeking, poor economic performance, and income inequality, the persistence of rent seeking in many countries begs the question of what drives entry into such activities. In many developing and transition countries, it is the relatively wealthy who choose rent-seeking activities—such as the government bureaucracy, the army, and the police—rather than engaging in productive and entrepreneurial activities. In other words, the rent seekers are exactly those who might otherwise become the first capitalists.

This paper presents a model that analyzes the relationships among rent-seeking behavior, wealth distribution, and economic performance when the government cannot protect property rights. We show how an individual's initial income (wealth) drives entry into rent seeking and how the size of the rent-seeking sector affects the level of distortion in an economy. In the model, participation in rent seeking is a costly activity. However, rent-seeking institutions provide protection against expropriation by others. With insecure property rights, the incentive to engage in rent seeking is relatively stronger for the wealthy, who have more to lose from expropriation than poorer individuals. In the presence of borrowing constraints and a fixed cost for entry into rent seeking, the size of the rent-seeking sector and corresponding efficiency losses in society depend on the initial distribution of wealth.

The fixed cost for rent seeking can be viewed as analogous to the purchase of weapons used for protection and for offense. In this interpretation, arms enable agents to protect their wealth from expropriation by others but also to extract rents from other agents. Hence, when property rights protection is poor or ineffective, as is the case in many developing and transition countries, the rich have an incentive to buy arms (i.e., enter into rent-seeking activities such as the government bureaucracy). The purchase of these arms, however, enables them to prey on other agents in the economy. Alternatively, one can think of this in terms of the costs of purchasing political power and influence. Political markets in the real world involve significant transaction costs of lobbying, obtaining exemptions, and large fixed costs of political organization (Downs, 1957; and Huntington, 1968). In this interpretation, wealthy agents or the elites can then use their political power to extract rents from the most productive sectors and to influence government decisions that are favorable to their interests.

The assumption of increasing returns in rent seeking or political influence is consistent with historical evidence. In Republican Rome and in 17th century France, an important prerequisite for engaging in tax farming was the availability of sufficient capital that enabled wealthy individuals to advance funds to rulers and to collect taxes (Braudel, 1982; and Levi, 1989). Baumol (1990) notes that in Mandarin China and in medieval Europe, government service—with its potential

[2]See Benabou (1997) for a survey on the inverse relationship between inequality and growth.

for illegal personal enrichment—was the principal career choice for many wealthy individuals. Engerman and Sokoloff (2002) provide numerous historical examples in which a small wealthy class was able to capture the state. Wealth requirements for voting and differences in the distribution of political power in colonial Latin America are widely perceived as affecting access to economic opportunities in ways that, in turn, have influenced long-term institutional development and perpetuated inequality (de Ferranti and others, 2003).

Wade (1985) found that individuals in India paid thousands of dollars for positions with the power to allocate supposedly free water to farmers, since these jobs gave them monopoly rights to charge for water. In many developing countries, civil service positions are purchased at high prices compared with the annual salary for the position. Casual observation suggests that in many countries it is common for government officials and politicians to own businesses run by either themselves or their relatives, and to protect such businesses from corrupt practices and other forms of expropriation by virtue of their positions. The association between wealth and rent seeking in the presence of poor property rights protection is also receiving increasing attention in Eastern Europe's transition to capitalism, particularly in the context of privatization (Sonin, 2003; and Hoff and Stiglitz, 2004).

This paper develops a simple general equilibrium model with imperfect capital markets and a nonconvex rent-seeking technology. Agents in this model have identical preferences and abilities and differ only with respect to their initial income. We assume that agents can operate in one of two sectors in the economy: rent seeking and production. Entry into rent seeking requires the payment of a fixed cost. Payment of this cost allows rent seekers to appropriate some portion of the surplus generated by productive economic activity and to save a desired portion of their initial income (through hoarding), thereby protecting their wealth from expropriation by others. We show that for a certain range of parameter values there is a unique interior equilibrium with rent seeking, which depends on the initial distribution of wealth in society.

In an extension, we introduce heterogeneity in productive ability. This introduces a trade-off, since wealthier agents can choose to become producers even though they can afford the costs of rent seeking. We show that rent seeking in such an economy typically involves lower social costs, but the main implications of our basic model generalize.

This paper is related to several strands of research. The role of investment indivisibilities and credit market imperfections in explaining the relationship between inequality and economic development has been studied by Galor and Zeira (1993) and Banerjee and Newman (1993), among others. Another literature explicitly examines the associations among wealth, political power, and redistribution. Verdier (1996) and Rodriguez (1999) assume fixed costs in political participation to generate increasing returns and the separation of the rich and poor in terms of political influence. Specifically, Verdier shows how the wealthy influence the direction of income transfers in society and highlights the role of the initial distribution of wealth in determining this pattern. In this respect, our analysis of rent seeking by the rich in the presence of insecure property rights can be viewed as complementary to his.

Our paper is closely related to economic models of rent seeking and economic development. In Baumol (1990); Murphy, Shleifer, and Vishny (1993); Acemoglu (1995); Baland and Francois (2000); and Mehlum, Moene, and Torvik (2003), for example, individuals choose between productive and rent-seeking activities by comparing their relative rewards. These rewards are, in turn, determined by the allocation of individuals between the two activities. Our paper extends the literature by providing an explanation for the observed association between wealth and rent seeking.

Finally, this paper is related to the literature on conflict. Grossman (1991, 1994) shows how inequality induces the poor to engage in predatory activities and the rich in investments in defense, diverting resources from productive activities. Skaperdas (1992) argues that the more productive groups of individuals may have a comparative disadvantage in the process of appropriative competition. The view proposed in this paper is that in the absence of adequate property rights protection, it is the rich who are likely to benefit from appropriation. Grossman and Kim (1997) assume that weapons are used either for predation or for defensive fortifications. In contrast, this paper views predation and the deterrence of predation as complementary activities.

I. Model

Environment and Technology

Consider an economy comprised of a continuum of two-period-lived agents distributed over the interval [0,1]. Each agent is endowed with $w > 0$ units of the consumption good when young. The initial distribution of goods endowment in the economy is represented by the cumulative distribution function $\Gamma:[\underline{w}, \bar{w}] \rightarrow [0,1]$. Agents are risk neutral and have identical preferences and abilities. They differ only in initial income. The lifetime expected utility of an agent is given by $U(c_1, c_2) = c_1 + E(\tilde{c}_2)$, where E is the expectations operator and c_1 and $\tilde{c}_2$ denote consumption of the economy's single good in each period of life. Second-period consumption is uncertain, because it depends on the degree of rent seeking in the economy.

Each producer has access to a standard concave production technology that yields $f(i)$ units of the consumption good in period 2 of the producer's life from an investment of i units in period 1, where $f(0) = 0$ and $f'(0) = \infty$. Producers, however, face appropriation of some share of their market production by rent seekers. There is no law enforcement (we briefly discuss law enforcement in Section II) and no government taxation.

Let $0 < \gamma < 1$ denote the exogenously given proportion of market production that can be extracted from each producer. The idea here is that in the absence of adequate enforcement, producers do not possess complete property rights over their output. We can, therefore, think of γ as the profit loss due to taxes, bribes, outright expropriation, or, more generally, as the cost of doing business in a rent-seeking society. In reality, this fraction may be an endogenous function of the level of economic development, the effectiveness of law enforcement, or the efficiency of the institutional background of the economy.

The description of the rent-seeking sector is similar to that in Mehlum, Moene, and Torvik (2003). Let π^P denote the probability of being approached by a rent seeker. The second-period expected return to a producer is then given by $(1 - \pi^P \gamma) f(i)$. Let n denote the mass (fraction) of rent seekers in the economy. Each rent seeker can expropriate a share γ of the production of $\lambda \geq 1$ producers in the economy. This assumption implies that a rent seeker can expropriate from more than one producer. However, each producer is only approached by a single rent seeker.[3] The probability π^P can then be defined as the ratio of the total number of rent-seeking cases in the economy divided by the mass of producers, $(1 - n)$. Assuming full information, π^P can be defined as

$$\pi^P = \min\left\{\frac{\lambda n}{(1-n)}, 1\right\}. \tag{1}$$

Entry into rent seeking requires a fixed cost of θ units of the consumption good when young. Payment of this cost allows rent seekers to tax market production in the second period of their lives. This assumption is vital to the results below and captures increasing returns to rent seeking. Because of capital market imperfections, agents are unable to borrow to finance entry into rent seeking using their future income as collateral. Therefore, in order to belong to the class of rent seekers, an agent must have a starting level of wealth $w \geq \theta$.

Rent seekers save through a simple technology that returns xs units of goods in period 2 of their lives for an investment of s units when young. The inputs of rent seekers are assumed to be unobservable, which implies that one rent seeker cannot expropriate the goods controlled by another. Therefore, one may think of x as the return on hoarding that allows rent seekers protection from theft by others. The net return from hoarding is low but positive, that is, $x > 1$.[4]

The probability π^R that a particular rent seeker is the first to approach a producer is defined by the ratio of the total mass of producers, $(1 - n)$, divided by the total rent-seeking cases in the economy, λn. This probability can be expressed as follows:

$$\pi^R = \min\left\{\frac{(1-n)}{\lambda n}, 1\right\}. \tag{2}$$

The ratio $(1 - n)/\lambda n$ also captures congestion in rent-seeking activity. When $(1 - n) > \lambda n$ (i.e., when the fraction of producers in the economy exceeds total rent-seeking cases), there is no crowding out among rent seekers, with probability 1 each rent seeker can expropriate from a producer ($\pi^R = 1$). When there are fewer

[3]A producer never gets approached by more than one rent seeker here. One interpretation of this assumption is that each rent seeker implicitly provides protection against extortion by others, as in Mehlum, Moene, and Torvik (2003). In our model, each rent seeker can extract rent from more than one producer, but we abstract from the issue of protection.

[4]As we note below, $f'(i) > x$ in a positive rent-seeking equilibrium when γ is high. In this case, rent seeking competes with productive sectors for resources, resulting in further misallocation of resources in the economy.

rent-seeking cases in the economy than there are producers, each rent seeker can capture the full potential rent given by the rent-seeking technology from the λ producers approached. However, when $\lambda n > (1 - n)$, crowding out occurs, since each rent seeker competes with others for rents and the expected rent to each falls.

Let $\tilde{n}$ denote the value of n for which the total fraction of rent-seeking cases in the economy equals the fraction of producers. The probability π^P that a producer is approached by a rent seeker can then be expressed as

$$\pi^P = \begin{cases} \lambda n/(1-n) & \text{for } 0 \le n \le \tilde{n} \\ 1 & \text{for } n \ge \tilde{n} \end{cases}, \tag{3}$$

and the probability π^R that a rent seeker expropriates from a producer as

$$\pi^R = \begin{cases} 1 & \text{for } 0 \le n \le \tilde{n} \\ (1-n)/\lambda n & \text{for } n \le \tilde{n} \end{cases}. \tag{4}$$

Notice that $\tilde{n} = \dfrac{1}{1+\lambda} \le ½$ for $\lambda \ge 1$.

Occupational Choice

The timing of the model is as follows. In the first period of their lives, individuals choose their occupation (production or rent seeking) and consume. At the beginning of the second period of their lives, there is a matching process between rent seekers and productive agents. Once that process resolves, individuals consume.

First, consider the optimization problem faced by an agent who chooses to become a producer. The budget constraints faced by a producer in each period of life are

$$c_1 = w - i$$

$$\tilde{c}_2 = \begin{cases} (1-\gamma) f(i) & \text{with probability } \pi^P, \\ f(i) & \text{with probability } (1-\pi^P) \end{cases}$$

where π^P is the probability of being approached by a rent seeker from equation (3). The producer maximizes expected utility, given an initial endowment *w:*

$$\underset{i}{Max}\,(w-i) + \pi^P (1-\gamma) f(i) + (1-\pi^P) f(i).$$

The first-order condition is given by

$$\left(1 - \gamma\pi^P\right) f'(i) = 1, \tag{5}$$

where $0 \leq \gamma\pi^P < 1$. For a positive probability of theft, the marginal return from investment exceeds 1: $f'(i) = 1/(1 - \gamma\pi^P) > 1$. If γ and π^P are significantly high, the marginal return from production also dominates the marginal return from hoarding, that is, $f'(i) > x$.

Let $i^* = i(\gamma\pi^P)$ denote the optimal investment choice that satisfies equation (5). Differentiating equation (5) with respect to $\gamma\pi^P$ shows that investment is decreasing in rent seeking because it lowers the expected return from investment.[5] Note also that i^* is the same for all producers and is independent of a producer's initial wealth.[6]

The indirect utility from being a producer can then be written as

$$V^P = w + \left[\left(1 - \gamma\pi^P\right) f(i^*) - i^*\right], \tag{6}$$

the last term being positive for optimal investment choice.

We turn next to the rent seeker's problem. Each rent seeker can expropriate a γ fraction of a producer's second-period income. We assume random matching between rent seekers and producers. Since all producers, irrespective of their initial endowment, earn the same income in the second period, the potential rent that can be appropriated from a producer is the same no matter which producer the rent seeker steals from. However, we later allow for heterogeneous income across producers (Section III). To simplify our future analysis, we assume that endowments are not *ex ante* observable; they are observed by rent seekers only when they attempt to expropriate producers' incomes. Hence, decisions about rent seeking are based on the expected rent that can be earned from a successful match with a producer, which is given by

$$R = \frac{\lambda \int_{\underline{w}}^{w^*} \gamma f(i^*) d\Gamma(w)}{\int_{\underline{w}}^{w^*} d\Gamma(w)},$$

where w^* (to be endogenously determined) is the cutoff level of wealth, w, for which agents choose to be producers. Since i^* is independent of wealth, this simplifies to

$$R = \lambda\gamma f(i^*). \tag{7}$$

The rent seeker's budget constraints are now given by

$$c_1 = w - \theta - s$$

$$\tilde{c}_2 = \begin{cases} R + xs & \text{with probability } \pi^R \\ xs & \text{with probability } (1 - \pi^R), \end{cases}$$

[5]Note that $di^*/d\left(\gamma\pi^P\right) = \dfrac{f'(i^*)}{\left(1 - \gamma\pi^P\right) f''(i^*)} < 0$ for a concave production function.

[6]This follows from our assumption of linear utility. Allowing for risk aversion does not fundamentally alter the insights of this paper but does complicate the analysis, especially for rent seekers' decisions.

where π^R is the probability of being the first to approach a producer (equation (4)), s represents the savings of the rent seeker, θ denotes the up-front fixed cost to rent seeking, and R (given in equation (7)) is the expected amount expropriated from a producer. The rent seeker maximizes expected utility, given an initial endowment w:

$$\underset{s}{Max}(w-\theta-s)+\pi^R(R+xs)+(1+\pi^R)xs.$$

The assumption that the gross return on hoarding exceeds 1 implies that a rent seeker will save the entire first-period income and consume only in the second period, that is,

$$s = w-\theta. \tag{8}$$

The indirect utility from being a rent seeker is then given by

$$V^R = x(w-\theta)+\pi^R R, \tag{9}$$

which, in equilibrium using equation (7), can be rewritten as

$$V^R = x(w-\theta)+\lambda\pi^R\gamma f(i^*). \tag{9'}$$

An individual chooses an occupation depending on which one generates higher lifetime utility. Let $\Omega(w)$ denote the difference between lifetime utility of an agent with endowment w if the agent engages in rent seeking and utility upon becoming a producer:

$$\Omega = V^R - V^P = x(w-\theta)+\lambda\pi^R\gamma f(i^*)-\left[(w-i^*)+\left(1-\gamma\pi^P\right)f(i^*)\right]. \tag{10}$$

The individual will choose to be a rent seeker if $\Omega > 0$. Differentiating Ω with respect to w and using $x > 1$ shows that the net utility from being a rent seeker is strictly increasing in the agent's wealth (see Figure 1).

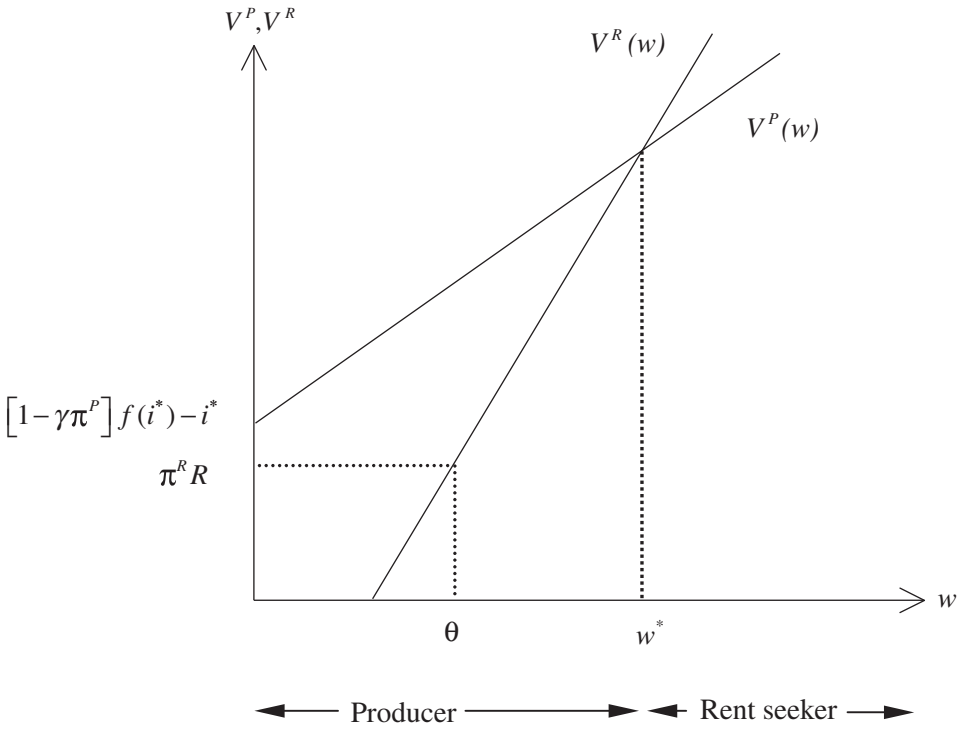
Equilibrium

In this subsection, we show that an individual's initial wealth determines occupational choice between rent seeking and production. Let w^* denote the threshold level of wealth such that $\Omega > 0$ for $w > w^*$, and n^* denote the equilibrium allocation of agents between rent seeking and production, such that

$$n^* = \int_{w^*}^{\bar{w}} d\Gamma(w) = 1-\Gamma(w^*). \tag{11}$$

An equilibrium in this environment is defined by a cutoff wealth level, w^*, and a number of rent seekers, n^*, such that agents with wealth w^* are indifferent between production and rent seeking (i.e., $V^R(w^*, n^*) = V^P(w^*, n^*)$, $i^*(n^*) = i(\gamma\pi^P(n^*))$ from equation (5), and equations (3), (4), and (11) hold.

Figure 1. Occupational Choice



We illustrate the general equilibrium using a diagram in (n^*, w^*) in Figures 2 and 3. Since an agent with wealth w^* is indifferent between the two occupations, the net utility from rent seeking is

$$\Omega(w^*, n^*) = x(w^* - \theta) + \lambda\pi^R(n^*)\gamma f(i^*) - \left[(w^* - i^*) + \left(1 - \gamma\pi^P(n^*)\right)f(i^*)\right] = 0, \tag{12}$$

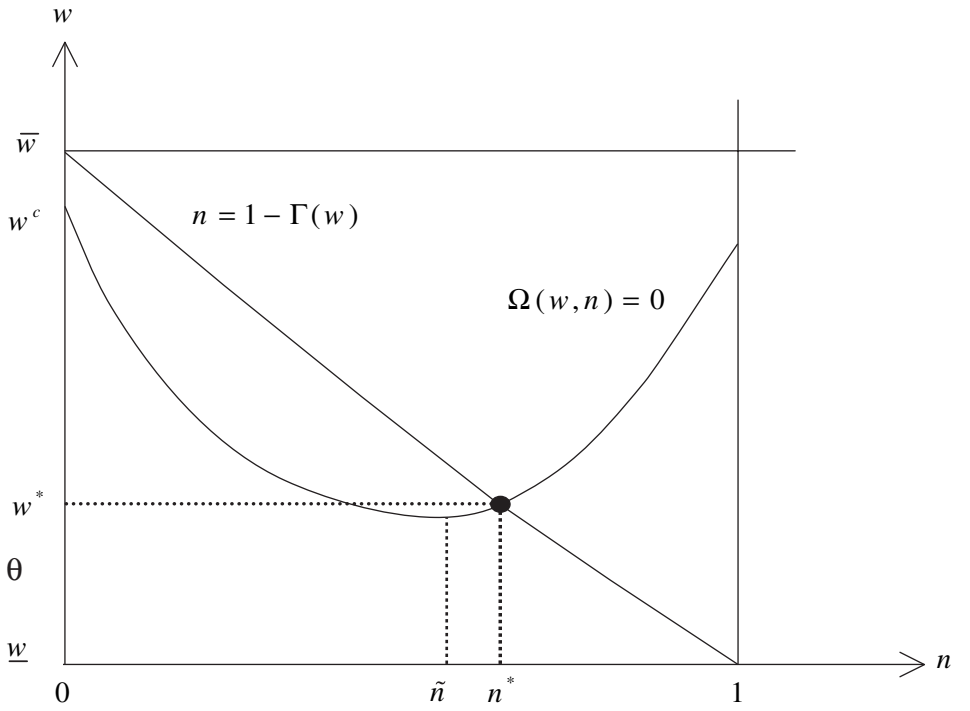
given n^*.

Note that as the fraction of producers in the economy decreases and the fraction of rent seekers increases, production declines but rents accruing to each rent seeker depend on whether or not crowding out sets in. For $n < \tilde{n}$, the fraction of producers in the economy exceeds total rent-seeking cases: There is no crowding out among rent seekers ($\pi^R = 1$), but producers face expropriation of their production with increasing probability ($\pi^P = \lambda n/(1-n)$), thereby decreasing the return to production. Using equations (3) and (4) and simplifying, equation (12) can be written as

$$\Omega(w^*, n^*) = w^*(x-1) - x\theta + i^* - \left[1 - \frac{\gamma\lambda}{1-n^*}\right]f(i^*) = 0. \tag{13}$$

$\Omega(w^*, n^*)$ is increasing in w^* and n^*. Hence, as illustrated in Figure 2, the zero net utility curve is downward sloping in (n^*, w^*) space for the range $n \leq \tilde{n}$.

Figure 2. Unique Rent-Seeking Equilibrium



With $n > \tilde{n}$, the probability of being approached by a rent seeker becomes 1 ($\pi^P = 1$), while the probability that a rent seeker can expropriate from producers starts to decline as rent seekers begin to crowd each other out ($\pi^R = (1 - n)/\lambda n$). Equation (12) now becomes

$$\Omega(w^*, n^*) = w^*(x-1) - x\theta + i^* - \left[1 - \frac{\gamma}{n^*}\right] f(i^*) = 0. \tag{14}$$

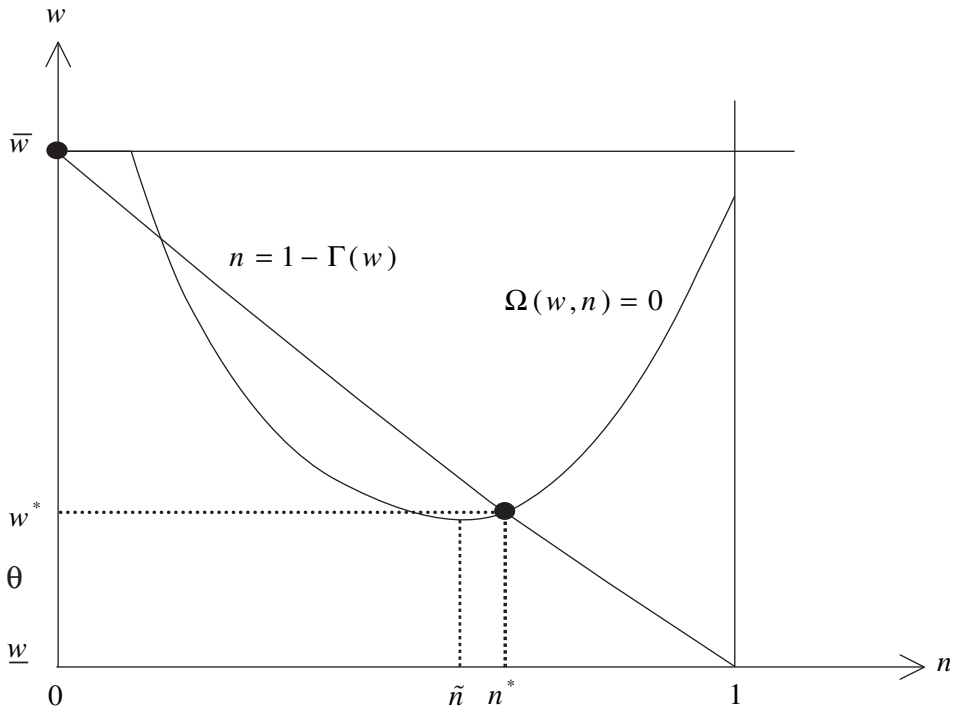
In this case, $\Omega(w^*, n^*)$ is increasing in w^* but decreasing in n^*. This means, for $n > \tilde{n}$, the zero net utility curve is upward sloping as in Figure 2.

Consider now the case where there is no rent seeking ($n = 0$).[7] At $n = 0$, equation (4) suggests that $\pi^R = 1$ and the first rent seeker to approach producers can capture $\lambda\gamma f(i^*)$ in rents, while $\pi^P = 0$. Let w^C denote the value of w that satisfies $\Omega(w, 0) = 0$. We can show that

$$w^C = \frac{1}{x-1}\left[x\theta - i^* + (1 - \lambda\gamma) f(i^*)\right].$$

[7]It is easy to see that an equilibrium in which all agents engage in rent seeking does not exist. With $n = 1$, $\pi^R = 0$ as rent seekers crowd each other out, while $\pi^P = 1$. As a result, the rent accruing to each rent seeker is zero. Moreover, marginal returns from production $f'(0)$ will dominate marginal returns from protected investment x.

Figure 3. Multiple Rent-Seeking Equilibria



For $w^C < \overline{w}$, as illustrated in Figure 2, the $\Omega(w, n) = 0$ curve intersects the $n = 1 - \Gamma(w)$ line once at the unique rent-seeking equilibrium with $n^* > 0$.[8] A sufficiently low θ or sufficiently high λ and γ make this equilibrium more likely. Intuitively, when costs of entering into rent seeking are low or the amount that can be expropriated is high, a career in rent seeking will be relatively more attractive compared with production. If λ is sufficiently greater than 1 such that each rent seeker can expropriate from a large number of producers, $\tilde{n}$ and, therefore, n^* would be lower than 0.5. This would correspond to the historical evidence of a small rent-seeking elite expropriating from a large segment of society (see Engerman and Sokoloff, 2002, for examples).

It is also possible that $w^C > \overline{w}$ at $n = 0$. If, for instance, the cost of entering into rent seeking, θ, is high relative to the expected return from expropriation (determined by λγ), we can get multiple equilibria in rent seeking and production. This case is illustrated in Figure 3, where the $\Omega(w, n) = 0$ curve intersects the $n = 1 - \Gamma(w)$ line twice. One equilibrium corresponds to the case of positive levels of rent seeking with $n^* > \tilde{n}$. A second equilibrium is one with no rent seeking ($n = 0$).

A third equilibrium also exists in Figure 3 with a positive level of rent seeking and production where $\Omega(w^*, n^*) = 0$. But such an equilibrium is unstable,

[8]It is also possible for n^* to be less than $\tilde{n}$ in Figure 2.

since a small change in n and w from their equilibrium values drives the economy to one of the two stable equilibria (with positive or zero rent seeking).

We summarize this discussion in the following two propositions (technical details are provided in the appendix).

Proposition 1. When the cost of rent seeking is low relative to expected returns (that is, $w^C > \bar{w}$), a unique equilibrium exists with a positive measure of agents engaged in rent seeking and production. When costs of rent seeking are relatively high ($w^C < \bar{w}$), a corner equilibrium with zero rent seeking definitely exists. In addition, an interior equilibrium with positive rent seeking may exist.

Now consider an equilibrium with positive rent seeking. As the discussion above makes clear, it is the relatively wealthy who engage in rent-seeking activities in such an equilibrium.

Proposition 2. In equilibrium, the greater the initial wealth of an agent, the more attractive a career in rent seeking relative to production.

Proposition 2 states that the rich become rent seekers. But it is not higher wealth alone that creates incentives for wealthy agents to enter into rent seeking. The ability to protect their wealth from expropriation by others (the slope of V^R is simply x) also plays a key role.[9] In the absence of credit markets, a higher initial wealth is of greater value to rent seekers than to producers, because it allows them to devote a larger amount of their endowment to hoarding once θ is paid. Specifically, the existence of a protection technology against expropriation by others implies that wealthier agents endure a smaller lifetime utility sacrifice in paying the fixed cost for rent seeking and hoarding, and hence are more willing to enter into this occupation. Therefore, in the presence of the rent-seeking fixed cost, θ, and in the absence of credit markets, only relatively wealthy agents choose to become rent seekers.

This result extends to alternative formulations of capital market imperfections or the rent-seeking technology. All that is needed is that the marginal benefit of entering into rent seeking is higher for wealthier individuals. Thus, despite the fact that agents are ex ante identical in terms of preferences and abilities, two classes of agents emerge in equilibrium. It is the initial income (w) of an agent that determines the agent's decision to engage in rent seeking or production, and how much to consume and save. Hence, the initial distribution of endowments determines aggregate output in the economy and the measure of agents engaged in rent seeking.

Note that the weaker property right protection is (that is, the higher γ), the more attractive rent seeking becomes as a profession and the less attractive production becomes. In Figure 3, the effect of a higher γ is to shift down the net utility curve, resulting in more agents switching to rent seeking (lower w^*, higher n^*).

When hoarding has a lower marginal product than marketed production ($x < f'(i^*)$), the greater the proportion of rent seekers, and the smaller the aggregate output. Hence, the total output in the economy is negatively related to the size of the rent-seeking sector. Net output—that is, aggregate output net of rent-seeking costs (deadweight losses)—is even lower.

[9]This slope will also depend on wealth if, for example, rents expropriated depend on the rent seeker's own wealth, as in Sonin (2003).

II. Key Assumptions

Having analyzed the equilibrium of this model, we now briefly discuss the assumptions underlying the model and how changes in parameter values affect equilibrium outcomes. The results of the model are robust to alternative specifications of the rent-seeking technology. In our model, returns to rent seeking, past the fixed cost θ, are not sensitive to the amounts invested. Alternatively, returns to rent seeking could be related to the amounts invested in such activities. Allowing rent seekers to expropriate a fixed proportion of producers' output simplifies our analysis but does not affect any of the major results.

The model also assumes a specific form for the function that relates the amount of rent obtained by each rent seeker from expropriation to the proportion of rent seekers in the economy. This assumption implies that the amount expropriated by each rent seeker is decreasing in n because of increased competition among rent seekers. This extreme form of crowding out, again, is not crucial for the results of this model.

In the model, agents are assumed to be unable to finance their entry into rent seeking through borrowing. Credit markets in many developing countries are typically characterized by high collateral requirements. An agent seeking to obtain a loan to finance the entrance fee into rent seeking would need to have substantial capital to meet the collateral requirements. This suggests that even in the presence of credit markets it is the relatively wealthy who will be able to obtain such loans.

The model also abstracts from law enforcement as the rent seekers' probability of being caught and punished is implicitly set equal to zero. Clearly, effective law enforcement would reduce the attractiveness of rent seeking. Law enforcement can easily be captured in the model by making γ a decreasing function of the economy-wide resources devoted to enforcement. If, for instance, such enforcement is financed through a lump sum tax on producers, the larger the size of the rent-seeking sector, the larger the tax required to finance a given level of enforcement would be. As a result, even with positive law enforcement, the implied trade-off for producers between lower expropriation and higher taxes suggests that rent seeking will not be eliminated in equilibrium.

Our results require two crucial assumptions: (1) that there are indivisibilities in rent seeking and (2) that rent seekers have access to a "protection" technology. The assumption of indivisibilities in rent seeking is vital to the results of the paper, as it captures the increasing returns necessary to generate a split between the poor and rich in terms of rent seeking. These assumptions together ensure that the distribution of initial income determines the choice between rent seeking and production.

III. Extensions

In the environment above, fixed costs of rent seeking confer to wealthier agents a comparative advantage in rent seeking. It is not surprising, then, that wealthier agents substitute away from production in a rent-seeking equilibrium.

But wealth is not the sole determinant of production possibilities in general. In particular, an individual's productivity as a producer could depend on innate

abilities that are unrelated to income or wealth. In this section, we consider an extension in which individuals differ in two ways: (1) their initial endowment (wealth) and (2) their ability as producers.[10]

Specifically, suppose an individual of ability a is able to produce $af(i)$ units of the final consumption good from an investment i. We posit that ability is uncorrelated with wealth. As before, the initial endowment is distributed according to the cumulative distribution function $\Gamma(w)$ on the $[\underline{w}, \overline{w}]$. An individual's ability is drawn *independently* from a cumulative distribution function $G(a)$ defined over $[\underline{a}, \overline{a}]$. The absence of correlation between wealth and ability introduces an interesting trade-off in occupational choices. Wealth will no longer be the sole determinant of occupational choice, as wealthier individuals may have a comparative advantage in production.

Consider the optimization problem facing a producer of ability a and wealth w. He chooses investment to maximize expected lifetime utility

$$U^P = w - i + \left(1 - \gamma\pi^P\right) af(i)$$

facing similar budget constraints as before. The first-order condition for investment choice, $1 = (1 - \gamma\pi^P)af'(i)$, gives the investment function

$$i^* = i\left(a, \gamma\pi^P\right), \tag{15}$$

which is increasing in the first argument as expected and decreasing in the second as before. Substituting these into the lifetime utility function gives a producer's indirect utility as

$$V^P = w + \left[\left(1 - \gamma\pi^P\right) af\left(i^*(a)\right) - i^*(a)\right], \tag{16}$$

where the term inside the brackets is positive given optimal investment choice and concavity of the production function.

Nothing changes in the rent seeker's optimization problem. He maximizes expected lifetime utility as before, taking as given the rent R that he expects to get from a successful random match with a producer. The indirect utility function is then

$$V^R = x(w - \theta) + \pi^R R, \tag{17}$$

as before, though rent R will be different in equilibrium.

[10]One interpretation of the basic model is that differences in endowments are due to differences in first-period abilities as producers. More able agents are wealthier in period 1 as a result. Subsequent to first-period production, agents face an investment choice between a productive asset (which yields $f(i)$ from investment i) and a low-productive one (with a fixed cost θ paid up front but a low return x in period 2, plus, possibly, rent-seeking opportunities). The productivity of this investment does not depend on first-period abilities. In this scenario, more able (and wealthier) individuals will prefer the low-productivity asset, resulting in smaller net output. This interpretation of the model is similar to the result of Murphy, Shleifer, and Vishny (1993) and Acemoglu and Verdier (1998) that rent seeking is particularly costly because talented individuals spend time rent seeking instead of in more productive occupations.

Consider now the occupational choice facing an individual with ability a and wealth $w \geq \theta$. The *net* utility that the individual enjoys from being a rent seeker is

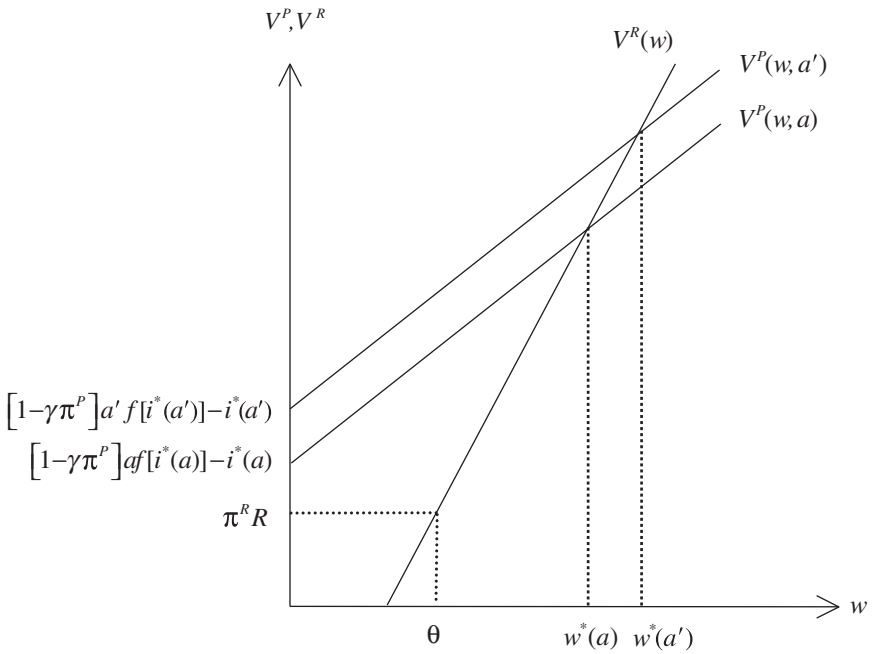
$$\Omega(w,a) = V^R - V^P = (x-1)w - x\theta + \pi^R R - \left[\left(1-\gamma\pi^P\right)af(i^*(a)) - i^*(a)\right], \tag{18}$$

which is increasing in wealth but decreasing in ability. Intuitively, the higher a person's productive ability, the more income that person can enjoy as a producer; while the higher the wealth, the easier it is to afford the cost of rent seeking.

But a high-wealth individual now finds rent seeking more attractive only if the individual's ability is low enough. To see this, consider Figure 4, which illustrates the indirect utility functions corresponding to ability levels a and $a'(> a)$. Since the higher-ability individual generates higher output, $V^P(w, a')$ lies uniformly above $V^P(w, a)$, the intercept gap capturing the differential productivity effect. As the diagram indicates, the lower-ability individual is willing to switch to rent seeking for a lower wealth level $w^*(a)$ than the higher-ability individual. Intuitively, wealth confers a comparative advantage in rent seeking, while ability confers a comparative advantage in production. Wealthy individuals of low ability, therefore, find rent seeking a more attractive occupation.

To formalize this intuition, note that an individual with wealth $w \geq \theta$ becomes a rent seeker only if $\Omega(w, a) \geq 0$, taking as given rents R and the number of agents in rent seeking n. Since Ω is a continuously differentiable increasing function of wealth and decreasing function of ability, we can rewrite this as $a \leq h(w)$ with $h' > 0$.

Figure 4. Occupational Choice with Differing Abilities ($a' > a$)



This means that the agent chooses to be a rent seeker only if the agent's ability as a producer is "low enough," that is, $a \leq h(w)$. This is true of any wealth level exceeding the minimum amount required to finance the rent-seeking operation. Thus, the measure of individuals corresponding to any wealth $w \geq \theta$ who become rent seekers is $\int_{\underline{a}}^{h(w)} dG(a)$.

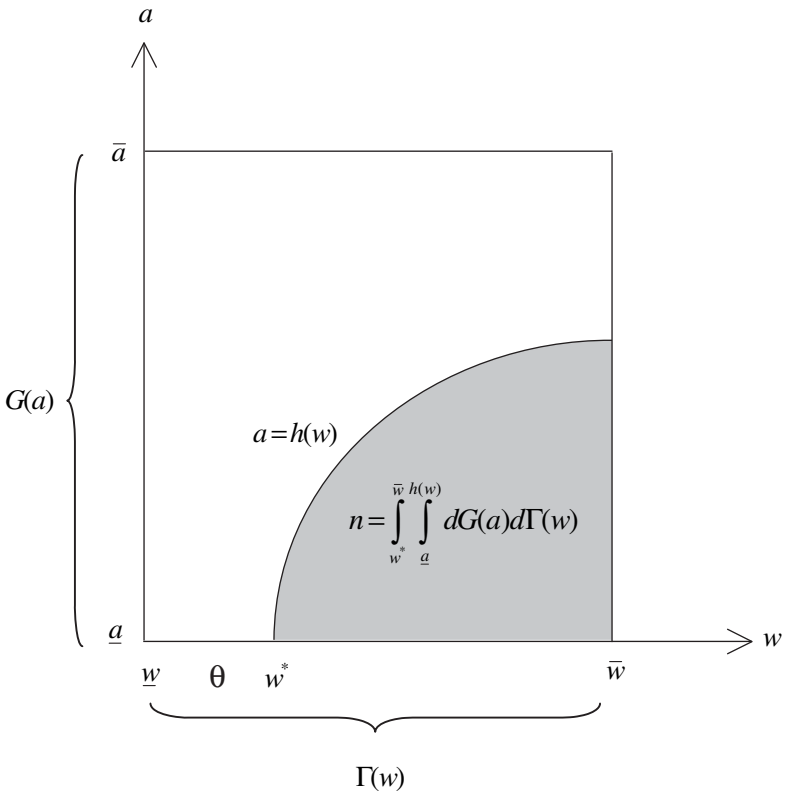
Now consider the lowest ability individual, $\underline{a}$, who has the most incentive to become a rent seeker as long as it is affordable. To do so, wealth has to be high enough relative to ability so that it adequately compensates for the fixed costs of rent seeking, that is,

$$w \geq h^{-1}(\underline{a}) \equiv w^*. \tag{19}$$

Hence, as Figure 5 illustrates, the measure of individuals in the entire population who choose rent seeking over production is simply

$$n = \int_{w^*}^{\bar{w}} \int_{\underline{a}}^{h(w)} dG(a)\, d\Gamma(w). \tag{20}$$

Figure 5. Fraction of Population That Becomes Rent Seekers When Abilities Differ



Finally, consider the rents that a rent seeker can expect to appropriate when successfully (but randomly) matched with a producer. Given the occupational choices above, this is given by

$$R = \frac{\lambda \int_{\underline{w}}^{w^*} \int_{h(w)}^{\bar{a}} \gamma f(i^*(a)) dG(a) d\Gamma(w)}{\int_{\underline{w}}^{w^*} \int_{h(w)}^{\bar{a}} dG(a) d\Gamma(w)}. \tag{21}$$

Since the expression for expected rents does not simplify any further, it is not easy solving for the general equilibrium of this model, even with linear utility and simple distribution functions. The intuition behind an interior rent-seeking equilibrium presented in the previous sections is, however, general. Given the differentiability and continuity of the utility and distribution functions, such an equilibrium will exist in this model. It will be defined by a threshold value w^*, the ability function $h(w)$, number of rent seekers n, and expected rents R that satisfy equations (15), (19), (20), and (21).

Corner equilibrium ($n = 0$, $R = 0$) with no rent seeking is also possible. In such an equilibrium, the net utility for an agent who is considering deviating to rent seeking is

$$\Omega(a, w; n = 0) = (x - 1) w - x\theta + \lambda\gamma \int_{\underline{a}}^{\bar{a}} f(i(a,0)) dG(a) - [af(i(a,0)) - i(a,0)],$$

where $i(a, 0)$ denotes optimal investment corresponding to ability a and zero rent seeking. We already know that the individual who faces the highest incentive to be a rent seeker is the one with highest wealth $\bar{w}$ and lowest ability $\underline{a}$. Hence, for no one willing to switch to rent seeking, a sufficient condition is that

$$\Omega(\underline{a}, \bar{w}; n = 0) = (x - 1) \bar{w} - x\theta + \lambda\gamma \int_{\underline{a}}^{\bar{a}} f(i(a,0)) dG(a) - \left[\underline{a} f(i(\underline{a},0)) - i(\underline{a},0)\right] \leq 0$$

or, equivalently,

$$\bar{w} \leq \frac{\left[\underline{a} f(i(\underline{a},0)) - i(\underline{a},0)\right] + x\theta - \lambda\gamma \int_{\underline{a}}^{\bar{a}} f(i(a,0)) dG(a)}{x - 1} \equiv w^c.$$

Given $\bar{w}$, this is more likely the higher the productivity of the lowest ability type ($\underline{a}$), the lower the return from hoarding (x), the higher the costs of becoming a rent seeker (θ), and the lower the productivity of rent seeking (γ, λ). As long as the condition above is satisfied, the economy will admit multiple levels of rent seeking as equilibrium outcomes. Typically, two stable equilibria with zero and positive rent seeking will result, as in Figure 3.

Focusing on the interior equilibrium, one implication of this extended model is obvious: The initial distribution is not the only determinant of rent seeking; it

depends also on the productivity of agents and the extent to which wealth and productivity are correlated.

Second, the model implies that costs of rent seeking in the previous model are exaggerated, since it "overstates" the incentives wealthier agents face to become rent seekers. At one extreme, if wealth and ability were perfectly correlated (for instance, because wealthier individuals could afford better education in period 1, which made them more productive in period 2), it is quite possible for rent seeking to disappear, in equilibrium; for example, if the education technology were highly productive.

In general though, innate ability (uncorrelated with wealth) also determines an individual's production possibilities. In the case analyzed above, positive rent seeking occurs in equilibrium. But the social costs of rent seeking are lower than before: First, a smaller fraction of the population with wealth exceeding θ become rent seekers, so the deadweight loss is lower, and second, these low-ability rent seekers do not cost aggregate output as much, since their comparative advantage does not lie in production.

The dynamic implications of our model are interesting. Think of the two-period model as a snapshot of a two-period, overlapping-generations economy in which successive generations are interlinked via wealth transfers (bequests). Assuming that wealthier parents transfer more to their offspring, our basic model with homogeneous abilities implies that persistent dynastic occupational choices are possible, as children of rent-seeking wealthy parents choose their parents' occupations.

If we allow for ability differences, too, and if intergenerational abilities are not (or weakly) correlated, that is no longer the case. Over time, we will observe dynastic mobility between the two occupations: As some producer dynasties become wealthier as a result of high abilities, there will be a tendency for their low-ability wealthy offspring to engage in rent seeking; and vice versa for dynasties that begin as rent seekers.

Finally, consider whether government policies can mitigate the effect of inequality on rent seeking via redistribution (assuming the government can observe endowments even though private agents cannot). Our model, whether ability is homogeneous or heterogeneous, implies that taxing the rich will raise their average cost for rent seeking, shifting them into production. But redistributing these tax revenues to the poor may be counterproductive if posttax endowments exceed θ, since it raises the incentive of the poor to engage in rent seeking. A better policy would be to use the tax revenues to subsidize production directly, which raises the return on investment for all.

Moreover, if the economy admits multiple rent-seeking equilibria (Figure 3), a *temporary* redistributive policy may have permanent effects if it is progressive enough. Taxing wealthier agents at a higher rate can significantly lower their net returns from rent seeking, so that rent seeking disappears in equilibrium. Since multiple equilibria result in this model from a coordination problem,[11] once the

[11]If all agents could coordinate a simultaneous move to production, the threat of expropriation would disappear, since there would be no rent seekers, and production would yield higher utility even to the wealthiest individuals.

economy moves away from rent seeking, lifting the redistributive tax schedule need not push the economy back to rent seeking as long as agents expect zero rent seeking in the future.

IV. Conclusion

This paper presents a simple economic model of rent seeking to examine the relationships among income distribution, rent-seeking behavior, and economic performance when the government cannot enforce property rights. We show that in the absence of credit markets only wealthy agents can overcome the nonconvexity in rent seeking. The wealthy are, therefore, "born into rent seeking." We analyze a model in which rent seeking not only enables agents to extract some portion of the proceeds from market production but also ensures protection of their wealth from appropriation by others. Therefore, wealthy agents avoid the tax from rent seekers by becoming rent seekers themselves, a result that accords well with both historical evidence and casual observation in developing countries.

The model implies that, in the absence of property rights protection, societies that have a more unequal distribution of wealth and are characterized by a small fraction of people who can afford entry into rent seeking will be the societies with greater social polarization and entrenched rent seeking by a few at the expense of the majority. At the other extreme, the model suggests that societies in which property rights are better enforced (for example, if γ is very low) will experience less polarization and higher economic performance.

We extended the model to analyze the robustness of our results when agents also differ in their ability as producers. Heterogeneous ability and wealth introduce a trade-off—agents specialize in rent seeking or production depending on where their comparative advantage lies. Since wealthier, but able, agents can prefer production over rent seeking, the rent-seeking equilibrium now involves lower social costs than when individuals have identical abilities.

Our model also has implications for redistribution and public education policies. Irrespective of whether ability is homogeneous or heterogeneous, our model implies that taxing the rich will raise their average cost for rent seeking, shifting them into production. However, simply redistributing these tax revenues to the poor may be counterproductive if it raises their posttax endowments and their incentive to engage in rent seeking. A better policy would be to use the tax revenues to target education resources to the poor, particularly if wealth and abilities are uncorrelated. If, on the other hand, wealth and abilities are perfectly correlated, it may be possible for the rent-seeking equilibrium to disappear if, for example, the education technology is highly productive. One policy conclusion that naturally arises from this case is that it may be worth targeting higher education to the rich to eliminate rent seeking. However, Dabla-Norris and Gradstein (2004) show that in societies with weak rule of law, the rich are more effective in appropriating a larger share of public education spending, thereby preventing the reduction of inequality. Therefore, overall incentives to engage in rent seeking may not be eliminated as long as the rich can also engage in rent seeking over public education funds. More generally, our model suggests that

improving access to high-quality education would raise the returns to production directly for all.

The model presented is essentially static in nature in that it describes the short-run equilibrium effect of the distribution of income on the decision to enter into rent-seeking activities. An important extension would be to consider the dynamics of the relationship between income distribution and rent seeking as an economy develops. Such an extension would allow us to analyze how the distribution of wealth can, in turn, be determined by the size of the rent-seeking sector and to explain the persistence of rent seeking and inequality in societies.

Another extension we plan to pursue in future work is the possibility for producers to invest in defending their output from appropriation. Such investment could be entirely private or, more interestingly, publicly funded out of taxes. This will provide a backdrop to study the endogenous evolution of property rights and how it changes with economic development and wealth accumulation. To do so, we need to move beyond our static framework with risk-neutral agents so that dynamic choices depend more meaningfully on the evolution of wealth and production possibilities.

APPENDIX

Proof of Proposition 1

Rewrite equation (11) as $w = \Gamma^{-1}(1 - n) \equiv \Phi(n)$, where Φ is monotonically decreasing. Similarly, from equations (13) and (14),

$$w = \Psi(n) \equiv \begin{cases} \frac{1}{x-1}\left[x\theta + \left(1 - \frac{\gamma\lambda}{1-n}\right) f(i^*) - i^*\right] & \text{for } n \le \tilde{n} \\ \frac{1}{x-1}\left[x\theta + \left(1 - \frac{\gamma}{n}\right) f(i^*) - i^*\right] & \text{for } n > \tilde{n} \end{cases},$$

which is nonmonotonic and, in particular, U-shaped, as shown in the text. The optimal investment level i^* solves $\left(1 - \frac{\gamma\lambda n}{1-n}\right) f'(i) = 1$ for $n \le \tilde{n}$, and $(1 - \gamma)f'(i) = 1$ otherwise. Equilibrium level of rent seeking n^* solves $Z(n^*) \equiv \Phi(n^*) - \Psi(n^*) = 0$. Once n^* is known, the equilibrium w^* can be determined via Φ or Ψ.

First, consider the case when $w^C < \overline{w}$. Here $\Phi(0) = \overline{w}$, $\Phi(1) = \underline{w}$, $\Psi(0) = w^C$, $\Psi(1) > \underline{w}$. Hence, $Z(0) > 0$ and $Z(1) < 0$. By the Intermediate Value Theorem (IVT), since $Z(n)$ is continuous on [0, 1] there exists at least one $n^* \in [0, 1]$ such that $Z(n^*) = 0$. In fact, by appealing to the U shape of Ψ, we know this equilibrium is unique.

When $w^C \ge \overline{w}$, on the other hand, let $n^C \ge 0$ such that $\Psi(n^C) = \overline{w}$. Now we have $Z(0) = 0$, $Z(n^C) < 0$, $Z(1) < 0$. Note that, as before, there exists $n^* = 0$ such that $Z(n^*) = 0$. Zero rent seeking is evidently an equilibrium in this case. In addition, other equilibria with positive rent seeking can also exist, since Z is nonmonotonic. For instance, since Ψ is U-shaped, it is possible that $\Psi(\tilde{n}) < \Phi(\tilde{n})$, in which case $Z(\tilde{n}) > 0$. By the IVT, we should have an intermediate $n' \in (0, \tilde{n})$ such that $Z(n') = 0$ and another $n^{**} \in (\tilde{n}, 1)$ such that $Z(n^{**}) = 0$. But only n^{**} is an equilibrium in the model, since n' is unstable (slight deviations in n will cause individuals to move toward zero or n^{**}). In other words, a sufficient condition for the existence of a unique interior rent-seeking equilibrium is $w^C < \overline{w}$.

Proof of Proposition 2

The differential equilibrium utility of any agent with wealth $w \geq w^*$, given equilibrium level of rent seeking n^*, is

$$\Omega(w;n^*) = \begin{cases} w(x-1) - x\theta + i^* - \left(1 - \frac{\gamma\lambda}{1-n^*}\right) f(i^*), & \text{for } n \leq n^* \\ w(x-1) - x\theta + i^* - \left(1 - \frac{\gamma}{n^*}\right) f(i^*), & \text{for } n > n^* \end{cases}.$$

Evidently, $\partial\Omega/\partial w = x - 1 > 0$. In other words, as long as individuals can afford the fixed cost of rent seeking ($w \geq w^*$), net utility from rent seeking over production increases in initial wealth w.

REFERENCES

Acemoglu, D., 1995, "Reward Structures and the Allocation of Talent," *European Economic Review,* Vol. 39 (January), pp. 17–33.

———, and T. Verdier, 1998, "Property Rights, Corruption and the Allocation of Talent: A General Equilibrium Approach," *The Economic Journal,* Vol. 108 (September), pp. 1381–403.

Baland, J., and P. Francois, 2000, "Rent-Seeking and Resource Booms," *Journal of Development Economics,* Vol. 61 (April), pp. 527–42.

Banerjee, A. V., and A. F. Newman, 1993, "Occupational Choice and the Process of Development," *Journal of Political Economy,* Vol. 101 (April), pp. 274–98.

Baumol, W. J., 1990, "Entrepreneurship: Productive, Unproductive, and Destructive," *Journal of Political Economy,* Vol. 98 (October), pp. 893–921.

Benabou, R., 1997, "Inequality and Growth," NBER Working Paper No. 5658 (Cambridge, Massachusetts: National Bureau of Economic Research).

Braudel, F., 1982, *Civilization and Capitalism 15th–18th Century* (New York: Harper and Row).

Dabla-Norris, E., and M. Gradstein, 2004, "The Distributional Bias of Public Education: Causes and Consequences," IMF Working Paper 04/214 (Washington: International Monetary Fund).

de Ferranti, David, and others, 2003, *Inequality in Latin America and the Caribbean: Breaking with History?* (Washington: World Bank).

Downs, A., 1957, *An Economic Theory of Democracy* (New York: Harper).

Easterly, W., 2002, "Inequality Does Cause Underdevelopment: New Evidence from Commodity Endowments, Middle Class Share, and Other Determinants of Per Capita Income," Center for Global Development Working Paper No. 1 (Washington: Center for Global Development).

Engerman, S., and K. Sokoloff, 2002, "Factor Endowments, Inequality, and Paths of Development Among New World Economies," NBER Working Paper No. 9259 (Cambridge, Massachusetts: National Bureau of Economic Research).

Galor, O., and J. Zeira, 1993, "Income Distribution and Macroeconomics," *Review of Economic Studies,* Vol. 60 (1), pp. 35–52.

Grossman, H., 1991, "A General Equilibrium Model of Insurrections," *American Economic Review,* Vol. 81 (September), pp. 912–21.

———, 1994, "Production, Appropriation, and Land Reform," *American Economic Review,* Vol. 84 (June), pp. 705–12.

———, and M. Kim, 1997, "Predation, Efficiency and Inequality," NBER Working Paper No. 6301 (Cambridge, Massachusetts: National Bureau of Economic Research).

Hall, R., and C. Jones, 1999, "Why Do Some Countries Produce So Much More Output Per Worker than Others?" *Quarterly Journal of Economics,* Vol. 114 (February), pp. 83–116.

Hoff, K., and J. Stiglitz, 2004, "After the Big Bang? Obstacles to the Emergence of the Rule of Law in Post-Communist Societies," *American Economic Review,* Vol. 94 (June).

Huntington, S., 1968, *Political Order in Changing Societies* (New Haven, Connecticut: Yale University Press).

Keefer, P., and S. Knack, 1997, "Why Don't Poor Countries Catch Up? A Cross-National Test of an Institutional Explanation," *Economic Inquiry,* Vol. 35 (July), pp. 590–602.

———, 2002, "Polarization, Politics, and Property Rights: Links Between Inequality and Growth," *Public Choice,* Vol. 111 (October), pp. 127–54.

Krueger, A. O., 1974, "The Political Economy of a Rent Seeking Society," *American Economic Review,* Vol. 64 (June), pp. 291–303.

Levi, M., 1989, *Of Rule and Revenue,* California Series on Social Choice and Political Economy 13 (Berkeley: University of California Press).

Mauro, P., 1995, "Corruption and Growth," *Quarterly Journal of Economics,* Vol. 110 (August), pp. 681–712.

Mehlum, H., K. Moene, and R. Torvik, 2003, "Predator or Prey?: Parasitic Enterprises in Economic Development," *European Economic Review,* Vol. 47 (April), pp. 275–94.

Murphy, K., A. Shleifer, and R. Vishny, 1993, "Why Is Rent-Seeking So Costly to Growth?" *American Economic Review,* Vol. 83 (May), pp. 409–14.

Olson, M., 1965, *The Logic of Collective Action; Public Goods and the Theory of Groups* (Cambridge, Massachusetts: Harvard University Press).

Rodriguez, F., 1999, "Inequality, Redistribution and Rent Seeking," Department of Economics Working Paper, University of Maryland (College Park: University of Maryland).

Skaperdas, S., 1992, "Cooperation, Conflict, and Power in the Absence of Property Rights," *American Economic Review,* Vol. 82 (September), pp. 720–39.

Sonin, K., 2003, "Why the Rich May Favor Poor Protection of Property Rights," *Journal of Comparative Economics,* Vol. 31 (December), pp. 715–31.

Verdier, T., 1996, "The Rise and Fall of Elites: A Theory of Economic Development and Social Polarization in Rent-Seeking Societies," CEPR Discussion Paper No. 1495 (London: Center for Economic Policy Research).

Wade, R., 1985, "Market for Public Office: Why the Indian State Is Not Better at Development," *World Development,* Vol. 13 (April), pp. 467–97.

IMF Staff Papers
Vol. 53, No. 1

Exchange Rate Regimes, Location, and Specialization

LUCA ANTONIO RICCI*

This paper investigates the effects of fixed versus flexible exchange rates on firms' location choices and on countries' specialization patterns. In a two-country, two-differentiated-goods monetary model, uncertainty arises after wages are set and prices are optimally chosen. The paper shows that countries are more specialized under flexible than fixed rates, which indicates that the pattern of specialization is not uniquely defined by trade models but also depends on the exchange rate regime. The creation of a currency area endogenously increases the desirability of such an area by reducing the asymmetry of shocks across member countries. The results also shed light on the effects of exchange rate variability on trade. [JEL F1, F31, F33, F4, L16, R12]

In the presence of price rigidities, countries tend to be more specialized under flexible exchange rates than under fixed exchange rates, thus suggesting that the pattern of specialization indicated by any trade model is not unique but also depends

*Previous drafts of this paper were written when Luca Antonio Ricci was a Visiting Fellow at Harvard University and at the University of Konstanz. He is grateful to Joshua Aizenman, Richard Baldwin, Olivier Blanchard, Alessandra Casella, Franceso Caselli, Tito Cordella, Donald Davis, Avinash Dixit, Guy Dumais, Hans Genberg, José Gonzales, Paul Krugman, Philip Lane, Giovanni Maggi, Paul Masson, Jeffrey Sachs, Carmelo Salleo, Jaume Ventura, and two anonymous referees for helpful discussions and comments. He thanks all the participants in conferences or workshops at the Center for Economic and Policy Research, European Meeting of the Econometric Society, Graduate Institute of International Studies, Harvard University, Innocenzo Gasparini Institute for Economic Research, La Sapienza University, University of Konstanz, University of Paris-I Panthéon-Sorbonne, and University of Louvain. He is indebted to Edward T. Smith for invaluable help, and to Sean Culhane for editorial assistance. Generous financial support from the Swiss National Research Funds and from the Human Capital Mobility program is gratefully acknowledged.

on the exchange rate regime. An important implication follows: the net benefits that can be expected from the creation of a currency area are endogenous to—and rise with—the creation of the currency area, as the latter induces sectoral dispersion and consequently reduces the degree of asymmetry of shocks across candidate countries. A second implication is that an increase in exchange rate variability would have an ambiguous effect on trade by increasing interindustry trade and reducing intra-industry trade. The implications highlight some possible real effects arising from the creation of the European Monetary Union (EMU).

Both the main results and the implications of this paper adequately capture some facts highlighted in the literature. The positive effect of flexible exchange rates on trade specialization is consistent with Fontagné and Freudenberg (1999), who find that exchange rate variability increases interindustry trade and reduces intra-industry trade (thus raising specialization) among European Union (EU) countries. The predicted positive impact of fixed exchange rates on the symmetry of shock is consistent with the results of Fatás (1997), who shows that the business cycle correlation of EU countries increased after the introduction of the European Monetary System. The endogeneity of the desirability of a currency union is consistent with Frankel and Rose (1998), who find that industrial countries that trade more with each other have more correlated business cycles; hence, if the EMU enhances trade among its member countries, shocks should become more symmetric.[1] The ambiguous effect of exchange rate variability on trade is highlighted in the empirical results of Clark and others (2004).

To illustrate, consider a world composed of two countries (1 and 2) engaging in both intra- and interindustry trade of two differentiated goods (*A* and *B*). Assume country 1 is a net exporter of good *A*. After prices are chosen, consider a shift of demand from good *B* to good *A*. Under fixed exchange rates, the shock equally affects all firms producing the same good, regardless of their location. Under floating, however, currency 1 appreciates, and the consequent substitution effect reduces the initial increase of demand experienced by firms producing *A* in country 1 and generates a further increase in the demand for varieties of *A* produced in country 2 (the opposite holds true for industry *B*.)

This implies that, on average, endogenous exchange rate movements provide a partial adjustment to shocks for firms located in the country that is a net exporter of the good they produce but generate further disturbance for firms producing the same good in the net importer country. As a consequence, under a flexible exchange rate regime, firms located in the country relatively specialized in (net exporter of) the good they produce experience a lower variability in sales than their competitors.

To the extent that firms dislike variability in sales, under flexible exchange rates the uneven sectoral adjustment to shocks gives firms an incentive to locate in the country that is relatively specialized in the good they produce. Under fixed exchange rates, however, all firms face the same variability in sales regardless of their location, so this incentive does not arise. Thus, countries should be more

[1]See, however, Imbs (1999), who finds that neither the degree of trade intensity nor a fixed exchange rate regime seems to matter for the synchronization of the business cycle.

specialized under flexible exchange rates than under fixed rates. This would suggest that nominal rigidities and shocks could induce an additional incentive for specialization, in addition to those suggested by trade and location theory.

The novelty of our results stems from the integration of basic elements of open macroeconomic, trade, and location theory (i.e., price rigidities, specialization patterns, and location choices), thus overcoming some limitations of each approach. Open macroeconomics usually neglects countries' specialization patterns.[2] Trade theory usually ignores the existence of short-run market rigidities and assigns no role to the nominal exchange rate. Location theory normally does not consider the effects of exchange rate regimes on the location choices of firms.[3] While several papers have analyzed the effects of exchange rate regimes on foreign direct investment (FDI), this literature (except for Aizenman, 1992) usually neglects the endogenous nature of the exchange rate and fails to capture the heterogeneity of sectors of production in the economy.[4] Hence, it does not allow us to infer how changes in FDI induced by exchange rate variability influence the pattern of specialization.

We develop a two-country, two-differentiated-good, one-factor monetary model in which countries engage in both inter- and intraindustry trade. Price rigidities, decreasing returns to scale, and international labor immobility allow us to tailor the model toward the representation of the short-run adjustment to shocks. The effectiveness of the exchange rate stems from the fact that countries do not have an identical production structure and that wages are set and prices are optimally chosen before the resolution of uncertainty. We ignore the existence of trade costs to neutralize the backward and forward linkages, and thus focus on the location incentives and specialization patterns induced by the sectoral impact of exchange rate adjustments.[5]

I. The Model

The model extends Blanchard and Kiyotaki (1987) to a two-country two-differentiated-good setup.[6] This is similar to a one-period version of the new open economy approach (see Obstfeld and Rogoff, 2000), but with two tradable goods.

[2]See, however, Faruqee (1996), which introduces nominal rigidities in a two-country model of imperfect competition and investigates the implications of different trade patterns for real exchange rate dynamics. Tille (2006) builds on Ricci (1997a) and Faruqee (1996) to investigate how the differential impact of exchange rate movements across sectors leads to substantial welfare differences across households in a given country.

[3]See, however, Ricci (1998) for agglomeration effects arising from exchange rate variability. For a review of the location literature, see Fujita, Krugman, and Venables (1999). For a survey of location theory in a historical perspective, see Fujita and Thisse (1996).

[4]Most of the investigations are based on a partial equilibrium analysis of the behavior of a single firm and find that when the variability of the exchange rate exogenously increases, a risk-averse firm will raise FDI (see Cushman, 1988; Goldberg and Kolstad, 1994; and Campa and Goldberg, 1995) or increase foreign production and decrease foreign sales (see Broll, Wahl, and Zilcha, 1995). In a two-country model, Aizenman (1992) finds that under fixed exchange rates, both domestic investment and foreign direct investment are higher than under flexible exchange rates.

[5]For a description of how these linkages shape economic geography, see Krugman (1991).

[6]We leave out some of the features of the Blanchard and Kiyotaki (1987) model. For our purposes, it is unnecessary to replicate their endogenous wage setting, to employ differentiated labor supply, or to introduce work in the utility function.

Consider a world composed of two countries (1 and 2) that are inhabited by an equal number of agents (L) and produce two types of goods (A and B). Each good is produced in n differentiated varieties (indexed by $i = 1, \ldots, n$ for good A and by $j = 1, \ldots, n$ for good B) worldwide, each variety being manufactured by a different firm. The only factor of production—labor—is homogenous, immobile across countries, and mobile across industries (this last assumption is not essential).

The two countries have a mirror-image production structure: a share η of the n firms in industry A are located in country 1, and an identical share η of the n firms in industry B are located in country 2. Without loss of generality, we assume $\frac{1}{2} < \eta < 1$, so that country 1 (country 2) is relatively specialized in good A (good B). A sector is defined as the part of an industry located in one country; hence, there are four sectors: $A1$, $A2$, $B1$, and $B2$. The parameter η may be thought of as capturing trade-theoretical reasons for a given specialization pattern.[7]

Uncertainty, Timing of Actions, Price Setting, and Monetary Rule

The formal analysis focuses on uncertainty arising from demand shocks. Monetary, productivity, and exchange rate shocks are discussed in the second subsection of Section II.

Before the resolution of uncertainty, workers of each country set the domestic wage (w_k, with $k = 1,2$) and commit to supply as much labor as demanded by firms at this wage level. Firms observe the wage and choose optimal prices as markup over expected marginal costs. Because the wage and the associated employment in the absence of shocks are not relevant to our analysis (what matters are the fluctuations around the initial equilibrium), we assume that in the absence of shocks the wage chosen would ensure full employment (i.e., every worker would supply one unit of labor).[8]

After the resolution of uncertainty, a new equilibrium in the goods and money markets is reached: taking wages and prices as given, consumers choose optimal consumption and money balances, and firms choose optimal employment levels. Monetary authorities do not pursue discretionary policies. When shocks occur in a flexible exchange rate regime, monetary authorities abstain from intervention and let the exchange rate adjust the money market and the trade balance. When shocks occur in a fixed exchange rate regime, authorities are committed to eliminate any pressure on the exchange rate by adjusting money supply in order to equilibrate the money market (trade may be unbalanced).[9]

[7]This mirror-image location of firms could be derived endogenously by introducing a comparative advantage or factor specificity. A previous draft of this paper allowed for a Ricardian comparative advantage in a new trade theory setup, as in Ricci (1997c, 1999), but the resulting mathematical complication obscured the intuition and required simulations to present qualitatively identical results.

[8]Menu costs or costs in the price-setting decision process may be the reason wages and prices are not adjusted once shocks are known. Nominal rigidities are quite common in recent international macroeconomics (see Obstfeld and Rogoff, 2000).

[9]It is irrelevant to our results whether the fixed exchange rate regime is managed symmetrically or asymmetrically. We solve the model for the symmetric case.

Consumer Maximization Problem

All individuals share the same utility function. A representative consumer of country k chooses nominal money balances (m_k) and consumption of varieties of goods A and B (c_{iAk} and c_{jBk}, respectively) so as to maximize the following random preferences:

$$U_k = \left(C_{Ak}^{\gamma} C_{Bk}^{1-\gamma}\right)^{\lambda} \left(m_k / P_k\right)^{1-\lambda}, \quad 0<\gamma<1,\ 0<\lambda<1,\ k=1,2, \tag{1}$$

with

$$C_{Ak} = \left(\sum_{i=1}^{n} c_{iAk}^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)}, \quad C_{Bk} = \left(\sum_{j=1}^{n} c_{jBk}^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)}, \quad \sigma>1,$$

where P_k is the true price index of consumption in country k, σ is the elasticity of substitution among varieties of the same good, and γ is the random share of expenditure on good A, whose mean value is 0.5 and whose percentage change ($\mu = d\gamma/\gamma$) is bounded in $(-\Omega, \Omega)$ with mean 0 and variance u^2.

The nominal wealth of a representative individual of country k (q_k) is the sum of income (y_k) and endowment of domestic currency (m_k^s). Each individual supplies labor to domestic firms at the given wage and receives profits from these firms; individual income is, therefore, a share $1/L$ of domestic firms' revenues. Individual endowment of money is a fraction $1/L$ of the domestic stock of money, which may vary under fixed exchange rates because of monetary intervention. The consumer's budget constraint is:

$$\sum_{i=1}^{n} p_{iA}^{k} c_{iAk} + \sum_{j=1}^{n} p_{jB}^{k} c_{jBk} + m_k^d = q_k \equiv m_k^s + y_k, \tag{2}$$

where p_{iA}^k and p_{jB}^k are the prices of variety i of good A and of variety j of good B, measured in the currency of the consumer's country k.

Aggregate demands of country k for money (M_k^d) and for each variety of good A produced in country f ($A_{fk}, f=1,2$) are given by (the expressions for good B can be derived analogously):

$$A_{11} = \frac{p_{A1}^{-\sigma}}{P_{A1}^{1-\sigma}} \gamma\lambda Q_1, \quad A_{21} = \frac{(e\,p_{A2})^{-\sigma}}{P_{A1}^{1-\sigma}} \gamma\lambda Q_1, \quad M_1^d = (1-\lambda) Q_1, \tag{3}$$

$$A_{12} = \frac{(p_{A1}/e)^{-\sigma}}{P_{A2}^{1-\sigma}} \gamma\lambda Q_2, \quad A_{22} = \frac{p_{A2}^{-\sigma}}{P_{A2}^{1-\sigma}} \gamma\lambda Q_2, \quad M_2^d = (1-\lambda) Q_2, \text{ with}$$

$$P_{A1} = P_{A1} = \left(n_{A1}\, p_{A1}^{1-\sigma} + n_{A2}\,(e p_{A2})^{1-\sigma}\right)^{1/(1-\sigma)},$$

$$P_{A2} = \left(n_{A1}\,(p_{A1}/e)^{1-\sigma} + n_{A2} p_{A2}^{1-\sigma}\right)^{1/(1-\sigma)} = P_{A1}/e,$$

$$Q_k = n_{Ak}\, p_{Ak}\, x_{Ak} + n_{Bk}\, p_{Bk}\, x_{Bk} + M_k^s = L\, q_k, \quad k=1,2,$$

where Q_k is the aggregate wealth of country k, P_{Ak} is true price index of good A in country k, e is the exchange rate defined as units of the currency of country 1 for one unit of the currency of country 2 (and is equal to 1 under a fixed exchange rate regime), and n_{ck} is the number of varieties of good c ($c = A, B$) produced in country k.

Firms' Maximization Problem

The market structure is the usual large group monopolistic competition based on Dixit and Stiglitz (1977) and often adopted in trade theory thereafter (see Helpman and Krugman, 1985). There is no free entry; endogenizing the number of firms would not alter the results.

Production functions are identical for all firms and exhibit diminishing returns to labor. The output (x_{ck}) of a typical firm producing a variety of good c ($c = A,B$) in location k ($k = 1,2$) is given by $x_{ck} = l_{ck}^{\alpha}$, where l_{ck} is the employment of such a firm. We assume for simplicity that $\alpha = 0.5$; this assumption is not essential as long as $0<\alpha<1$.

Before the resolution of uncertainty, a typical firm of sector ck takes the wage and other firms' behavior as given and chooses its price (p_{ck} in its domestic currency) to maximize expected profits (π_{ck})[10]:

$$E[\pi_{ck}] = E\left[p_{ck}\, x_{ck}^{d} - w_k\, l_{ck}\right] = E\left[p_{ck}\, x_{ck}^{d} - w_k \left(x_{ck}^{d}\right)^{1/\alpha}\right], \quad c = A,B;\ k = 1,2, \tag{4}$$

where E is the expectation operator and x_{ck}^{d} is the demand for one variety of good c produced in k:

$$x_{Ak}^{d} = A_{k1} + A_{k2}, \quad x_{Bk}^{d} = B_{k1} + B_{k2}.$$

The profit-maximizing price (p_{ck}) for good c produced by a typical firm of country k is therefore set as a markup over expected marginal cost:

$$P_{ck} = \frac{\sigma}{\sigma - 1} E\left[\frac{w_k}{\alpha}\left(x_{ck}^{d}\right)^{(1-\alpha)/\alpha}\right] = \frac{\sigma}{\sigma - 1} E\left[\frac{w_k}{\alpha}\left(l_{ck}\right)^{1-\alpha}\right], \quad c = A,B;\ k = 1,2, \tag{5}$$

where σ approximates, for n large enough, the perceived elasticity of demand.

After the resolution of uncertainty, firms choose optimal employment. As prices and wages are now given, the profit function is rising in output (around the initial equilibrium) and firms will find it optimal to satisfy demand, thus bringing the goods market into equilibrium ($x_{ck}^{d} = x_{ck}$).[11] The profit function is also concave

[10]Producer currency pricing is the prevailing assumption in recent open economy literature. Alternatively, some authors have focused on the pricing-to-market assumption (for example, Faruqee, 1995; and Betts and Devereux, 2000). See Obstfeld and Rogoff (2000) for a summary of the arguments against the pricing-to-market assumption.

[11]Firms would not find it optimal to satisfy very large increases in demand, since marginal cost would rise above price. We rule out this possibility by assuming that the shocks are opportunely bounded, so that the qualitative outcome of a comparative statics exercise would correspond to the outcome of a simulation exploiting the full nonlinearities of the model.

in output, implying that firms dislike variability in sales. On any given variety, the law of one price applies.

Equilibrium in the Absence of Shocks

In the absence of shocks, the only difference between the two countries is given by their symmetric pattern of specialization (η). Wages are set at the same level (w) in both countries. Because all firms face an identical wage and a marginal cost function, which is linear in output, they all choose the same price, which can be normalized to 1 ($p_{ck} = p = 1$, with $c = A,B$; $k = 1,2$). Prices being equal, an identical share of expenditure will be allocated to each variety. All firms have identical employment, output, and profits. Both countries have the same aggregate income and wealth, and every consumer demands all varieties in the same amount. After normalizing the exchange rate and the price of each variety to 1, we obtain the following (where M^{WR} is the equilibrium world real stock of money):

$$p = e = 1, \quad x = l^{1/2} = \left(\frac{L}{n}\right)^{1/2} = \frac{\lambda}{1-\lambda}\frac{M^{WR}}{2n}, w = \frac{\sigma-1}{\sigma}\frac{1}{2x}, \quad p = \frac{\sigma+1}{2\sigma}x. \tag{6}$$

II. Uncertainty and Equilibrium Location

In this section, we investigate the equilibrium distribution of firms that occurs if firms are allowed to choose location before any other action takes place. To this purpose, for each exchange rate regime, we first study how shocks affect firms' sales and expected profits in each sector for a given location structure (η). We then compare firms' expected profits within the same industry across locations and discuss the equilibrium location structure. The exercise is based on comparative statics calculations.

Equilibrium Location Pattern

To discuss the effect of demand shocks under a fixed exchange rate regime, we assume a given location distribution (η). Given the assumptions outlined in Section I, before the occurrence of shocks, the economy is described by the set of relations discussed in the last subsection of Section I. Under fixed exchange rates, any demand shock (μ) gives rise only to a direct demand effect deriving from the change in preferences. While the money market is equilibrated by the intervention of monetary authorities, the new goods market equilibrium is:

$$X_{Ak} = \mu, \quad X_{Bk} = -\mu, \tag{7}$$

where $X_{ck} = dx_{ck}/x_{ck}$. Considering the distribution of shocks, ex ante all firms face the same variability of output and the same expected profits independently of their location:

$$E[\pi_{ck}] = x - wE[x(1+X_{ck})]^2 = \pi - w\,x^2\,u^2, \tag{8}$$

with x and π defined in the final subsection of Section I. Hence, firms are indifferent to location. We can thus state the following:

Proposition 1. Under a fixed exchange rate regime, any distribution of firms is an equilibrium location pattern.

More generally, if η were to be endogenously determined, say, via factors such as comparative advantage or location theory, a fixed exchange rate regime would not induce a deviation from such an equilibrium allocation of resources.

Under flexible rates, the exchange rate would react to balance trade. Following any demand shock, the direct demand effect is accompanied by the substitution effect owing to the exchange rate adjustment, which affects firms differently according to their industry and location. For example, an increase in demand for good *A* would tend to improve the trade balance of country 1 (which is specialized in *A*) and to deteriorate the balance of country 2, thus prompting an appreciation of the currency of country 1. This would, in turn, generate a substitution effect away from all goods produced by country 1 and in favor of those produced by country 2. Such a substitution effect would dampen the initial shock for firms producing *A* in country 1 (or *B* in country 2), but it would exacerbate the shock for firms producing *B* in country 1 (or *A* in country 2). More generally, the possibility of shocks would induce firms to anticipate a lower variability in sales in the location where they would be part of the net exporting sector. Since firms dislike variability in sales, they would prefer such a location. Formally, the goods and money market equilibria require

$$de/e = -2(2\eta - 1)z\mu, \quad X_{ck} = g_{ckD}\mu, \tag{9}$$

with

$$z = 1/(1 + 4\eta(1-\eta)(\sigma - 1)), \quad 0 < g_{A1D} = -g_{B2D} = 2(1-\eta)\sigma z < 1,$$
$$g_{A2D} = -g_{B1D} = 2\eta\sigma z > 1,$$

where the change in the exchange rate ensures trade balance, which reflects changes in sales in the four different sectors. By confronting the coefficients g_{ckD}, we can derive that firms located in the country relatively specialized in their industry (i.e., firms producing varieties of *A* in country 1 or of *B* in country 2) face a lower variability in sales and higher expected profits than firms of the same industry located in the other country (i.e., producing *A* in country 2 or *B* in country 1):

$$E[\pi_{ck}] = x - wE[x(1 + X_{ck})]^2 = \pi - wx^2 g_{ckD}^2 u^2,$$
$$E[\pi_{A1}] - E[\pi_{A2}] = E[\pi_{B2}] - E[\pi_{B1}] = wx^2 4(2\eta - 1)\sigma^2 z^2 u^2 > 0. \tag{10}$$

The difference in expected profits across locations for the same industry increases with the degree of specialization (η) and with the variance of the shocks

(u^2). Because firms enjoy higher expected profits if they are located in the country relatively specialized in the good they produce, countries are fully specialized in equilibrium. We can thus state the following:

Proposition 2. Under a flexible exchange rate regime, firms have an incentive to locate in the country that is relatively specialized in the good they produce. In the absence of any other location incentive (such as comparative advantage), the equilibrium location pattern is given by full specialization.

The extreme specialization under flexible rates is obviously due to the exogenous nature of the initial distribution of firms (η).[12] More generally, when other trade or location incentives for specialization are present, the equilibrium location pattern would be derived by weighing, for the marginal firm: (1) the incentive to locate in the country that is relatively specialized in the good the firm produces, in order to benefit from the adjustment role of the exchange rate; and (2) the efficiency loss associated with the departure from the location choice dictated by the comparative advantage or other location incentives. As a result, countries would be more specialized under flexible than under fixed exchange rates.

Other Sources of Uncertainty

In this section we argue that similar location incentives arise from monetary, exchange rate, and productivity shocks.[13]

Monetary and exchange rate shocks

Under a fixed exchange rate regime, any change in money stock is reflected in an equal change in expenditure on both goods, independently of the country in which the shock originated. All firms experience the same change in sales; hence, the same variability in sales and the same expected profits. Any distribution of firms is an equilibrium.

Under a flexible exchange rate regime, the exchange rate movements associated with monetary and exchange rate shocks induce expenditure shifts across countries. Each firm located in the country that has the largest market share of the good that the firm produces will bear a smaller share of the expenditure shift than its foreign competitors. Such a firm would therefore face a lower variability in sales and higher expected profits. To the extent that countries are net exporters of the good of which they have the largest market share, under flexible rates, firms would have an incentive to locate in the country relatively specialized in the

[12]The results would apply also in the presence of endogenous monetary policy, as long as such a policy would not find it optimal to fully counteract shocks.

[13]For a formal analysis of these shocks, see the Working Paper version of this paper (Ricci, 1997a). In all cases but one (exchange rate shocks), the exchange rate is an endogenous variable.

good they produce, and the only equilibrium location pattern would be full specialization.[14]

Supply shocks

Supply shocks are very similar to demand shocks, once we allow for an automatic response of monetary policy. In the presence of price rigidities, it is reasonable to assume that monetary authorities adjust domestic money supply in the same direction as the change in average domestic productivity, to accommodate changes in expenditure and reduce employment fluctuations. For simplicity, consider the case in which money is adjusted at the same rate as productivity.

Under a fixed exchange rate regime, the monetary accommodation allows expenditure on all goods to change by the world average productivity growth. All firms face the same variability in sales, and any location pattern is an equilibrium.

Under a flexible exchange rate regime, the country whose average productivity rises relative to that of the other country experiences a depreciation of its currency. The consequent substitution effect eases the adjustment of firms located in the country relatively specialized in the good they produce and constitutes an element of further disturbance for the other firms. Firms located in the country relatively specialized in their industry experience a lower variability in sales and higher expected profits. The equilibrium location of firms is full specialization.

Also for monetary, exchange rate, and productivity shocks, the presence of comparative advantage or other location incentives would imply that specialization under flexible rates would not be full but would be higher than under fixed rates.

III. Conclusions

This paper shows that fixed and flexible exchange rate regimes are associated with different location incentives when demand, supply, monetary, and exchange rate shocks arise in the presence of short-run price rigidities. Countries tend be more specialized under flexible exchange rates than under fixed exchange rates. In fact, when real shocks occur, flexible exchange rates provide a partial adjustment to the firms located in the country whose aggregate shocks to net exports are positively correlated with the firm's shocks and generate further disturbance to the other firms. This effect does not occur under fixed rates. A similar intuition would apply to the cases of monetary and exchange rate shocks.

Our findings imply that the pattern of specialization suggested by any trade model is not unique but depends on the exchange rate regime. If one considers that firms' location choices are normally influenced by factors such as comparative advantage, economic geography, and so on, our model would suggest that under fixed exchange rates the equilibrium location distribution would be dictated by only those factors. Under flexible exchange rates, however, additional location incentives would arise: the equilibrium location pattern would need to be

[14]In a setup with nonsymmetric countries, there would also be agglomeration effects (see Ricci, 1998).

derived by weighing, for the marginal firm, the incentive to locate in the country that is relatively specialized in the good the firm produces and the efficiency loss associated with the departure from the location dictated by trade and geography factors.

This paper provides a theoretical argument for the endogeneity of the optimum currency area (OCA) criterion; that is, the net benefits that can be expected from a currency area increase with the creation of the currency area, because this reduces international specialization and the asymmetry of shocks across candidate countries.[15] This would suggest that the creation of the European Monetary Union could induce a relocation of economic activity that would make countries more similar and shocks more symmetric.

This paper also offers two explanations for the puzzle related to the effects of exchange rate variability on trade. Whereas most theoretical analyses (based on partial equilibrium models) suggest a negative effect, most empirical studies found no effect or a small negative one (see, for example, Gagnon, 1993; Dell'Ariccia, 1999; and Clark and others, 2004). First, a partial equilibrium analysis may overstate the negative effect of the exchange rate variability, as it fails to recognize that such variability is partly an endogenous response to shocks. Second, under flexible rates, economic uncertainty (whether in fundamentals or in the exchange rate itself) unevenly affects firms in different industries, suggesting that a small impact at the aggregate level may hide large differences at the industrial level.

The model has been kept simple to avoid unnecessary mathematical complication and to highlight the intuition; however, we believe that our results would hold qualitatively under more general assumptions or in a more complicated setup. For example, one could extend the setup to a multiperiod framework (as long as in every period some shocks arise after prices are optimally chosen); introduce financial assets (as long as contingent claims markets are incomplete); consider that firms maximize expected utility from profits (as long as international equity markets are imperfect);[16] relax price rigidities (as long as firms dislike variability in sales, the result would hold for demand and monetary shocks);[17] allow for foreign direct investment (the location incentives we described would apply to both the main firm and its foreign subsidiaries); endogenously derive the number of firms in each industry (by allowing for free entry and introducing a fixed production cost); consider firms' relocation choices (the overall effect would depend on the extent of relocation costs versus the difference in expected operating profits

[15]See Bayoumi (1994) and Ricci (1997b) for models of the net benefits of a currency area. For surveys of the OCA literature, see de Grauwe (1992), Masson and Taylor (1992), and Tavlas (1994). Frankel and Rose (1998) argued the endogeneity of the OCA criterion on the basis of an empirical analysis, and refer to the mechanism described in this paper (or Ricci, 1997a) for the support of a formal model.

[16]In this case, however, prices would be set differently across sectors (as expected utility from profits is not linear in sales), and the model would need to be solved via simulations.

[17]Firms may dislike variability in sales for additional reasons such as the cost of firing workers, bankruptcy costs, and the cost of maintaining stocks of goods in order to smooth periods of excess demand. Moreover, firms would behave in a risk-averse manner if the owners are risk-averse and face incomplete financial markets or if managers are risk-averse and face an imperfect labor market or get nonmarketable payoffs (such as satisfaction or reputation).

across locations); or allow for probabilistic creation and destruction of firms, to analyze the effect of changes in regimes. All these interesting extensions are possibilities for future studies.

REFERENCES

Aizenman, Joshua, 1992, "Exchange Rate Flexibility, Volatility, and Domestic and Foreign Direct Investment," *IMF Staff Papers,* Vol. 39, No. 4, pp. 890–922.

Bayoumi, Tamim, 1994, "A Formal Model of Optimum Currency Areas," *IMF Staff Papers,* Vol. 41, No. 4, pp. 537–54.

Betts, Caroline, and Michael Devereux, 2000, "Exchange Rate Dynamics in a Model of Pricing-to-Market," *Journal of International Economics,* Vol. 50 (February), pp. 215–44.

Blanchard, Olivier, and Nobuhiro Kiyotaki, 1987, "Monopolistic Competition and the Effects of Aggregate Demand," *American Economic Review,* Vol. 77 (September), pp. 647–66.

Broll, Udo, Jack Wahl, and Itzhak Zilcha, 1995, "Indirect Hedging of Exchange Rate Risk," *Journal of International Money and Finance,* Vol. 14 (October), pp. 667–78.

Campa, Josè, and Linda Goldberg, 1995, "Investment in Manufacturing, Exchange Rates and External Exposure," *Journal of International Economics,* Vol. 38 (May), pp. 297–320.

Clark, Peter, Azim Sadikov, Natalia Tamirisa, Shang-Jin Wei, and Li Zeng, 2004, *A New Look at Exchange Rate Volatility and Trade Flows,* IMF Occasional Paper No. 235 (Washington: International Monetary Fund).

Cushman, David, 1988, "Exchange Rate Uncertainty and Foreign Direct Investment in the United States," *Weltwirtschaftliches Archiv,* Vol. 124 (2), pp. 322–36.

de Grauwe, Paul, 1992, *The Economics of Monetary Integration* (New York: Oxford University Press).

Dell'Ariccia, Giovanni, 1999, "Exchange Rate Fluctuations and Trade Flows: Evidence from the European Union," *IMF Staff Papers,* Vol. 46 (September), pp. 315–34.

Dixit, Avinash, and Joseph Stiglitz, 1977, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review,* Vol. 67 (June), pp. 297–308.

Faruqee, Hamid, 1995, "Pricing to Market and the Real Exchange Rate," *IMF Staff Papers,* Vol. 42, No. 4, pp. 855–81.

———, 1996, "Real Exchange Rate and the Pattern of Trade," *Journal of International Money and Finance,* Vol. 15, No. 2, pp. 313–36.

Fatás, Antonio, 1997, "EMU: Countries or Regions? Lessons from the EMS Experience," *European Economic Review,* Vol. 41 (April), pp. 743–51.

Fontagné, Lionel, and Michael Freudenberg, 1999, "Endogenous Symmetry of Shocks in a Monetary Union," *Open Economies Review,* Vol. 10, pp. 263–87.

Frankel, Jeffrey, and Andrew Rose, 1998, "The Endogeneity of the Optimum Currency Area Criteria," *The Economic Journal,* Vol. 108 (July), pp. 1009–25.

Fujita, Masahisa, Paul Krugman, and Anthony Venables, 1999, *The Spatial Economy: Cities, Regions, and International Trade* (Cambridge, Massachusetts: MIT Press).

Fujita, Masahisa, and Jacques-François Thisse, 1996, "Economics of Agglomeration," *Journal of the Japanese and International Economies,* Vol. 10 (December), pp. 339–78.

Gagnon, Joseph, 1993, "Exchange Rate Variability and the Level of International Trade," *Journal of International Economics,* Vol. 34 (May), pp. 269–87.

Goldberg, Linda, and Charles Kolstad, 1994, "Foreign Direct Investment, Exchange Rate Variability and Demand Uncertainty," *International Economic Review,* Vol. 36 (4), pp. 855–73.

Helpman, Elhanan, and Paul Krugman, 1985, *Market Structure and Foreign Trade* (Cambridge, Massachusetts: MIT Press).

Imbs, Jean, 1999, "Fluctuations, Bilateral Trade and the Exchange Rate Regime," Cahiers de Recherches Économiques du DEEP No. 9906 (Lausanne, Switzerland: Département d'économétrie et d'économie politique, Université de Lausanne).

Krugman, Paul, 1991, *Geography and Trade* (Cambridge, Massachusetts: MIT Press).

Masson, Paul R., and M. P. Taylor, 1992, "Common Currency Areas and Currency Unions: An Analysis of the Issues," CEPR Discussion Paper No. 617 (London: Center for Economic Policy Research).

Obstfeld, Maurice, and Kenneth Rogoff, 2000, "New Directions for Stochastic Open Economy Models," *Journal of International Economics,* Vol. 50 (February), pp. 117–53.

Ricci, Luca Antonio, 1997a, "Exchange Rate Regimes and Location," IMF Working Paper 97/69 (Washington: International Monetary Fund).

———, 1997b, "A Model of an Optimum Currency Area," IMF Working Paper 97/76 (Washington: International Monetary Fund); also "Un modèle simple de zone monétaire optimale," *Economie et Prévision,* Vol. 128.

———, 1997c, "A Ricardian Model of New Trade and Location Theory," *Journal of Economic Integration,* Vol. 12 (March), pp. 47–61.

———, 1998, "Uncertainty, Flexible Exchange Rates, and Agglomeration," IMF Working Paper 98/09 (Washington: International Monetary Fund); also *Open Economies Review* (forthcoming).

———, 1999, "Economic Geography and Comparative Advantage: Agglomeration Versus Specialization," *European Economic Review,* Vol. 43 (February), pp. 357–77.

Tavlas, George, 1994, "The Theory of Monetary Integration," *Open Economies Review,* Vol. 5 (March), pp. 211–30.

Tille, Cedric, 2006, "On the Distributional Effect of Exchange Rate Fluctuations" *Journal of International Money and Finance* (forthcoming).

IMF Staff Papers
Vol. 53, No. 1

Exchange Rate Pass-Through in the Euro Area

HAMID FARUQEE*

Exchange rate pass-through in a set of euro area prices along the pricing chain is examined in this paper. First, a vector autoregression (VAR) approach is used to analyze the joint time-series behavior of the euro exchange rate and a system of area-wide prices in response to an exchange rate shock. Second, the impulse-response functions from the VAR estimates are used to identify—in a "new open-economy macroeconomics model"—the key behavioral parameters that best replicate the pattern of exchange rate pass-through in the euro area. A key finding is that traded goods—both extra-area exports and imports—behave as though they are predominately priced in euros. The area-wide findings are compared with those for other major industrial economies. [JEL F41, F31, E31]

Against the backdrop of strong global growth, a low-flying recovery in the euro area has struggled to gain altitude, weighed down in part by a substantially stronger euro that had appreciated by roughly 45 percent against the U.S. dollar and 25 percent on a trade-weighted basis over the past three years (since its 2002:Q1 trough) before retreating some in mid-2005. With a struggling recovery, European concerns about the economic impact of past euro appreciation have figured prominently. Specifically, in the context of low area-wide inflation and soft domestic demand, the possible effects of recent exchange rate movements on inflation and trade have taken on renewed interest and significance. In assessing the likely consequences, a key determining factor is the nature of exchange rate pass-through in the euro area.

*Hamid Faruqee is a Senior Economist in the European Department of the International Monetary Fund. Special thanks are due to Susanna Mursula for her outstanding research assistance and to Bankim Chada, Ehsan Choudhri, Dalia Hakura, Albert Jaeger, and Alessandro Rebucci for helpful discussions and comments.

More difficult to assess, however, are the behavioral features that underlie the nature and extent of pass-through and the responsiveness of trade flows to the exchange rate. The relevance of exchange rate pass-through for inflation is straightforward. But these issues are also important in determining the strength of expenditure-switching effects from relative price signals.[1] Incomplete pass-through, for example, could delay or diminish the response in external variables and produce a certain degree of "exchange rate disconnect."[2] Thus, ascertaining the degree of and underlying behavior behind pass-through in the euro area is a key input for assessing its likely economic impact on growth and inflation.

Several economic explanations have been put forth to account for incomplete exchange rate pass-through—a feature that has strong empirical support for a large number of economies, including the euro area.[3] With nominal rigidity and local currency pricing (LCP), destination prices can change very little in the face of exchange rate variation.[4] With pricing-to-market behavior, segmented markets allow firms to stabilize their destination prices (via changing markups) to preserve foreign market share. In the presence of local distribution costs, firms may also face offsetting factors when the exchange rate changes, leading to international price discrimination and incomplete pass-through.[5] These factors can help account for differential responses between first-stage pass-through (e.g., in import prices) and second-stage pass-through (e.g., in consumer prices). Moreover, these considerations have been shown to have important implications for optimal monetary and exchange rate policies.[6]

This paper examines exchange rate pass-through and its behavioral determinants in the euro area. The methodology proceeds in two parts. First, the empirical analysis follows a vector autoregression (VAR) approach, in which the time-series behavior of the euro exchange rate and a system of euro area prices are examined. Specifically, the empirical analysis investigates exchange rate pass-through in a *set* of prices along the pricing chain. Second, the impulse-response functions (IRFs) from the VAR estimation are used to calibrate in a new open economy macroeconomics model the key behavioral parameters that can help reproduce the pattern of pass-through and external adjustment in the euro area.

The use of a VAR approach to examine exchange rate pass-through has several advantages compared with single-equation methods. Previous studies typically have focused on pass-through into a single price (e.g., import or consumer prices) without further distinguishing between the types of underlying exchange rate shocks (e.g., permanent or transitory) that may be arriving. By investigating exchange rate pass-through into a set of prices along the pricing chain, the VAR analysis characterizes not only absolute but *relative* pass-through in upstream and downstream

[1]See Obstfeld (2002) and Engel (2002) for recent reviews of these issues.

[2]See Krugman (1989) and Devereux and Engel (2002).

[3]See Goldberg and Knetter (1997) for a survey. See Kieler (2001), Hüfner and Schroder (2002), Anderton (2003), and Hahn (2003) for euro area estimates.

[4]See Betts and Devereux (1996, 2000) and Devereux and Engel (2002).

[5]See Corsetti and Dedola (2002) and Choudhri, Faruqee, and Hakura (2002).

[6]See, for example, Corsetti and Pesenti (2002) and Devereux and Engel (2003).

prices. Second, the VAR methodology potentially allows one to identify specific structural shocks affecting the system. In this case, a structural exchange rate shock is identified through a Cholesky decomposition of innovations, in which exchange rate fluctuations at higher frequencies are assumed to be largely driven by asset market—rather than goods market—disturbances.

Using this identification scheme, one can map the empirical results into a well-defined shock in an economic model of incomplete pass-through. Specifically, the estimated IRFs from an exchange rate shock in the VAR are matched to the relative response patterns generated by the corresponding asset market disturbance in the analytical pass-through model. By minimizing the distance between these sets of IRFs, one can identify key structural parameters that underlie the overall pattern of pass-through. The key parameters include (1) the pricing behavior of firms (i.e., the extent of local currency pricing), (2) the degree of nominal rigidity, and (3) the extent of local distribution and trading costs generating international price discrimination. A comparison of these behavioral parameters across the major industrial countries can then be made.

I. Empirical Estimates

The empirical analysis focuses on euro area prices and the exchange rate at a monthly frequency. The time span covers the period from 1990 through 2002. All series are expressed in logarithms. For the euro exchange rate *s,* the nominal effective series is defined for the European Central Bank's "narrow" group of partner economies.[7] For factor prices (i.e., wage earnings *w*), the series derive from quarterly data on nominal compensation per employee, extrapolated to a monthly frequency.[8] For trade prices, import prices *pm* and export prices *px* are based on unit values in extra-area manufacturing trade.[9] Producer prices *py* are based on the producer price index for manufacturing, excluding construction and energy. Consumer prices *pc* are based on the core consumer price index (CPI)—that is, excluding energy and unprocessed foods—from the harmonized index of consumer prices (HICP), although the implications with headline inflation are also discussed.[10] The exclusion of energy prices is motivated primarily by the standard finding that their pass-through behavior differs from that of other goods; by eliminating them, we facilitate a more seamless transition to the pass-through model later that emphasizes differentiated goods with imperfect substitutability.[11]

[7]Partners are Australia, Canada, Denmark, Hong Kong SAR, Japan, Korea, Norway, Singapore, Sweden, Switzerland, the United Kingdom, and the United States.

[8]Price and wage data are from Eurostat and are not seasonally adjusted. The areawide measures reflect aggregates of the 11 participating countries through December 2000. Thereafter, the chained series include Greece as the 12th member country.

[9]Manufactured goods include Sections 5–8 of the Standard International Trade Classification. The use of unit values (i.e., aggregrate, implicit deflators) is not ideal, but data limitations on direct trade prices necessitate their use.

[10]For Canada, Japan, the United Kingdom, and the United States, all corresponding monthly series were drawn from the IMF's *International Financial Statistics,* except wages, which were obtained from the Organization for Economic Cooperation and Development's Analytical Database.

[11]See, for example, Campa and Goldberg (2002).

The use of extra-area trade data helps avoid potential pitfalls from inferring pass-through for the euro area from estimates for individual member countries. To the extent that intra-area trade systematically differs from extra-area trade with respect to pass-through behavior, aggregating country estimates to generate an areawide measure could suffer from a "fallacy of composition." Hüfner and Schroder (2002), for example, construct pass-through estimates for euro area consumer prices by summing over individual country estimates, based on each country's weight in the areawide HICP. The analysis is then forced to make some "correction" of the estimates due to the presence of intra-area trade.

Before turning to the VAR estimation of euro area pass-through, some preliminary tests of the data were conducted. Unit root and stationarity tests indicate that these nominal variables are nonstationary in levels but stationary in first differences, suggesting that they are integrated-of-order-one or $I(1)$ series.[12] Furthermore, residual-based co-integration tests do not find evidence of co-integration among the variables (see Appendix). Given potential nonstationarity and lack of co-integration in the data in levels, estimating the VAR in first differences is appropriate.

VAR Methodology

The VAR approach examines the joint historical time-series behavior of the euro exchange rate and a system of euro area prices. Specifically, the reduced-form VAR(p) can be written as follows:

$$Y_t = c + A(L)Y_{t-1} + \mu_t;$$
$$E[\mu_t \mu_t'] = \Omega, \tag{1}$$

where $Y = [\Delta s\ \Delta w\ \Delta pm\ \Delta px\ \Delta py\ \Delta pc]'$; c is a vector of deterministic terms (i.e., monthly time dummies); A is a matrix polynomial of degree p in the lag operator L; and μ is the (6×1) vector of reduced-form residuals with variance-covariance matrix Ω. The exchange rate is placed first in the order of variables, reflecting the presumption that exchange rate innovations at monthly frequency are primarily driven by exogenous asset market disturbances.[13] For prices, the ordering after the exchange rate is motivated by the pricing chain, from factor input prices to trade prices to wholesale producer prices and retail consumer prices. The ordering among price variables after the exchange rate does not matter for the subsequent analysis of the exchange rate shock.

[12]The differenced series for euro area consumer and import prices were borderline nonstationary (see Appendix). But, as is well known, unit root and stationarity tests have low power, making it difficult to distinguish between stationary and unit root processes in finite samples.

[13]The identification scheme largely follows Choudhri, Faruqee, and Hakura (2005). That analysis also includes the interest rate in the VAR in order to further distinguish between the effects of interest rate and exchange rate shocks on prices. McCarthy (2000) and Hahn (2003) also use a Cholesky decomposition to examine pass-through based on a somewhat different model.

To recover the underlying exchange rate shock, the Cholesky decomposition of the matrix Ω is used to produce orthogonalized innovations ε. These disturbance terms are expressed in terms of the reduced-form VAR innovations as follows:

$$C\varepsilon_t = \mu_t, \tag{2}$$

where C is the unique lower triangular Cholesky matrix with 1*s* along its principal diagonal. Because the exchange rate appears first in the VAR, the recursive structure in equation (2) imposes the assumption that orthogonalized innovations to the exchange rate depend only on the residuals from the exchange rate equation and not from the other equations. This identification allows for a simple correspondence between the VAR estimates and a well-defined shock in the model described later. For prices, the corresponding disturbance term will represent a mix of shocks, including the structural exchange rate shock.[14]

An alternative identification scheme would place the exchange rate last (or near last) in the VAR. This ordering is motivated by the view that prices (and quantities) are predetermined in the very short run and, thus, cannot respond to an exchange rate shock; whereas the exchange rate can respond to various shocks.[15] Although this restriction may be valid for many prices, it may not be appropriate for others. More to the point, this ordering imposes a specific pass-through pattern in the estimates and, ultimately, certain behavioral features in the model that the current analysis seeks to investigate. Nevertheless, sensitivity analysis is reported later for reorderings of the VAR.

The implications of the identifying restriction can be further understood from the structural representation of the VAR:

$$F(L)Y_t = k + \varepsilon_t, \tag{3}$$

where $F(L)$ is a matrix polynomial of degree $p + 1$, k is a transformation of deterministic terms (i.e., $Ck = c$), and ε is the vector of structural shocks. One can show that the identification scheme based on the Cholesky decomposition introduces the following restriction: In the first equation (i.e., for the exchange rate Δs), the coefficients on contemporaneous price changes Δw, Δpm, Δpx, Δpy, and Δpc are equal to zero. Granger causality and block exogeneity tests find that (lagged) euro area prices have no predictive value for the euro exchange rate.[16] The VAR identification scheme takes this result one step further, assuming that concurrent price innovations also do not help explain exchange rate innovations.

[14]For the first variable (i.e., the exchange rate), the orthogonalized disturbance term is given by $\varepsilon_{1t} = \mu_{1t}$. For the jth variable ($j > 1$) in the VAR, the corresponding shock term is given by $\varepsilon_{jt} = \mu_{jt} - c_{j,1}\varepsilon_{jt} \ldots - c_{j,j-1}\varepsilon_{jt}$, where $c_{j,i}$ correspond to the entries of the Cholesky matrix (see Hamilton, 1994).

[15]See, for example, Peersman and Smets (2001).

[16]The exchange rate, however, helps predict (at least) trade prices (see Appendix). Restricted VAR estimates (not reported) excluding lagged prices from the exchange rate equation yield very similar results to those reported here.

The economic justification for this identifying assumption can be understood as follows. Exchange rates—especially at higher frequencies—are essentially driven by asset market rather than goods market disturbances. In the presence of noise traders, these short-run fluctuations may have little to do with economic fundamentals, including the price variables considered here.[17] The econometric assumption, though, is not necessarily incompatible with a more "fundamentalist" view. Under the asset market view of exchange rate determination, the predictable component is usually deemed small, and short-run changes in the exchange rate are likely to be dominated by "news."[18] In the universal presence of reporting lags, however, data releases on price indices, when they are made available, typically offer very little new information to move exchange rate markets. Furthermore, the empirical justification for this identification scheme is well known, given the overwhelming failure of exchange rate models to outperform a simple random walk over short horizons.[19]

Based on the reduced-form estimates of the VAR and the Cholesky decomposition to identify structural shocks, accumulated IRFs to a unit exchange rate shock are shown in Figure 1.[20] The horizontal axis measures the time horizon in terms of months after the shock; the vertical axis measures the deviation in (log) prices from their baseline levels.

Figure 1 shows that the impact effects of an exchange rate shock on prices are small in the euro area. Prices tend to be predetermined or very sticky (in local currency) initially in response to a depreciation in the euro effective exchange rate. Over time, the degree of exchange rate pass-through generally rises, although minimally so in factor and retail prices. Wholesale producer prices tend to rise more than retail consumer prices, but the greatest response is in trade prices. Twelve to 18 months after the shock, export prices respond by almost half the response in import prices, which moves in proportion to the exchange rate, suggesting full pass-through. Consequently, the (manufacturing) terms of trade for the euro area tend to worsen or decline in response to an exchange rate depreciation.[21] Choudhri, Faruqee, and Hakura (2005) report similar relative pass-through findings in a quarterly VAR for Group of Seven (G-7) countries that further include interest rates to account for monetary policy shocks.[22]

---

[17]See, for example, Jeanne and Rose (2002) and Devereux and Engel (2002).

[18]See Mussa (1984).

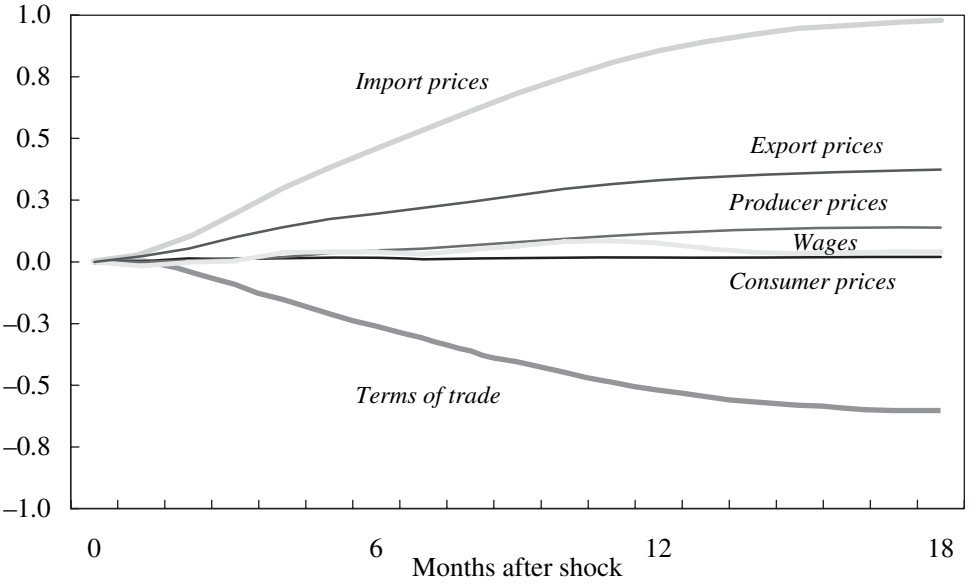
[19]See the seminal paper by Meese and Rogoff (1983). More recently, see, for example, Flood and Rose (1995) and Cheung, Chinn, and Pascual (2002). For a review, see Sarno and Taylor (2002).

[20]Given the large number of parameters in the VAR, a parsimonious lag structure is sought to conserve degrees of freedom. The lag length p is chosen by starting with given maximum lag length and sequentially testing the incremental significance of dropping an additional lag based on the likelihood ratio test. Derived from bootstrapping methods, confidence intervals for the IRFs are shown in Figure 3.

[21]Obstfeld and Rogoff (2000) provide evidence that the terms of trade decline with a currency depreciation—an observation that appears at odds with the implications of strict LCP models. See Lane (2001) for a review.

[22]The exchange rate path is typically found to be less persistent (i.e., more mean-reverting) in the quarterly VAR when interest rates are included (and first in the ordering), but the (absolute and relative) pass-through effects are more similar to those reported here.

Figure 1. Exchange Rate Shock on Euro Area Prices



Source: IMF Staff estimates.

Normalizing the price responses in Figure 1 by the path of the exchange rate response, Table 1 shows dynamic pass-through elasticities for euro area prices. After 18 months, pass-through rates in export and import prices are about one-half and one, respectively. Wage pass-through is relatively very small at 5 percent. Pass-through in wholesale prices is nearly 20 percent, significantly exceeding the pass-through in retail prices.[23] Using headline rather than core inflation (or including longer lags) would raise the degree of pass-through in consumer prices to near 10 percent at 18 months, but the pattern of relative pass-through remains intact.

Full pass-through in euro area import prices over time may be a somewhat surprising result.[24] Alternative specifications (e.g., with coefficient restrictions or VAR reorderings) tend to yield smaller but still high pass-through elasticities between 0.7 and 1. In a recent paper that also examines extra-area manufacturing trade prices, Anderton (2003) also finds generally high (albeit not full) pass-through between 0.5 and 0.7, based on a single-equation approach.[25] Given the identification of near-permanent exchange rate shocks here, it is not surprising that import

[23]Gagnon and Ihrig (2001) find low degrees of pass-through (around 5 percent) in consumer prices for 20 industrial countries, and argue that pass-through has been declining.

[24]These estimates should be taken as indicative, given the standard errors of the impulse-response functions. Based on bootstrapped standard errors, the 90 percent confidence interval for import prices is the widest, suggesting pass-through at 18 months in the range 0.6 to 1.4. The median value of this band suggests a slightly lower degree of import price pass-through (i.e., 0.95 at 18 months); otherwise, the median and estimated IRFs closely coincide.

[25]The time pattern is similar, though, with most of the price adjustment transpiring by five quarters. Based on a quarterly VAR, Hahn (2003) reports similar findings on the degree and time pattern of pass-through in the euro area non-oil import deflator.

Table 1. Euro Area Pass-Through Elasticities
(Percent change in prices divided by percent change in exchange rate)

	$t = 1$	$t = 6$	$t = 12$	$t = 18$
CPI	0.00	0.01	0.02	0.02
PPI	0.00	0.04	0.11	0.17
Wage	–0.02	0.04	0.07	0.05
Px	0.02	0.18	0.31	0.45
Pm	0.03	0.42	0.81	1.17

Source: IMF staff estimates.

Notes: CPI refers to consumer price index; PPI refers to producer price index; Wage refers to nominal compensation per employee; *Px* and *Pm* refer to export and import prices. Based on impulse-response functions from six-variable vector autoregression (VAR) estimated on monthly data from 1990 through 2002.

price pass-through in this VAR analysis lies somewhat above those single-equation estimates.

Cross-Country Evidence

Before relating the empirical findings to the theoretical model, it is useful to compare the pass-through results for the euro area to the other major industrial economies. Repeating the VAR exercise for the United States, Japan, the United Kingdom, and Canada produces the pass-through coefficients for trade prices shown in Table 2.

Table 2. Pass-Through Elasticities in Trade Prices: International Comparisons

	$t = 1$	$t = 6$	$t = 12$	$t = 18$
			Import Prices	
Euro area	0.03	0.42	0.81	1.17
United States	0.06	0.15	0.18	0.30
Japan	0.61	0.56	0.57	0.57
United Kingdom	0.28	0.58	0.57	0.60
Canada	0.68	0.54	0.62	0.68
			Export Prices	
Euro area	0.02	0.18	0.31	0.45
United States	0.00	0.00	0.06	0.12
Japan	0.62	0.50	0.48	0.47
United Kingdom	0.16	0.47	0.46	0.50
Canada	0.35	0.19	0.35	0.44

Source: IMF staff estimates.

Notes: Based on impulse-response functions from six-variable VAR estimated from 1990 through 2002. For euro area, import and export prices are based on unit values in manufacturing trade. For others, import and export prices are based on unit values in total trade.

As evident from the table, pass-through is incomplete, particularly in the short run. For the United States, pass-through in trade prices at the time of the shock is near zero, quite similar to the response of euro area. But while pass-through rises significantly over time in the euro area, the increase is much smaller for the United States. For the other countries, pass-through to import and export prices is higher on impact than for the United States and the euro area. For Canada and Japan, import price pass-through estimates are about 60–70 percent initially and remain around those levels over time; for the United Kingdom, import price pass-through eventually reaches 60 percent, but is initially half that. Pass-through in export prices is eventually around 50 percent for these three countries, similar to the euro area.

Sensitivity Analysis

As is well known, impulse responses from a VAR can be sensitive to the ordering of variables. In the six-variable VAR system, there are 6!=720 possible orderings under the Cholesky decomposition used to identify structural shocks. In general, when the reduced-form residuals from the VAR do not display high cross correlations, the order of factorization makes little difference.[26] Otherwise, the results can be sensitive to the choice of ordering. Since the focus here is solely on the effects of an exchange rate shock, suborderings among the price variables—given the exchange rate's order in the VAR—do not matter. Thus, the relevant reorderings to consider surround the placement of the exchange rate in the VAR and the combinations of price variables selected to appear either before or after the exchange rate, given its order in the VAR.[27]

Changes taken from this subset of possibilities, however, can be shown to have very little impact on the results for the euro area (and the United States); the results are robust to reorderings of the VAR. Take the extreme cases. Compared with the case in which the exchange rate appears first in the VAR, placing the exchange rate last in the VAR produces very similar pass-through elasticities for the euro area (see Appendix).[28] Similarly, placing different combinations of prices before or after the exchange rate does not materially affect the results. For the other countries, the results largely obtain so long as the trade prices appear after the exchange rate in the VAR. Otherwise, import and export prices would be predetermined (i.e., have zero response on impact) by *construction,* which, as shown in Table 2, does not reflect their behavior under less restrictive assumptions. Given the objective of investigating without prejudice the pass-through behavior in these key prices, making fewer (possibly erroneous) data restrictions a priori is clearly desirable.

[26]From the variance-covariance matrix, the correlations between residuals are less than 0.2, with the notable exceptions of the exchange rate and trade prices and between trade prices themselves.

[27]This reduces the number of relevant reorderings considerably. The possible cases can be enumerated as follows: $\sum_{x=0}^{5} 5Cx = 32$, where nCr denotes "n choose r."

[28]The pass-through results are very similar despite the fact that the impulse-response path for the exchange rate itself differs across the two orderings, displaying less persistence in the latter case where the exchange rate shock depends recursively on the reduced-form residuals from all the price equations. The interpretation of the exchange rate shock becomes more difficult in this case, since it incorporates a mix of innovation terms.

II. Model of Incomplete Pass-Through

This section describes the analytical framework used to interpret the VAR evidence in terms of underlying economic behavior. The stylized model generates incomplete pass-through by drawing on many of the common themes found in the new open economy macroeconomics (NOEM) paradigm, following the seminal work of Obstfeld and Rogoff (1995).[29] A detailed derivation of the microfounded model can be found in Choudhri, Faruqee, and Hakura (2005). A sketch of the model follows.

Imperfect competition characterizes the production and allocation of two differentiated goods—a traded intermediate good and a nontraded final (consumption) good. One differentiated primary factor (labor) and intermediate inputs enter the production of traded goods. A domestic retail sector relying on labor services is needed to transform intermediate goods into final consumption goods in each country. This structure gives rise to a pricing chain with five price indices: retail consumer prices, wholesale producer prices, import and export prices, and factor prices (i.e., wages).

Beginning with the demand side and consumer preferences from the home country's perspective, expected lifetime utility of a household, indexed by $l \in [0, 1]$, is assumed to be

$$E_t \sum_{\tau=t}^{\infty} \beta^{\tau-t} \left[\frac{1}{1-\rho} C_\tau(l)^{1-\rho} - \frac{1}{1+\mu} L_\tau(l)^{1+\mu} \right], \tag{4}$$

where $C_\tau(l)$ and $L_\tau(l)$ are the household's consumption basket and labor supply. The consumption basket, reflecting household preferences, is a constant-elasticity-of-substitution (CES) aggregate over individual varieties, indexed by $c \in [0, 1]$, of the final good and is given by

$$C_t = \left[\int_0^1 C_t(c)^{1-1/\varepsilon} \, dc \right]^{\varepsilon/(\varepsilon-1)}. \tag{5}$$

Each consumption good variety, meanwhile, is produced according to the following technology:

$$C_t(c) = Q_t(c)^{\gamma} N_{Ct}(c)^{1-\gamma}, \tag{6}$$

where $Q_t(c)$ is an index of both home and foreign varieties of the intermediate good, and $N_{Ct}(c)$ is a bundle of differentiated labor services. This technology combines traded goods with nontraded labor services to produce consumption services.

The technology for producing a home variety, indexed by *y,* of the intermediate good is given by:

$$Y_t(y) = Z_t(y)^{\alpha} N_{Yt}(y)^{1-\alpha}, \tag{7}$$

[29]See Lane (2001) for a survey.

where $Z_t(y)$ and $N_{Yt}(y)$ are composites of differentiated intermediate and labor inputs. From these two equations, note that intermediate goods (indexed by Q) are used directly in the production of the final consumer good, while others (indexed by Z) are used in the production of other intermediate goods.

Specifically, each home intermediate variety is allocated across the following four sectors:

$$Y_t(y) = Y_{QHt}(y) + Y_{ZHt}(y) + Y^*_{QMt}(y) + Y^*_{ZMt}(y), \tag{8}$$

where $Y_{QHt}(y)$ and $Y_{ZHt}(y)$ denote the amounts of the home variety y used to meet domestic final and intermediate demands, and $Y^*_{QMt}(y)$ and $Y^*_{ZMt}(y)$ denote the exports of the home variety y used to satisfy foreign final and intermediate demands.

Using $Y_{QMt}(y^*)$ and $Y_{ZMt}(y^*)$ to denote the corresponding amounts of an imported foreign variety, indexed by $y^* \in [0, 1]$, the intermediate CES input bundles in the home country are defined as

$$Q_t = \left[v^{1/\sigma} Q_{Mt}{}^{1-1/\sigma} + (1-v)^{1/\sigma} Q_{Ht}{}^{1-1/\sigma}\right]^{\sigma/(\sigma-1)}, \tag{9}$$

$$Z_t = \left[v^{1/\sigma} Z_{Mt}{}^{1-1/\sigma} + (1-v)^{1/\sigma} Z_{Ht}{}^{1-1/\sigma}\right]^{\sigma/(\sigma-1)}, \tag{10}$$

$$Q_{Mt} = \left[\int_0^1 Y_{QMt}(y^*)^{1-1/\varepsilon} dy^*\right]^{\varepsilon/(\varepsilon-1)}, \; Z_{Mt} = \left[\int_0^1 Y_{ZMt}(y^*)^{1-1/\varepsilon} dy^*\right]^{\varepsilon/(\varepsilon-1)}, \tag{11}$$

$$Q_{Ht} = \left[\int_0^1 Y_{QHt}(y)^{1-1/\varepsilon} dy\right]^{\varepsilon/(\varepsilon-1)}, \; Z_{Ht} = \left[\int_0^1 Y_{ZHt}(y)^{1-1/\varepsilon} dy\right]^{\varepsilon/(\varepsilon-1)}. \tag{12}$$

Here, the elasticity of substitution σ between domestic and foreign baskets of the intermediate good is allowed to be different than the elasticity ε between varieties within each bundle. Similar equations would apply to the foreign country.

Imported varieties (in both countries) are also assumed to go through distribution channels before their use, and the distribution process requires local labor services. Following Corsetti and Dedola (2002), one unit of an imported variety y^* in the home country requires δ units of the labor service bundle, $N_{Mt}(y^*)$, so that[30]

$$N_{Mt}(y^*) = \delta\left[Y_{QMt}(y^*) + Y_{ZMt}(y^*)\right]. \tag{13}$$

[30]The Leontief-type distribution process is needed only for moving intermediate goods across borders. For simplicity, the distribution cost for imports is the same whether they are used in the production of the intermediate or the final good. Note, however, that the consumption "technology" in equation (6) specifies a Cobb-Douglas process for transforming both home and foreign intermediates into a final consumer good.

Firms that face these local trading or distribution costs δ have an incentive for pricing to market (PTM) or price discrimination across local and export markets.[31] With this dependence on local currency wages, changes in the exchange rate lead to incomplete pass-through in trade prices and the PTM behavior described by Krugman (1987) through a cost or supply mechanism.[32] The introduction of pricing to market along these lines helps limit the effects of import price pass-through in retail prices and helps generate greater overall persistence in the degree of incomplete pass-through over time.[33]

Given this structure, household labor supply employed in the production of intermediate and final goods and distribution services,

$$L_t(l) = L_{Yt}(l) + L_{Ct}(l) + L_{Mt}(l), \tag{14}$$

and aggregate labor service bundles in the three activities are combined as follows:[34]

$$N_{Yt} = \left[\int_0^1 L_{Yt}(l)^{1-1/\varepsilon}\, dl\right]^{\varepsilon/\varepsilon-1}, \quad N_{Ct} = \left[\int_0^1 L_{Ct}(l)^{1-1/\varepsilon}\, dl\right]^{\varepsilon/(\varepsilon-1)},$$
$$N_{Mt} = \left[\int_0^1 L_{Mt}(l)^{1-1/\varepsilon}\, dl\right]^{\varepsilon/(\varepsilon-1)}. \tag{15}$$

Marginal costs for final and intermediate goods derive from the employment of labor services and the costs of intermediated goods. From the production technologies in equations (6) and (7), these are

$$MC_{Ct} = P_{Qt}^{\gamma} W_t^{1-\gamma} \Big/ \left(\gamma^{\gamma}(1-\gamma)^{1-\gamma}\right), \quad MC_{Yt} = P_{Zt}^{\alpha} W_t^{1-\alpha} \Big/ \left(\alpha^{\alpha}(1-\alpha)^{1-\alpha}\right), \tag{16}$$

where P_{Qt} and P_{Zt} are the familiar price indices associated with the CES bundles for intermediate inputs Q_t and Z_t in equations (9) and (10), and W_t is the aggregate wage.

Given the CES demands for each variety derived (ultimately) from household preferences, monopolistically competitive firms set price as a markup over marginal costs to maximize profits. But prices and wages are not fully flexible. Instead, they

[31]Investigating extra-area import prices, Anderton (2003) finds that foreign suppliers attach a significant weight to the PTM strategy in efforts to maintain market share. Herzberg, Kapetanios, and Price (2003) find PTM to be the dominant consideration for U.K. import prices. Kieler (2001) provides comparative estimates for several industrial countries.

[32]See also Kasa (1992) and Faruqee (1995) for analyses that introduce market-specific costs as a way to generate incomplete pass-through and PTM behavior. An alternative approach focusing on varying markups and demand elasticities through translog (i.e., non-CES) preferences can be found in Bergin and Feenstra (2001).

[33]The home import price (in local currency terms) of a foreign variety, after and before distribution costs, is $\tilde{P}_{Mt}(y^*) = P_{Mt}(y^*) + \delta W_t$ and thus depends partly on the domestic wage W.

[34]Individual labor demands facing households depend on relative wages and aggregate labor demand: $L_t^d(l) = (N_{Ct} + N_{Yt} + N_{Mt})(W_t(l)/W_t)^{-\varepsilon}$. See Choudhri, Faruqee, and Hakura (2005).

are updated only infrequently based on Calvo-type adjustment.[35] Specifically, the (fixed) probability that a firm will leave its price unchanged at a point in time is equal to π. Correspondingly, the average interval over which prices remain fixed is $\pi/(1-\pi)$.[36] This framework generates staggered price-setting behavior in the economy.

With nominal rigidity in an open economy, the question arises as to whether prices are sticky in terms of domestic or foreign currency. The two limiting cases are referred to as (1) local currency pricing (LCP), in which all prices are rigid in the destination or buyer's currency; and (2) producer currency pricing (PCP), in which all prices are rigid in the origin or seller's currency. Allowing for a hybrid case, both types of pricing behavior are nested in the model. Specifically, the parameters ϕ and ϕ^* denote the share of domestic and foreign firms following PCP behavior respectively. If $\phi = 1$, one has strict PCP; $\phi = 0$ represents strict LCP.

In the PCP case, a home producer would set its home export price $P_{Xt}(y)$ at the point of origin; under LCP, the same firm would set the foreign import price $P^*_{Mt}(y)$ at the destination, where $P^*_{Mt}(y) = P_{Xt}(y)/S_t$ and S_t is the nominal exchange rate. Using superscripts P and L to denote prices setting under each type of behavior, the values X^P_{Yt} and X^P_{Xt} represent (under PCP) the domestic and export contract prices under Calvo-type adjustment chosen by firms at time t to maximize the present discounted value of profits:

$$E_t \sum_{\tau=t}^{\infty} DR_{t,\tau}\pi^{\tau-t}\left\{\left(X^P_{Yt} - MC_{Y\tau}\right)\left[\left(Z_{H\tau} + Q_{H\tau}\right)P^{\epsilon}_{Y\tau}\left(X^P_{Yt}\right)^{-\varepsilon}\right]\right.$$
$$\left.+\left(X^P_{Xt} - MC_{Y\tau}\right)\left[\left(Z^*_{M\tau} + Q^*_{M\tau}\right)P^{*\varepsilon}_{M\tau}\left(X^P_{Xt}/S_\tau + \delta^* W^*_\tau\right)^{-\varepsilon}\right]\right\}, \tag{17}$$

where $DR_{t,\tau}$ is the stochastic discount rate (consistent with preferences), $\pi^{\tau-t}$ is the probability that a price set at t will remain in place at τ, and the square-bracketed expressions represent the domestic and foreign demands for the home variety, conditional on the price set at t.[37] The analogous objective functions under LCP (and for retail firms) can be similarly derived.

Solving equation (17) and its analog for retailers and for foreign firms—and under both PCP and LCP—yields the following central (log-linearized) pricing equations for aggregate producer, consumer, import, and export prices:

$$p_{Y,t} = \pi p_{Y,t-1} + (1-\pi)x_{Y,t}, \tag{18}$$

$$p_{C,t} = \pi p_{Y,t-1} + (1-\pi)x_{C,t}, \tag{19}$$

[35]See Calvo (1983). See Kollmann (2001) for a NOEM analysis with sticky wages and prices but without distribution costs.

[36]Choudhri, Faruqee, and Hakura (2002) also examine the cases of wage and/or price flexibility (i.e., $\pi = 0$) and find that these model variants generally fall short in explaining the empirical impulse-responses.

[37]The firms' stochastic discount factor is consistent with that of households, based on the Euler condition in consumption: $DR_{t,\tau} = \beta^{\tau-t}\frac{P_{Ct}C^{\rho}_t}{P_{C\tau}C^{\rho}_\tau}$ and $E_t\beta\frac{P_{Ct}C^{\rho}_t}{P_{C,t+1}C^{\rho}_{t+1}} = \frac{1}{1+r_t}$, where r is the rate of interest.

$$p_{M,t} = p^*_{Y,t} + \phi^* s_t + \pi\left(p_{M,t-1} - p^*_{Y,t-1} - \phi^* s_{t-1}\right) + (1-\pi)\left[x_{D,t} + (1-\phi^*)x_{St}\right], \tag{20}$$

$$p_{X,t} = p_{Y,t} + (1-\phi)s_t + \pi\left(p^*_{M,t-1} - p_{Y,t-1} + \phi\, s_{t-1}\right) + (1-\pi)\left[x^*_{D,t} - (1-\phi)x_{St}\right], \tag{21}$$

where lowercase letters denote logarithms of variables. In the presence of Calvo-type staggered price adjustment, aggregate price indices display inertial dynamics depending on the parameter related to the probability of price (non)adjustment π. In the import and export price-setting equations, the behavior is also influenced strongly by the shares ϕ, ϕ^* of LCP versus PCP firms. The underlying contracted prices (or components) entering these expressions are given by

$$x_{Y,t} = \beta\pi E_t x_{Y,t+1} + (1-\beta\pi)mc_{Y,t}, \tag{22}$$

$$x_{C,t} = \beta\pi E_t x_{C,t+1} + (1-\beta\pi)mc_{C,t}, \tag{23}$$

$$x_{S,t} = \beta\pi E_t x_{S,t+1} + (1-\beta\pi)s_t, \tag{24}$$

$$x_{D,t} = \beta\pi E_t x_{D,t+1} + (1-\beta\pi)\psi_{CD}\left(w_t - s_t - mc^*_{Y,t}\right). \tag{25}$$

The last equation (and its counterpart, $x^*_{D,t}$) reflects the role of distribution costs in the setting of trade prices, where ψ_{CD} is a weight related to the importance of unit distribution costs δ relative to marginal production costs for imports.[38] Along with a similar inertial equation for wages, equations (18)–(21) represent the major behavioral price-setting equations in the model, and $\theta = \{\phi, \phi^*, \pi, \delta\}$ is the vector of key structural parameters used to help map model dynamics into those generated by the VAR with respect to exchange rate pass-through.

To close the model, monetary policy is determined by an interest rate rule, and uncovered interest rate parity is assumed to hold (up to an error term):

$$r_t = \varphi_0 + \varphi_1 r_{t-1} + \varphi_2 E_t \Delta p_{Ct} + \varphi_3 \Delta s_t, \tag{26}$$

$$s_t = E_t s_{t+1} + r_t^* - r_t + \xi_t, \tag{27}$$

where ξ represents an exchange rate shock, perhaps reflecting the impact of noise traders.[39]

[38]Specifically, one can derive $\psi_{CD} = \dfrac{\delta W/\varepsilon}{MC^*_Y/S + \delta W/\varepsilon}$, evaluated at steady-state values. Details of the log-linearization are available from the author upon request.

[39]In a noise-trader model, market participants can be subject to a stochastic bias in their expectations, $E^f_t s_{t+1} - E_t s_{t+1} = \xi_t$, generating a UIP or exchange rate shock in equation (27).

Benchmarks for parameters held fixed in the model follow the calibration in Choudhri, Faruqee, and Hakura (2005), and are generally based on average estimated values for the major industrial countries—including euro area members Germany, France, and Italy.[40] Monetary policy, for example, is specified by an interest rate rule targeting expected inflation in consumer prices and allowing for interest rate smoothing, with the corresponding parameter weights based on the average estimates for the (non-U.S.) G-7 countries.

III. Matching Theory and Empirics

To match the empirical impulse-response patterns, an asset market shock ξ to uncovered interest rate parity (UIP) in the model is considered. The scale and persistence of ξ is calibrated to reproduce the VAR's accumulated impulse-response path for the exchange rate from a structural exchange rate shock ε_s.[41] Simulating the model's price responses to this UIP shock generates the analytical impulse-response functions. To find the optimal structural parameter values in the model consistent with the empirical evidence, the following loss function is minimized:[42]

$$\min_{\{\theta\}} \left[IR(P,\xi;\theta) - IR^{VAR}(P,\varepsilon_s) \right]' \left[IR(P,\xi;\theta) - IR^{VAR}(P,\varepsilon_s) \right], \tag{28}$$

where $IR(P, \xi; \theta)$ denotes the vector of simulated impulse responses for the price vector $P = \{w, px, pm, pc, py\}$ from an ξ shock in the model conditional on the key set of structural parameters $\theta = \{\phi, \phi^*, \pi, \delta\}$.[43] $IR^{VAR}(P, \varepsilon_s)$ denotes the corresponding vector of accumulated impulse responses to a structural exchange rate shock ε_s in the VAR. The time horizon for the impulse-response paths is 18 months.

Figure 2 displays the combinations of LCP and PCP behavior in domestic and foreign firms that best replicate the price responses—particularly for trade prices—derived from the VAR estimates for each of the major industrial economies. The vertical axis measures the extent of PCP behavior ϕ^* in foreign firms (i.e., producing home imports), and the horizontal axis measures the extent of PCP behavior ϕ in domestic firms (i.e., producing home exports).

In the figure, the origin represents local currency pricing at home and abroad, while the opposing corner represents the case of universal producer currency pricing. The 45-degree line connecting them reflects symmetric combinations of local

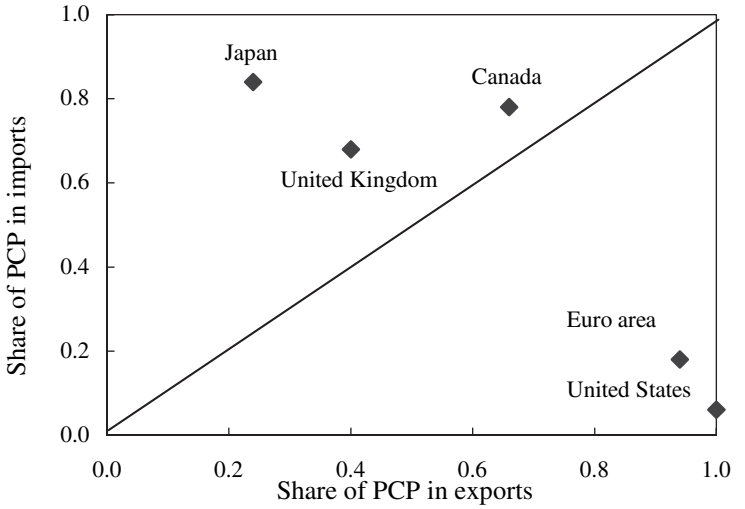
[40]Reflecting the higher (monthly) frequency, the discount factor and interest rate are adjusted to produce consistent annualized rates with Choudhri, Faruqee, and Hakura (2005).

[41]The exchange rate shocks from the VAR have permanent or highly persistent effects on the log level of the exchange rate. Consequently, the ξ shocks in the model are constructed to be permanent or highly persistent as well.

[42]The standard errors for the accumulated impulse-response trajectories are similar, suggesting that minimizing the unweighted least squares should yield parameters broadly similar to the weighted least squares approach used in Smets and Wouters (2002). The weighted approach would give less weight to import prices in favor of consumer prices.

[43]The elasticity of substitution σ between traded goods, broadly defined, was also chosen to minimize the distance between IRFs and found to be fairly low ($\sigma = 2$), in line with Choudhri, Faruqee, and Hakura (2002); Smets and Wouters (2002); and the references cited therein. The elasticity of substitution θ between specific varieties was also allowed to vary within a range ($\theta = 4$ to 10), consistent with plausible markups.

Figure 2. Pricing Behavior of Domestic and Foreign Firms



and producer currency pricing. In terms of symmetric outcomes, the midpoint along the 45-degree line tends to outperform either extreme empirically, as shown by Choudhri, Faruqee, and Hakura (2002). The intuition is as follows: Exclusive reliance on LCP behavior in both countries produces stable import prices at the destination, which generally squares well with the data, at the expense of unstable export prices at the point of origin, which does not. The reverse is true in the case of PCP. As is evident from the VAR results in Table 2, both export and import prices in a given currency show incomplete pass-through. Consequently, a mix of the two behaviors fares better.

Figure 2 indicates, though, that the opposing diagonal is more relevant empirically in describing the pricing behavior of traded goods across countries. Specifically, *asymmetric* pricing behavior between domestic and foreign firms appears more consistent with the data. Allowing for LCP to a greater extent in certain trade prices and PCP to a greater extent in others tends to be more consistent than equal degrees of the two. Specifically, the smaller, more open economies and Japan gravitate toward pricing in foreign currencies (i.e., LCP in their exports and PCP in their imports); whereas the larger, more insular economies of the United States and the euro area gravitate toward pricing in their domestic currencies (i.e., LCP in their imports and PCP in their exports). As further supporting evidence, the implicit currency pricing behavior suggested by Figure 2 appears very consistent with independent data on invoice currencies for these countries.[44]

For the United Kingdom, pricing behavior is consistent with an outward orientation. Domestic firms exporting to foreign markets tend toward LCP and set and

[44]Bekx (1998) reports the following 1995 shares of import and export prices that are respectively invoiced in domestic currency (figures in percent): Germany (52,75), Japan (23,36), the United Kingdom (43,62), and the United States (81,92). Note that the percentages for Germany refer to deutschemarks rather than all euro area legacy currencies.

stabilize prices in terms of foreign or destination currencies, safeguarding export market share. Import prices also tend to be rigid in terms of foreign currencies, as foreign exporting firms are less sensitive to fluctuations in their destination currency prices. Canada is similar to the United Kingdom, with foreign currencies playing a large role in price-setting behavior of traded goods, albeit closer to the symmetric or equal weighting case.

Despite its relative economic size and modest trade openness, Japan exhibits considerable outward orientation in its pricing behavior. From Table 2, recall that Japanese import and export prices show considerable pass-through on impact akin to those of the smaller, more open economies considered here. These results suggest, in the model, that Japanese exporting firms predominantly engage in local currency pricing, while foreign firms largely engage in producer currency pricing. This finding accords with the conventional wisdom regarding the behavior of Japanese and foreign (predominantly U.S.) exporting firms described in Giovannini (1988), Marston (1990), and others.[45]

For the larger, more insular economies, excluding Japan, both domestic and foreign firms tend to price in the dominant currency associated with the large economy. For foreign firms exporting to the United States and to the euro area, this translates into destination or local currency pricing in dollars and euros, respectively. For U.S. and euro area firms exporting abroad, this translates into origin or producer currency pricing. For the United States, this result is well established in the literature.

The similarity between the United States and the euro area is somewhat striking. While it is generally accepted that U.S. firms exporting to the euro area (and elsewhere) primarily invoice in dollars and engage in dollar-currency pricing, less is known about the pricing behavior of euro area firms.[46] The results here suggest that the euro area already behaves in some measure like the United States with respect to the predominant role of domestic currency pricing for traded goods. The finding is largely driven by the similar degree of low pass-through on impact in U.S. and areawide import and export prices shown in Table 2. The use of manufacturing trade prices for the euro area accentuates this similarity; although using total extra-area trade prices (which show higher pass-through) would only slightly alter the relative placements shown in Figure 2.

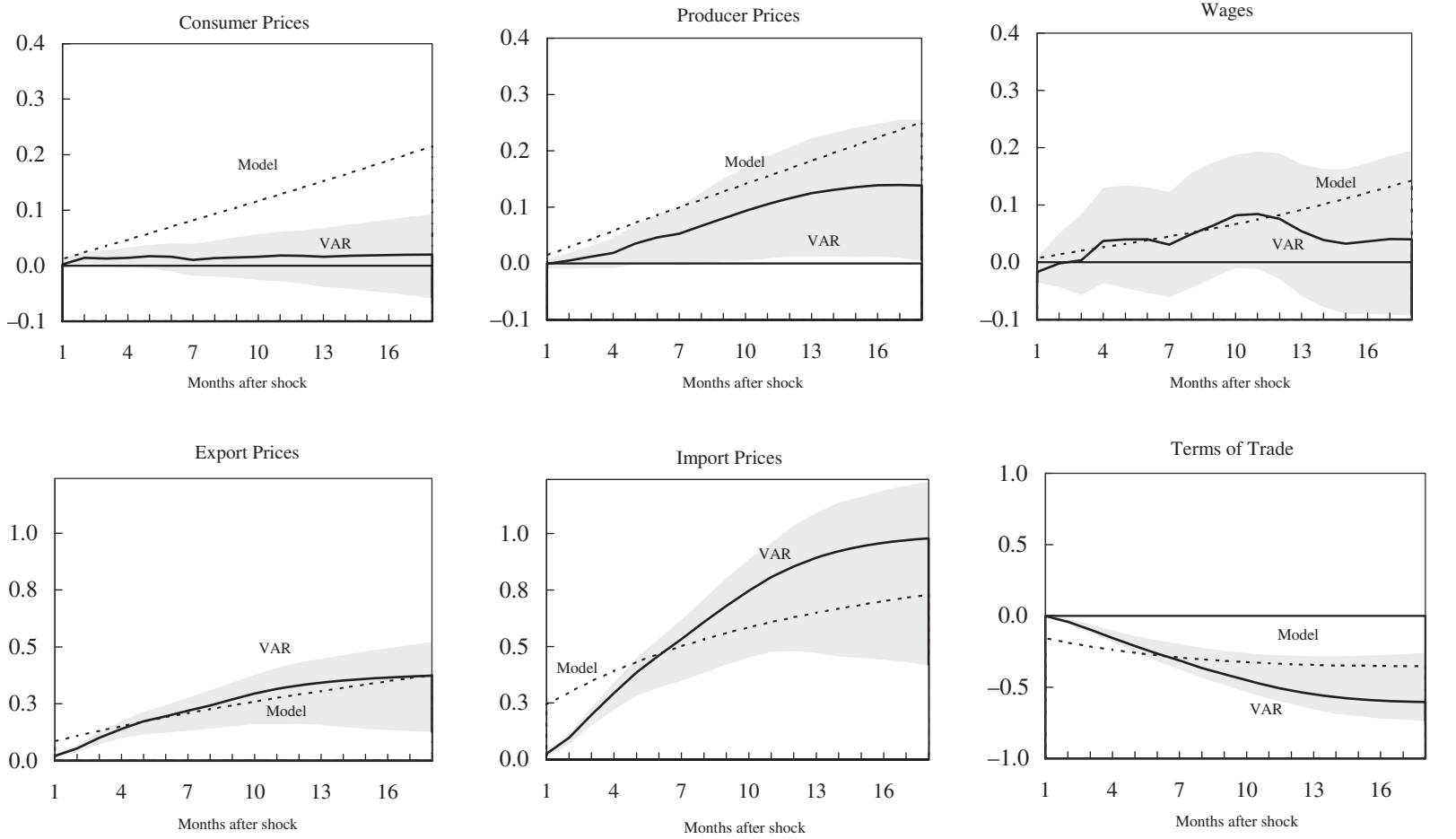
Some important differences between the United States and the euro area are worth highlighting. Specifically, the dynamic response of prices suggests significantly higher pass-through over time (particularly in import prices) for the euro area. Interestingly, this translates, in the model, into a higher estimate for the Calvo parameter π, reflecting greater nominal flexibility.[47] Imposing a similar π value would shift the euro area in Figure 3 toward the other countries and away from the

[45]See Dominguez (1999) for a related discussion of the limited role of the yen as an international currency, particularly as an invoicing currency. In principle, the choice of invoice currency is distinct from the issues of local versus producer currency pricing, but these issues share considerable overlap in practice.

[46]Some have suggested that the international role of the euro in this regard will expand over time. See Bekx (1998) and Devereux, Engel, and Tille (1999).

[47]The estimates for π typically range from 5 percent to 10 percent, suggesting an average duration between price changes of three to six quarters. The π estimates for the euro area (U.S.) are at the higher (lower) end of this range, suggesting shorter (longer) durations between price changes.

Figure 3. Euro Area Empirical and Theoretical Impulse-Response Paths to Unit Exchange Rate Shock
(*90 percent confidence intervals in shaded area*)



Source: IMF staff estimates.

Table 3. Implicit Estimates of Local "Distribution" Costs
(Percent of final goods price)

United States	Euro Area	Japan	United Kingdom	Canada
45–65	55–65	65–70	60–65	50–65

Source: IMF staff estimates.

Note: Estimates based on minimum distance between calibrated model and country VAR impulse-response functions for alternative degrees of price rigidity.

United States in terms of the mix of LCP versus PCP firms, although the euro area would remain more closely aligned with the United States.

In all countries, exchange rate pass-through in consumer prices is significantly lower than in import prices. In the model, this wedge between intermediate and final goods is captured by local distribution, marketing, and trading costs. Consequently, implicit estimates of these costs are fairly substantial. Country estimates of distribution costs—measured as a share of the final goods price—are shown in Table 3. The estimates are expressed as broad ranges, reflecting the fact that certain combinations of distribution costs and the degree of nominal rigidity produced similar fits of the model. The overall range is from 45 percent for the United States to 70 percent for Japan.[48] Burstein, Neves, and Rebelo (2003) estimate a 40 percent share for distribution costs in the United States.

Table 4 reports a summary of the "optimized" structural parameters that minimize the distance between the IRFs from the model and each country's VAR. The table also provides an *R*-squared value, based on the minimized loss or sum of squared deviations between model and VAR normalized by the total sum of squares for prices from the VAR impulse-response trajectories. The fit of the pass-through model is generally very good overall, albeit less so for the United States. In the case of the euro area, the impulse responses from the model compared with those of the VAR, including 90 percent confidence intervals, are shown in Figure 3.

Local currency pricing and PTM behavior notwithstanding, expenditure switching effects still operate in these economies, although short-run trade elasticities may be small.[49] Based on the calibrated model, an exchange rate depreciation can be shown to improve the trade balance overall, once both volume and pass-through effects are factored in.[50] Given its generally small effects on the trade balance,

[48]High estimates of implicit distribution costs in Japan may reflect aspects of its peculiar distribution system—including restrictive government regulations, predominance of small stores, complex marketing channels, and pervasive constraints on vertical integration. See Flath (2003).

[49]The classic Marshall-Lerner-Robinson (MLR) condition requires that the sum of trade elasticities exceed unity for a depreciation to improve the trade balance under traditional pass-through assumptions (i.e., zero and full pass-through in export and import prices, respectively). For example, Isard and others (2001) use elasticity benchmarks of 0.7 and 0.9 for export and import volumes. When pass-through is incomplete, however, the MLR condition need not apply.

[50]Using the calibrated model described here, Faruqee (2003) examines the trade impact of a uniform dollar decline on the external positions of the major industrial economies.

Table 4. Summary of Behavioral Parameter Estimates[1]

	United States	Euro Area	Japan	United Kingdom	Canada
ϕ	1.00	0.94	0.24	0.40	0.66
ϕ^*	0.04	0.18	0.80	0.68	0.78
π	0.05	0.10	0.05	0.05	0.10
δ[2]	0.65	0.60	0.65	0.65	0.60
R-squared	0.61	0.91	0.98	0.91	0.88

Source: IMF staff estimates.

[1]Optimized values based on minimum distance between calibrated model and country VAR impulse-response functions. *R*-squared based on sum of squared deviations between model and VAR impulse responses normalized by total sum of squares from VAR impulse responses over 18 months.

[2]Local distribution costs expressed in percent of final goods price.

however, the exchange rate adjustment required to achieve a moderate degree of external adjustment may be substantial in the absence of adjustment in other variables.

IV. Conclusions

Incomplete exchange rate pass-through in prices is a well-documented empirical regularity for many economies, including the euro area. A better understanding of the economic behavior underlying limited pass-through is an important consideration for investigating the implications of currency fluctuations and the role of monetary and exchange rate policy. The analysis here has sought to examine pass-through in a set of euro area prices along the pricing chain by using a VAR approach to identify the effects of an exogenous exchange rate shock. Mapping the effects of this structural shock into an analytical framework helps identify behavioral features that could help account for the nature of incomplete pass-through in the euro area. On the basis of the analysis, the following conclusions can be drawn.

- Short-run pass-through is low in the euro area for a wide range of prices. Similar to the United States, the impact effect of an exchange rate shock on factor and trade prices, and on wholesale and retail prices, is near zero.
- Pass-through tends to rise over time in the euro area, although the extent of wage and consumer price pass-through remains comparatively small. Pass-through in producer and export prices is somewhat higher, but the highest degree of pass-through (near unity) is reflected in euro area import prices. The differences in relative pass-through in import prices and consumer prices, for example, suggest that the roles of the retail sector and local distribution costs are important for price determination.
- The pattern of pass-through in trade prices suggests a fair degree of asymmetry with respect to the pricing behavior of domestic and foreign firms operating in the euro area. Specifically, the impulse-response patterns suggest a high degree

of local currency pricing in import prices and producer currency pricing in export prices. For the euro area (and the United States), this suggests that these traded goods are priced predominantly in euros (and dollars). For Japan and smaller, more open economies, the behavior of pass-through is more consistent with importers and exporters operating significantly in foreign currencies.

- Local currency pricing and pricing to market behavior notwithstanding, expenditure switching effects still operate in these economies, although short-run trade elasticities can be small. Nevertheless, the model suggests that an exchange rate depreciation improves the trade balance, once both volume and pass-through effects are factored in. Given its generally small effects on the trade balance, however, the exchange rate adjustment required to achieve a moderate degree of external adjustment may, in some circumstances, be substantial in the absence of adjustment in other variables.

APPENDIX

Unit Root and Stationarity Tests

Test statistics based on the Phillip and Perron (1998) nonparametric unit root test and the stationarity test of Kwiatowski, Phillips, Schmidt, and Shin (KPSS; 1992) are shown in Table 5. The alternative hypothesis with the Phillips-Perron test and the null hypothesis with the KPSS test is trend stationarity. The tests suggest that the levels (first-differences) of these variables are nonstationary (stationary). The borderline cases, in which the tests give conflicting answers, are CPI inflation (where the tests reject a unit root and stationarity, respectively) and import price inflation (where tests fail to reject a unit root and stationarity, respectively).

Table 5. Unit Root and Stationarity Tests
(Euro area monthly data, 1990–2002)

Variable	Phillips-Perron Z_t Test	KPSS η_τ Test	Order of Integration
NEER	–2.34	0.34**	I(1)
Δ NEER	–8.46**	0.07	I(0)
CPI	–2.89	0.74**	I(1)
Δ CPI	–10.87**	0.36**	I(0) or I(1)
PPI	–1.66	0.32**	I(1)
Δ PPI	–5.86**	0.09	I(0)
Px	–0.92	0.27**	I(1)
Δ *Px*	–3.29*	0.10	I(0)
Pm	–2.38	0.15**	I(1)
Δ *Pm*	–2.61	0.11	I(0) or I(1)
Wage	–2.46	0.63**	I(1)
Δ Wage	–6.56**	0.07	I(0)

Source: IMF staff estimates.

Notes: *(**) indicates significance at the 10 (5) percent level. NEER refers to nominal effective exchange rate. Phillips-Perron unit root test includes deterministic time trend under the alternative. The Kwiatowski and others (KPSS; 1992) stationarity test includes deterministic time trend under the null.

Table 6. Co-Integration Tests
(Euro area monthly data, 1990–2002)

Number of Regressors	$\hat{P}_z$ Test (demeaned)	$\hat{P}_z$ Test (demeaned and detrended)
n = 5	115.10 (225.23)	120.43 (284.01)

Source: IMF staff estimates.

Notes: Multivariate trace statistic based on Phillips and Ouliaris (1990); 10 percent critical value given in parentheses.

Co-Integration Tests

Given that the unit root and stationarity tests suggest that the log levels of the exchange rate and various price measures are nonstationary, co-integration tests are conducted to examine whether a linear combination of these variables is stationary. Table 6 reports the results of the Phillips and Ouliaris (1990) residual-based test for co-integration with window size = 2. Other window sizes produce similar results. The tests fail to reject the null of no co-integration.

Granger Causality Tests

Granger causality tests were conducted to examine whether changes in euro area exchange rates and prices have predictive content for each other. Simple bivariate tests indicate that the nominal effective exchange rate Granger-causes (i.e., helps predict) several price measures, but prices fail to Granger-cause exchange rates. The exchange rate is found to be a significant predictor for trade prices and has predictive content for wages at shorter lag length (see Table 7).

Block Exogeneity Tests

To generalize Granger causality tests to a multivariate context, consider the following partitioned VAR(p) system:

Table 7. Bivariate Granger Causality Tests
(Euro area monthly data, 1990–2002)

Price Measure	Lags	H_0 : Exchange Rate Does Not Cause Price Measure	Lags	H_0 : Price Measure Does Not Cause Exchange Rate
CPI	12	0.66	2	0.96
PPI	14	0.99	2	0.87
Px	13	0.06	2	0.27
Pm	13	0.00	2	0.70
Wage[1]	14	0.54	2	0.47

Source: IMF staff estimates.

Notes: Reported numbers are p-values on the relevant exclusion restriction (F-test). Lag length selection based on Akaike Information Criterion.

[1]At one lag, the exchange rate Granger-causes wages ($p = 0.04$).

Table 8. Block Exogeneity Tests
(Euro area monthly data, 1990–2002)

Lag Length	$H_0 : A_2 = 0$	$H_0 : B_1 = 0$
$p = 7$	32.42	50.48
	(0.59)	(0.04)

Source: IMF staff estimates.

Notes: Test statistic is based on likelihood ratio test with degrees of freedom correction and is distributed as $\chi^2(5p)$. Significance levels are given in parentheses.

$$\begin{aligned} \Delta s_t &= c_1 + A_1 X_{1t} + A_2 X_{2t} + \varepsilon_{1t} \\ \Delta P_t &= c_2 + B_1 X_{1t} + B_2 X_{2t} + \varepsilon_{2t} \end{aligned},$$

where c_1 and c_2 represent a constant term and monthly time dummies, P represents the (5×1) vector of price variables, X_1 and X_2 are $(p \times 1)$ and $(5p \times 1)$ vectors of lagged changes in exchange rates and prices, respectively, with conformable matrices of autoregressive coefficients A_1, A_2, B_1, and B_2. Block exogeneity test results are shown in Table 8.

Sensitivity Analysis

As seen in Table 9, the implied pass-through elasticities from the two VARs (i.e., with the exchange rate first and last in the ordering) are fairly similar. On impact, euro area prices are essentially predetermined in response to the exchange rate shock in the two VARs. It should be stressed, however, that the exchange rate shock has zero contemporaneous effect on prices (i.e., prices are exactly predetermined) *by construction* in the latter VAR, based on the ordering of

Table 9. Euro Area Pass-Through Elasticities Across VAR Reorderings

	$t = 1$	$t = 6$	$t = 12$	$t = 18$
		(Exchange Rate First in VAR)		
CPI	0.00	0.01	0.02	0.02
PPI	0.00	0.04	0.11	0.17
Wage	–0.02	0.04	0.07	0.05
Px	0.02	0.18	0.31	0.45
Pm	0.03	0.42	0.81	1.17
		(Exchange Rate Last in VAR)		
CPI	0	0.02	0.02	0.03
PPI	0	0.05	0.11	0.18
Wage	0	0.04	0.08	0.05
Px	0	0.10	0.20	0.30
Pm	0	0.30	0.63	0.97

Source: IMF staff estimates.

Notes: Entries report percent change in price measure divided by percent change in exchange rate in response to unit exchange rate shock from Cholesky decomposition of innovations.

variables. In the case of the euro area (and the United States) this restriction appears valid empirically. Hence, the reordering makes little difference. For other industrial countries, however, this is generally not true for trade prices, as shown by Table 2, and should thus not be imposed a priori. So long as the exchange rate appears *before* trade prices in the VAR for these countries, alternative impulse-response functions will closely match those reported in the text.

REFERENCES

Anderton, Robert, 2003, "Extra-Euro Area Manufacturing Import Prices and Exchange Rate Pass-Through," ECB Working Paper No. 219 (Frankfurt: European Central Bank).

Bekx, Peter, 1998, "The Implications of the Introduction of the Euro for Non-EU Countries," Euro Papers 26 (Brussels: European Commission).

Bergin, Paul, and Robert C. Feenstra, 2001, "Pricing-to-Market, Staggered Contracts, and Real Exchange Rate Persistence," *Journal of International Economics,* Vol. 54 (August), pp. 333–59.

Betts, Caroline, and Michael B. Devereux, 1996, "The Exchange Rate in a Model of Pricing-to-Market," *European Economic Review,* Vol. 40 (April), pp. 1007–21.

———, 2000, "Exchange Rate Dynamics in a Model of Pricing-to-Market," *Journal of International Economics,* Vol. 50 (February), pp. 215–44.

Burstein, Ariel T., João C. Neves, and Sergio Rebelo, 2003, "Distribution Costs and Real Exchange Rate Dynamics During Exchange-Rate-Based Stabilizations," *Journal of Monetary Economics,* Vol. 50 (September), pp. 1189–214.

Calvo, Guillermo, A., 1983, "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics,* Vol. 12 (September), pp. 383–98.

Campa, Jose M., and Linda S. Goldberg, 2002, "Exchange Rate Pass-Through into Import Prices: A Macro or Micro Phenomenon?" Staff Reports No. 149 (New York: Federal Reserve Bank of New York).

Cheung, Y., M. Chinn, and A. Pascual, 2002, "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?" NBER Working Paper No. 9393 (Cambridge, Massachusetts: National Bureau of Economic Research).

Choudhri, Ehsan U., Hamid Faruqee, and Dalia S. Hakura, 2002, "Explaining the Exchange Rate Pass-Through in Different Prices," IMF Working Paper 02/224 (Washington: International Monetary Fund).

———, 2005, "Explaining the Exchange Rate Pass-Through in Different Prices," *Journal of International Economics,* Vol. 65 (March), pp. 349–74.

Choudhri, Ehsan U., and Dalia S. Hakura, 2001, "Exchange Rate Pass-Through to Domestic Prices: Does the Inflationary Environment Matter?" IMF Working Paper 01/194 (Washington: International Monetary Fund).

Corsetti, Giancarlo, and Luca Dedola, 2002, "Macroeconomics of International Price Discrimination," ECB Working Paper No. 176 (Frankfurt: European Central Bank).

Corsetti, Giancarlo, and Paolo Pesenti, 2002, "Self-Validating Optimum Currency Areas," NBER Working Paper No. 8783 (Cambridge, Massachusetts: National Bureau of Economic Research).

Devereux, Michael, and Charles Engel, 2002, "Exchange Rate Pass-Through, Exchange Rate Volatility, and Exchange Rate Disconnect," *Journal of Monetary Economics,* Vol. 49 (July), pp. 913–40.

———, 2003, "Monetary Policy in the Open Economy Revisited: Price Setting and Exchange-Rate Flexibility," *Review of Economic Studies,* Vol. 70 (4), pp. 765–83.

———, and Cedric Tille, 1999, "Exchange Rate Pass-Through and the Welfare Effects of the Euro," NBER Working Paper No. 7382 (Cambridge, Massachusetts: National Bureau of Economic Research).

Dominguez, Kathryn, 1999, "The Role of the Yen," in *International Capital Flows,* ed. by Martin Feldstein (Chicago: University of Chicago Press), pp. 133–71.

Engel, Charles, 2002, "Expenditure Switching and Exchange Rate Policy," in *NBER Macroeconomics Annual 2002,* ed. by Mark Gertler and Kenneth Rogoff (Cambridge, Massachusetts: MIT Press).

Faruqee, Hamid, 1995, "Pricing to Market and the Real Exchange Rate," *IMF Staff Papers,* Vol. 42 (December), pp. 855–81.

———, 2003, "Exchange Rate Pass-Through and External Adjustment in the Euro Area," in *Euro Area Policies, Selected Issues,* IMF Country Report No. 03/298 (Washington: International Monetary Fund).

Flath, David, 2003, "Regulation, Distribution Efficiency, and Retail Density," NBER Working Paper No. 9450 (Cambridge, Massachusetts: National Bureau of Economic Research).

Flood, Robert P., and Andrew K. Rose, 1995, "Fixing Exchange Rates, A Virtual Quest for Fundamentals," *Journal of Monetary Economics,* Vol. 36 (August), pp. 3–37.

Gagnon, Joseph, and Jane Ihrig, 2001, "Monetary Policy and Exchange Rate Pass-Through," International Finance Discussion Papers No. 704 (revised March 2002) (Washington: Board of Governors of the Federal Reserve System).

Giovannini, Alberto, 1988, "Exchange Rates and Traded Goods Prices," *Journal of International Economics,* Vol. 24 (February), pp. 45–68.

Goldberg, Penelopi, and Michael Knetter, 1997, "Goods Prices and Exchange Rates: What Have We Learned?" *Journal of Economic Literature,* Vol. 35 (September), pp. 1243–72.

Hahn, Elke, 2003, "Pass-Through of External Shocks to Euro Area Inflation," ECB Working Paper No. 243 (Frankfurt: European Central Bank).

Hamilton, James D., 1994, *Time Series Analysis* (Princeton, New Jersey: Princeton University Press).

Herzberg, Valerie, George Kapetanios, and Simon Price, 2003, "Import Prices and Exchange Rate Pass-Through: Theory and Evidence from the United Kingdom," Bank of England Working Paper No. 182 (London: Bank of England).

Hüfner, Felix P., and Michael Schroder, 2002, "Exchange Rate Pass-Through to Consumer Prices: A European Perspective," Center for European Economic Research Discussion Paper 02-20 (Mannheim, Germany: Center for European Economic Research).

Isard, Peter, Hamid Faruqee, G. Russell Kincaid, and Martin Fetherston, 2001, *Methodology for Current Account and Exchange Rate Assessments,* IMF Occasional Paper No. 209 (Washington: International Monetary Fund).

Jeanne, Olivier, and Andrew K. Rose, 2002, "Noise Trading and Exchange Rate Regimes," *Quarterly Journal of Economics,* Vol. 117 (May), pp. 537–69.

Kasa, Kenneth, 1992, "Adjustment Costs and Pricing-to-Market Theory and Evidence," *Journal of International Economics,* Vol. 32 (February), pp. 1–30.

Kieler, Mads, 2001, "Tracking the Pass-Through of External Shocks to Euro Area Inflation," in *Monetary and Exchange Rate Policies of the Euro Area—Selected Issues,* prepared by A. Jaeger, K. Ross, Z. Kontolemis, M. Kieler, and G. Meredith, IMF Country Report No. 01/201 (Washington: International Monetary Fund), pp. 52–67.

Kollmann, Robert, 2001, "The Exchange Rate in a Dynamic-Optimizing Business Cycle Model with Nominal Rigidities: A Quantitative Investigation," *Journal of International Economics,* Vol. 55 (December), pp. 243–62.

Krugman, Paul, 1987, "Pricing to Market When the Exchange Rate Changes," in *Real-Financial Linkages Among Open Economics,* ed. by Sven W. Arndt and J. David Richardson (Cambridge, Massachusetts: MIT Press), pp. 49–70.

———, 1989, *Exchange Rate Instability* (Cambridge, Massachusetts: MIT Press).

Kwiatowski, Denis, Peter Phillips, Peter Schmidt, and Yongcheol Shin, 1992, "Testing The Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics,* Vol. 54 (October–December), pp. 159–78.

Lane, Philip, 2001, "The New Open Economy Macroeconomics: A Survey," *Journal of International Economics,* Vol. 54 (December), pp. 235–66.

Marston, Richard, 1990, "Pricing to Market in Japanese Manufacturing," *Journal of International Economics,* Vol. 29 (November), pp. 217–36.

McCarthy, Jonathan, 2000, "Pass-Through of Exchange Rates and Import Prices to Domestic Inflation in Some Industrialized Economies," Staff Reports 111 (New York: Federal Reserve Bank of New York).

Meese, Richard, and Kenneth Rogoff, 1983, "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics,* Vol. 14 (February), pp. 3–24.

Mussa, Michael, 1984, "The Theory of Exchange Rate Determination," in *Exchange Rate Theory and Practice,* ed. by John F. O. Bilson and Richard C. Marston (Chicago: University of Chicago Press), pp. 13–78.

Obstfeld, Maurice, 2002, "Exchange Rates and Adjustment: Perspectives from the New Open Economy Macroeconomics," NBER Working Paper No. 9118 (Cambridge, Massachusetts: National Bureau of Economic Research).

———, and Kenneth Rogoff, 1995, "Exchange Rate Dynamics Redux," *Journal of Political Economy,* Vol. 103 (June), pp. 624–60.

———, 2000, "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" NBER Working Paper No. 7777 (Cambridge, Massachusetts: National Bureau of Economic Research).

Peersman, Gert, and Frank Smets, 2001, "The Monetary Transmission Mechanism in the Euro Area: More Evidence from VAR Analysis," ECB Working Paper No. 91 (Frankfurt: European Central Bank).

Phillips, P.C.B., and S. Ouliaris, 1990, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica,* Vol. 58 (January), pp. 165–93.

Phillips, P.C.B., and P. Perron, 1998, "Testing for a Unit Root in Time Series Regressions," *Biometrika,* Vol. 75(2), pp. 335–46.

Sarno, L., and M. Taylor, 2002, *The Economics of Exchange Rates* (Cambridge, United Kingdom: Cambridge University Press).

Smets, Frank, and Raf Wouters, 2002, "Openness, Imperfect Exchange Rate Pass-Through and Monetary Policy," *Journal of Monetary Economics,* Vol. 49 (July), pp. 947–81.

IMF Staff Papers
Vol. 53, No. 1

Parity Reversion in Real Exchange Rates: Fast, Slow, or Not at All?

PAUL CASHIN AND C. JOHN MCDERMOTT*

This paper tests for purchasing power parity (PPP) using real effective exchange rate data for 90 developed and developing countries in the post–Bretton Woods period. Support for PPP is found, since the majority of countries experience finite deviations of real exchange rates from parity. The speed of parity reversion is found to be typically much faster for developed countries than for developing countries and to be considerably faster for countries with flexible nominal exchange rate regimes compared with countries having fixed nominal exchange rate regimes. [JEL C22, F31, F41]

Exchange rates have been at the center of policy and academic debates in developing and developed countries, especially since the floating of developed-country exchange rates in the early 1970s, which marked the commencement of the post–Bretton Woods period. Despite these debates, several key empirical questions regarding the stylized facts of exchange rates remain largely unresolved, particularly in the little-researched area of real exchange rate behavior in developing countries (Edwards and Savastano, 2000). Several of these exchange rate–related questions are tackled in this paper: Do real exchange rates really display parity-reverting behavior? Is purchasing power parity (PPP) an appropriate (very) long-run benchmark for

*Paul Cashin is Deputy Division Chief in the Caribbean I Division of the IMF's Western Hemisphere Department; John McDermott is Associate Professor, School of Economics and Finance, Victoria University, New Zealand. The authors thank Robert Flood, Sam Ouliaris, Christopher Plantier, Carmen Reinhart, Kenneth Rogoff, Lauren Rosborough, Lucio Sarno, Miguel Savastano, Mark Taylor, and seminar participants at the 12th Meeting of the New Zealand Econometrics Study Group and the International Monetary Fund for helpful comments and suggestions on earlier versions of the paper, and Chi Nguyen and Manzoor Gill for excellent research assistance.

assessing real exchange rate developments? Does the behavior of real exchange rates differ among developed and developing countries? Are there important differences in the extent of parity reversion when comparing countries with fixed nominal exchange rate regimes and those with flexible nominal exchange rate regimes? Can we explain cross-country heterogeneity of parity reversion using countries' structural characteristics, and, if so, which characteristics appear to be important? In this study we examine the empirical support for PPP through time-series analysis of the persistence of shocks to the real effective exchange rates of 90 developed and developing countries in the post–Bretton Woods period.

The theory of PPP, in its most rudimentary form, states that there is an equilibrium level to which exchange rates converge, so that foreign currencies should have the same purchasing power.[1] Therefore, long-run PPP is inconsistent with a unit root in real exchange rates. The reason for this is that a shock to a unit root process will have permanent effects on all future values of the series, potentially without bound.[2] Unfortunately, formal statistical tests that compare the unit root (UR) model against the alternative of a stationary autoregressive (AR) model typically lead to a failure to reject the hypothesis of a unit root.

There are a number of econometric problems with using the UR/AR model to test for PPP. First, least squares estimators of AR models bias empirical results in favor of finding PPP. This downward bias becomes particularly acute when the AR parameter is close to unity. As lower values of the AR parameter imply faster speeds of adjustment following a shock, this will also result in a downward bias to estimates of half-lives of shocks. This near unit root bias is likely to be particularly relevant for real exchange rates, as they are often found to be stationary yet exhibit shocks that are highly persistent. Second, unit root tests tend to be uninformative regarding the speed of parity reversion, because a rejection of the unit root null could still be consistent with a stationary model of real exchange rates that has highly persistent shocks. This leads to problems about how to interpret results and often yields arbitrary conclusions that are dependent on the predisposition of the researcher.

This paper makes several contributions to the literature. First, it reports median-unbiased estimates of the half-lives of real exchange rate adjustment for 90 developed and developing countries over the post–Bretton Woods period. To the best of our knowledge, this study is the first to apply median-unbiased estimation methods to developing-country real exchange rate data.[3] Second, we use Andrews' (1993) unbiased model-selection rule to draw conclusions as to the presence (or absence) of parity reversion of real exchange rates in the post–Bretton Woods period. In particular, finite half-lives constructed from bias-corrected estimation indicate that

[1]Since the present study of PPP employs data on price indices rather than price levels, we are examining the relevance of relative PPP: the notion that the percentage change in the nominal effective exchange rate should compensate for the inflation differential between the home country and a weighted average of partner countries. This test of relative PPP is more stringent than the usual test, which uses first differences of the real exchange rate.

[2]For surveys on PPP and exchange rate economics, see Froot and Rogoff (1995), Sarno and Taylor (2002), and Taylor (2003).

[3]Recent work on testing PPP and the stationarity of real exchange rates in developing countries includes Bahmani-Oskooee (1995) and Montiel (1997), among others; studies concentrating on Latin American countries include Devereux and Connolly (1996) and Calvo, Reinhart, and Végh (1995).

real exchange rate shocks are mean-reverting. Third, we break down the half-lives estimated by geographic region, income level, dominant exportable, and exchange rate regime. Fourth, we attempt to explain the heterogeneity in the duration of parity reversion by examining bivariate correlations between estimated half-lives and potential nominal factors (inflation and choice of nominal exchange rate regime) and potential real factors (trade openness, productivity growth, and government spending). Finally, this study extends Cashin and McDermott (2003) by covering a larger sample of country real exchange rates for the post–Bretton Woods period, and attempting to explain the factors driving the observed heterogeneity of the duration of exchange rate deviations from parity.

We have several key findings. First, using post–Bretton Woods data on the real effective exchange rates of 90 industrial and developing countries and least squares estimation of unit root models, we replicate Rogoff's (1996) consensus estimate of the half-life of deviations from PPP of three to five years. Second, using median-unbiased estimates, we find that for 40 countries in our sample, deviations of the real exchange rate from parity are best viewed as being permanent. However, the majority (50) of the countries have finite half-life estimates—this evidence of real exchange rate reversion to parity is consistent with PPP holding in the post–Bretton Woods period. Third, the median half-life of parity deviations for industrial countries (eight years) is much shorter than that for developing countries (permanent). Fourth, the median half-life of parity deviation for countries with fixed nominal exchange rate regimes (permanent) is considerably longer than that for countries with flexible nominal exchange rate regimes (six years). Fifth, two key structural characteristics that affect the persistence of shocks to real exchange rates are differences in a country's inflation experience and the extent of nominal exchange rate volatility. Finally, the existing literature on measuring parity reversion in developing countries has largely concentrated on the Latin America experience (which is dominated by high-inflation countries). This sample-selection bias has resulted in the erroneous finding that parity reversion in developing countries is faster than that observed in developed countries.

I. Measures of Speed of Parity Reversion

The presence of unit roots in real exchange rates is incompatible with long-run PPP. Consequently, unit root tests that focus on whether shocks are mean-reverting or not have been widely used to test for long-run PPP. However, it is preferable to use measures of the speed of reversion in evaluating PPP, because a very persistent real exchange rate can still be incompatible with how PPP is interpreted in theories of exchange rate determination. Conversely, failure to reject a unit root test may occur as a result of the well-known low power properties of unit root tests, when in fact PPP may hold in longer-term data. As a consequence, recent papers examining the persistence of exchange rate shocks in the post–Bretton Woods period have used estimates of the half-life of parity deviations (Andrews, 1993; Andrews and Chen, 1994; and Cheung and Lai, 2000b).[4]

[4]Biased and bias-corrected point estimates of the half-life of shocks to economic time series have also been used by Cashin, Liang, and McDermott (2000) in modeling the persistence of shocks to world

We use the AR(p) model to measure the degree of persistence in real exchange rates. The AR(p) model (also known as an Augmented Dickey-Fuller, or ADF, regression) takes the form

$$q_t = \mu + \alpha q_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta q_{t-1} + \varepsilon_t \quad \text{for } t = 1,\ldots,T, \tag{1}$$

where the observed real exchange rate series is q_t: $t = -p, \ldots, T$, μ the intercept, α the autoregressive parameter (where $\alpha \in (-1,1]$), and ε_t are the innovations of the model. The lagged first differences are included to control for the presence of serial correlation. The α parameter in equation (1) can be corrected for the downward bias present in such regression by using the methods suggested by Andrews and Chen (1994).[5] The bias correction delivers an impartiality property to the decision-making process, because there is an equal chance of under- or overestimating the AR parameter. Moreover, an unbiased estimate of α will allow us to calculate an unbiased scalar estimate of persistence—the half-life of a unit shock.

Model Selection Rule

Testing the PPP proposition is a controversial subject that has not yet been resolved. Most modern time-series tests of PPP are based on testing whether $\alpha < 1$ in AR models of the real exchange rate, as in equation (1). Unit root tests based on classical hypothesis tests tend to have a null hypothesis where PPP does not hold—that is, the null hypothesis is that PPP deviations are permanent. Further, because of the low power properties of these tests (at least against reasonable alternatives), the tests often fail to reject the null. In particular, slow (albeit positive) reversion of real exchange rates toward PPP yields unit root tests that provide little information against relevant alternative hypotheses (Froot and Rogoff, 1995). The typically large estimated confidence intervals around the point estimate of α make it clear that the level of uncertainty about the value of the "true" α is very high. Nevertheless, practitioners often use hypothesis tests as a formal model-selection rule.

Andrews (1993) offers a solution to this impasse by introducing a statistical procedure whereby the probability of selecting the "true" model is at least as large as the probability of selecting the false model. Unbiased model-selection procedures have an impartiality property that may be useful if the selection of one model or another (such as the trend stationary or unit root model) is a contentious issue. In essence, the issue is to select one of two models defined by $\alpha \in I_a$ and $\alpha \in I_b$, where I_a and I_b are intervals partitioning the parameter space $(-1, 1]$ for α, with $I_a = (-1, 1)$ and $I_b = \{1\}$ (Andrews, 1993, p. 152).

commodity prices; by McDermott (1996), Murray and Papell (2002), and Cashin and McDermott (2003) in modeling the persistence of parity deviations in developed-country real exchange rates; and by Cashin, McDermott, and Pattillo (2004) in examining the persistence of terms of trade shocks.

[5]See Cashin and McDermott (2003) for details on the downward bias of autoregressive parameters derived from least squares estimates of AR models and Andrews' (1993) method for bias-correcting the least squares estimator.

Calculating Half-Lives

The main objective of using the autoregressive model is to extract the degree of persistence inherent in real exchange rates. Specifically, we use the half-life of shocks implied by such models to measure persistence. The half-life is defined as the duration of time it takes for a unit impulse to dissipate permanently by one half from the occurrence of the initial shock (Cheung and Lai, 2000b). That is, the half-life (denoted by l_h) is given by $IR(l_h) \leq \frac{1}{2}$ such that $l_{h+k} < \frac{1}{2}$ for $k = 1, 2, \ldots$, where the impulse-response function is given by $IR(h) = f_{11}^{(h)}$ for $h = 0, 1, 2, \ldots$, and $f_{11}^{(h)}$ denotes the (1,1) element of F^h, where F is the $(p \times p)$ matrix

$$F \equiv \begin{bmatrix} \alpha+\psi_1 & \psi_2-\psi_1 & \psi_3-\psi_2 & \cdots & \psi_{p-1}-\psi_{p-2} & -\psi_{p-1} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

and

$$\alpha_i \equiv \begin{cases} \alpha+\psi_1 & i=1, \\ \psi_i-\psi_{i-1} & i=2,\ldots,p-1, \\ -\psi_{p-1} & i=p. \end{cases}$$

We could also construct two-sided 90 percent (or one-sided 95 percent) confidence intervals for the half-life estimates, as in Cashin, Liang, and McDermott (2000); Cheung and Lai (2000b); and Murray and Papell (2002). Because such intervals are typically very wide, the null hypothesis that the half-life is permanent is rarely rejected. Such a finding leaves us in the awkward position of having to provisionally accept that PPP does not hold or resigning ourselves to the fact that we cannot make a decision at all because the degree of uncertainty attached to the estimate of the half-life is too high. Murray and Papell (2002) favor the latter option.

However, empirical researchers will typically (for better or worse) make a judgment as to whether the model is acceptable or not. Failure to reject the null of a unit root often invokes the argument that the test was not powerful enough, and researchers then proceed as if the null is false. This approach seems unsatisfactory and is one that will often deliver different conclusions from different researchers, because each is now using an arbitrary decision rule—an undisclosed and implicit decision rule at that. Instead, a more natural procedure is to use the unbiased model-selection rule of Andrews (1993), which offers an objective means to determine which model best represents the data. Moreover, because the rule is clear and explicit, different researchers can use it to verify results in an objective manner. Use of Andrews' (1993) unbiased model-selection rule and median-unbiased estimation

means that we have used a 0.5 (unbiased) test of the null hypothesis of a finite half-life versus the alternative hypothesis of a permanent half-life (see Cashin and McDermott, 2003).

There are other methods we could have used to examine the cross-country experiences of the persistence of parity deviations. One approach is to use long-span time series to analyze the persistence of real exchange rates. However, this approach is clouded by the differing nominal exchange rate regimes used by any given country over time. In order to provide a useful test of the validity of the notion of purchasing power parity in the post–Bretton Woods period, a test needs to be derived using data specifically from that period. Another approach would be to use multivariate generalizations of unit root tests of the real exchange rate (through panel unit root analyses of long-run PPP). This approach also has substantial problems. In particular, the use of panel unit root tests has been criticized because authors have typically presumed that rejection of the joint null hypothesis of unit root (nonreversion) behavior of the whole panel of real exchange rates implies that all real exchange rates are stationary. In actuality, rejection of the null only implies that at least one of the real exchange rates is stationary (or mean-reverting). A third approach would be to use nonlinear methods, as Taylor and Sarno (1998) do. In some sense, this approach complements the approach used in this paper.

II. Data

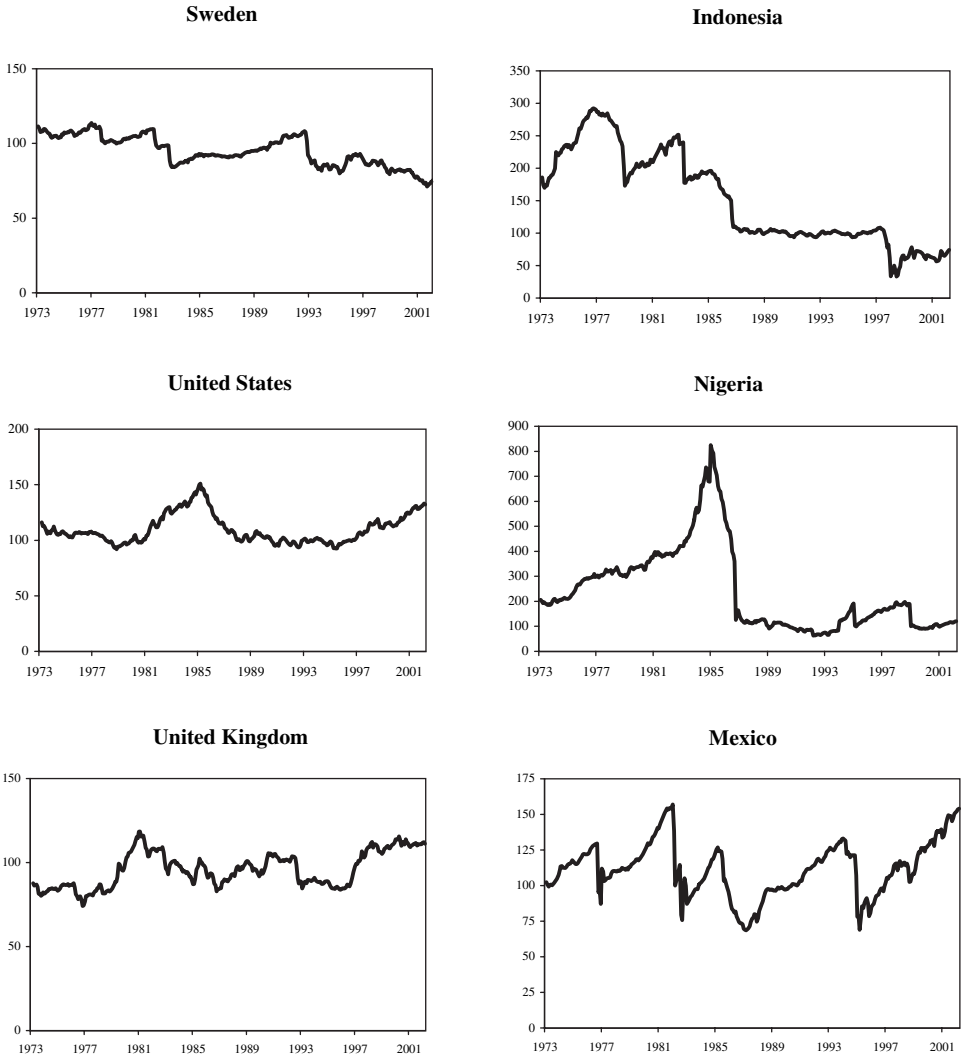
In this and the following section, we will investigate the properties of real exchange rate persistence. The data used to estimate the near unit root model are monthly time series of the real exchange rate obtained from the International Monetary Fund's *International Financial Statistics* (*IFS*) from the third quarter of 1973 to the third quarter of 2002 (the post–Bretton Woods period), which gives a total of 348 observations (see Appendix I for additional details). Appendix I also lists the derivation and description of the variables used in Section IV to explain the cross-country heterogeneity of the duration of real exchange rate shocks. The definition of the real exchange rate is the real effective exchange rate (REER) based on consumer prices (line *rec*), for which 20 industrial and 70 developing countries were selected.[6]

The REER data for six representative countries are set out in Figure 1—an increase in the REER series indicates a real appreciation of the country's currency.[7] The data reveal that most countries have real exchange rates that exhibit drift or nonstationarity—that is, there are periods of sustained deviation from PPP. The evolution of REER appears to be a highly persistent process; for most countries, the REER does not appear to converge to any particular equilibrium value, with the possible exception of Sweden. Also, sharp movements in the REER during the

[6]For a detailed explanation and critique of how the Fund's REER indices are constructed, see Wickham (1993) and Zanello and Desruelle (1997).

[7]The 20 industrial countries and 70 developing countries are listed in Table 1. A decline (depreciation) in a country's REER index indicates a rise in its international competitiveness (defined as the relative price of domestic tradable goods in terms of foreign tradables).

Figure 1. Real Effective Exchange Rate, Selected Countries, 1973:3–2002:3 (1990 = 100)



Source: International Monetary Fund.

1980s and 1990s are a relatively frequent occurrence, especially for developing countries, such as Indonesia, Nigeria, and Mexico.

III. Empirical Results

In this section, we present our estimates of the persistence of parity deviations. We then examine the cross-country heterogeneity of the persistence results by grouping the countries using several key characteristics, such as income level, type of nominal exchange rate regime, type of dominant exportable, and geographic location.

Biased Least Squares Estimates of Half-Lives of Parity Reversion

The results for the half-life of the duration of shocks to the REER, calculated from the least squares estimates of α in ADF regression of equation (1), are set out in Table 1. Across all countries, the average (median) half-life of parity reversion is 4.04 years. This result is consistent with Rogoff's (1996) consensus of half-lives of parity reversion of between 36 and 60 months (three to five years), and with Cheung and Lai's (2000b) finding of median half-lives of 3.3 years for industrial countries in the post–Bretton Woods period.

Median-Unbiased Estimates of Half-Lives of Parity Reversion

The half-lives of PPP deviations calculated above (using the least squares estimator) are reasonably close to past studies but are likely to be biased downward and in favor of finding that PPP holds in the REER data. Consequently, we remove this bias by calculating median-unbiased estimates for the autoregressive parameter in equation (1).

The median-unbiased estimates of half-lives derived from the AR parameter in ADF regressions are reported in Table 1. In comparison with their least squares counterparts, they are typically much longer, ranging from as little as one month (Bolivia) to infinity (Belgium and Sweden, among others). Across all countries, the average (median) half-life of parity reversion is 8.17 years, in excess of the average least squares AR(p) half-life of 4.04 years.[8] This implies a rate of parity reversion of only 8 percent a year, rather than the 16 percent a year calculated using least squares.

Using the Andrews unbiased model-selection rule, we find that 50 of the countries are subject to finitely persistent shocks to their REER (which is consistent with the reversion of REER to parity), whereas 40 of the countries experience permanent shocks to their REER series. The interpretation of this rule is that for any given country, there is a 50 percent probability that the interval from zero to the estimated median-unbiased half-life contains the true half-life. There is also a 50 percent probability that the interval from the estimated median-unbiased half-life to infinity contains the true half-life. Let us take the examples of Iceland (short-lived half-life) and Togo (infinite half-life). For Iceland, while there is a 50 percent probability that the interval from zero to 1.1 years contains the true half-life, there is also a 50 percent probability that the interval from 1.1 years to infinity contains the true half-life. For Togo, while there is a 50 percent probability that the interval with a finite upper bound contains the true half-life, there is a 50 percent probability that the true half-life will be infinite (Table 1). Using the Andrews (1993) unbiased model-selection rule, the finite (Iceland) and infinite (Togo) estimates of the half-lives indicate that while shocks to Iceland's REER are transitory, shocks to Togo's REER are best viewed as being permanent.

[8]In calculating group and all-country median half-lives of deviation from parity, a permanent deviation is defined as one that is at least as long as the span of the data. In our sample, the data span 29 years, so permanent deviation from parity (infinite shock) is set to equal 30 years.

Table 1. Half-Lives of Reversion to Parity (years), Real Effective Exchange Rates, March 1973–March 2002, Advanced and Developing Countries

Country (1)	Biased Half-Life (2)	Unbiased Half-Life (3)	Time to Peak (4)	Time to Unity (5)
	ADVANCED COUNTRIES			
Belgium	7.00	∞	3.00	∞
Sweden	5.83	∞	0.33	∞
Portugal	4.92	8.42	0.92	3.17
Austria	3.92	8.25	0.83	3.25
Italy	3.67	8.17	1.00	3.17
Spain	3.58	8.17	0.75	3.25
Australia	3.67	8.00	0.08	0.67
Canada	5.42	8.00	1.25	4.50
Japan	5.00	8.00	1.00	4.17
United States	4.25	7.83	1.00	4.58
Ireland	2.75	7.50	0.25	0.83
Finland	3.75	5.50	1.33	3.50
Germany	3.25	5.50	0.83	2.50
Netherlands	3.42	5.50	0.83	2.58
United Kingdom	3.25	5.42	0.33	2.25
Korea, Rep. of	2.08	3.17	0.17	0.83
Switzerland	1.75	2.67	0.08	0.92
Norway	1.83	2.50	0.08	0.33
New Zealand	1.75	1.83	0.08	0.92
Iceland	1.00	1.08	0.08	0.08
Advanced countries median	3.63	7.67	0.79	
	AFRICA			
Burkina Faso	5.58	∞	0.08	NCU
Chad	10.92	∞	17.58	∞
Central African Republic	11.75	∞	0.08	NCU
Gabon	14.17	∞	3.67	∞
Lesotho	∞	∞	1.42	∞
Madagascar	8.33	∞	0.17	NCU
Mauritania	∞	∞	0.33	NCU
Mauritius	4.17	∞	0.17	NS
Morocco	9.67	∞	0.17	NS
Senegal	8.17	∞	0.50	NCU
Togo	6.42	∞	0.08	NCU
Tunisia	10.67	∞	2.08	∞
Niger	∞	∞	0.08	NCU
Gambia, The	1.17	∞	0.25	NCU
Ethiopia	4.83	∞	0.83	∞
Tanzania	6.33	14.17	1.17	8.42
Cameroon	3.08	7.25	0.25	S
Seychelles	2.67	6.83	0.17	0.25
Nigeria	4.25	5.58	1.50	3.83
Congo, Rep. of	1.42	4.83	0.08	S
Kenya	2.00	4.33	0.08	0.25
Sierra Leone	1.00	4.33	0.08	0.25

Table 1 *(continued)*

Country (1)	Biased Half-Life (2)	Unbiased Half-Life (3)	Time to Peak (4)	Time to Unity (5)
Côte d' Ivoire	2.17	3.25	0.25	S
Malawi	2.25	3.17	0.08	0.75
Ghana	1.75	2.75	0.08	0.17
Zambia	0.67	1.50	0.08	0.08
Uganda	1.00	1.25	0.08	S
Rwanda	0.50	1.17	0.08	0.17
Sudan	0.92	1.00	0.33	0.42
Congo, Dem. Rep. of	0.42	0.42	0.08	S
Africa median	4.21	22.08	0.17	
WESTERN HEMISPHERE				
Colombia	8.00	∞	3.25	∞
Dominican Republic	5.25	∞	0.58	NS
Ecuador	6.33	∞	0.92	∞
El Salvador	∞	∞	0.75	∞
Guatemala	5.42	∞	0.50	∞
Haiti	19.33	∞	0.08	NS
Honduras	5.17	∞	0.67	∞
Paraguay	4.50	∞	0.17	NS
Venezuela	6.25	∞	0.17	∞
Panama	10.08	∞	0.08	NCU
Peru	1.00	∞	0.17	0.33
Guyana	12.50	∞	∞	∞
Jamaica	4.42	8.33	1.25	3.50
Trinidad and Tobago	3.92	7.75	1.00	4.17
Uruguay	3.25	7.25	0.08	0.33
Costa Rica	3.08	3.83	0.83	1.33
Argentina	2.75	3.58	0.75	2.08
Barbados	2.58	3.50	0.17	0.92
Chile	2.50	2.75	0.92	1.17
Brazil	1.75	2.50	0.08	0.42
Mexico	1.75	1.92	0.50	0.50
Bolivia	0.08	0.08	0.08	S
Western Hemisphere median	4.46	∞	0.54	
ASIA, MIDDLE EAST, AND EUROPE				
Fiji	4.75	∞	0.08	NCU
Hungary	∞	∞	1.83	∞
India	17.67	∞	16.75	∞
Indonesia	11.92	∞	0.75	NS
Malaysia	8.25	∞	0.33	∞
Malta	7.08	∞	20.00	∞
Myanmar	∞	∞	∞	∞
Pakistan	∞	∞	2.00	∞
Papua New Guinea	1.83	∞	0.08	NCU
Thailand	6.17	∞	0.17	NS
Nepal	9.33	∞	0.17	NCU

Table 1 *(concluded)*

Country (1)	Biased Half-Life (2)	Unbiased Half-Life (3)	Time to Peak (4)	Time to Unity (5)
Philippines	3.50	8.08	0.42	2.17
Syrian Arab Republic	4.42	7.92	0.75	3.25
Samoa	3.33	7.83	0.92	S
Turkey	1.92	5.67	0.08	0.17
Iran, Islamic Republic of	3.08	5.42	0.67	0.92
Sri Lanka	2.50	5.17	0.17	1.50
Egypt	2.00	2.42	0.42	0.25
Asia, Middle East, and Europe median	5.46	∞	0.54	
All countries median	4.04	8.17	0.33	

Source: Authors' calculations.

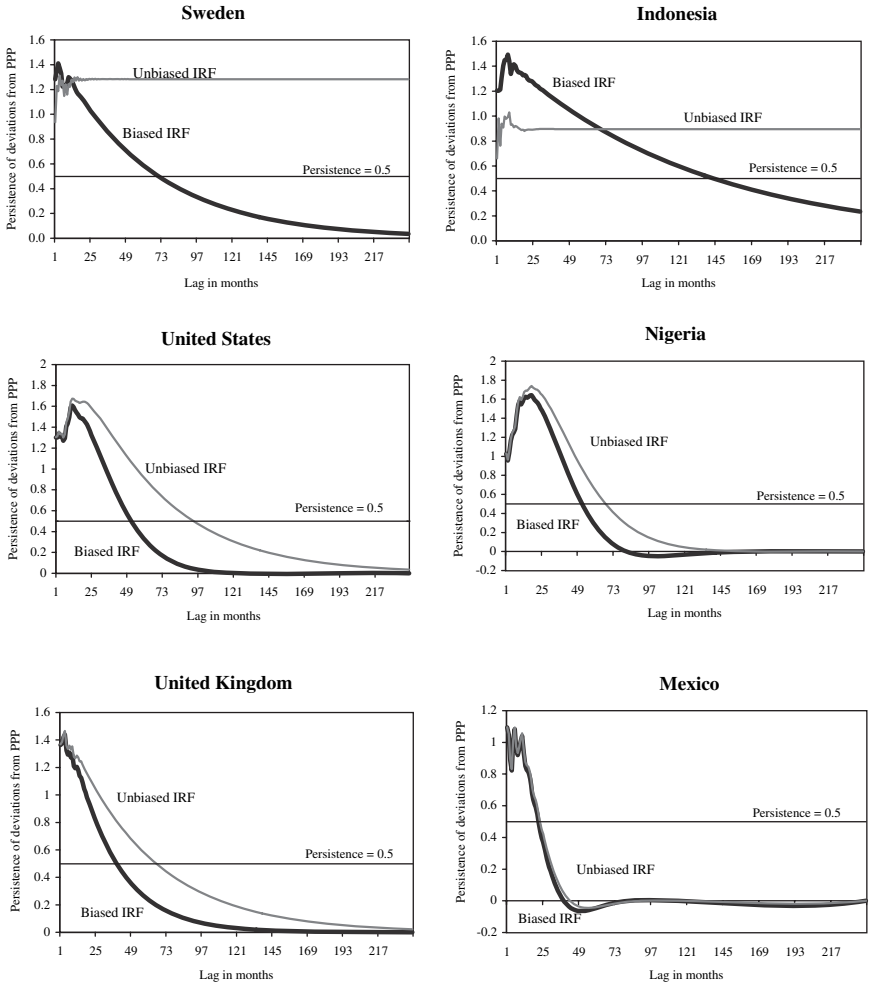
Notes: Column (2): Biased half-life of parity deviation, based on least squares estimation of the Augmented Dickey-Fuller regression of equation (1). The half-life for AR(*p*) models is calculated from the impulse-response functions (IRF) and is defined as the time taken for a unit impulse to dissipate permanently by one-half from the occurrence of the initial shock. Column (3): Median-unbiased half-life of parity deviation, based on median-unbiased estimation of the Augmented Dickey-Fuller regression of equation (1), as given by Andrews and Chen (1994). The half-life is as described for column (2) above. Column (4): "Time to peak" is the number of months following the unit shock to the real exchange rate that the IRF reaches its peak. Column (5): "Time to unity" is the number of months following the unit shock to the real exchange rate that the IRF crosses unity. NCU denotes non-stationary series—the IRF never crosses unity. NS denotes nonstationary series—the IRF is above unity, then crosses unity but does not decay. The symbol ∞ denotes classic nonstationary series—the IRF rises above unity and stays there. S denotes stationary series—the IRF never rises above unity and has a monotonic decay. In calculating group and all-country medians, infinity (∞) is set to equal 30 years.

We conclude that the majority of countries in our sample have real exchange rates that revert (albeit sometimes slowly) to PPP and, thus, that PPP holds in the post–Bretton Woods period. Our conclusion differs from that of some previous authors, because we use an unbiased model-selection rule that has a probability of correctly selecting the unit root model (when the true $\alpha = 1$) of about 0.5. This contrasts with the Monte Carlo experiments of Sarno and Taylor (2002), who show that standard method yields a probability of rejecting the null hypothesis of a unit root in the real exchange rate (when the rate is, in fact, mean-reverting) of only 0.11.

Impulse Response Functions and Half-Lives of Shocks to Parity

Figure 2 sets out the impulse-response functions (IRFs) for three country pairs: Sweden and Indonesia (which exhibit permanent shocks to their real exchange rate), the United States and Nigeria (which exhibit hump-shaped IRFs with finite half-lives), and the United Kingdom and Mexico (which exhibit monotonic decay to real exchange rate shocks and finite half-lives). The half-life can be read from the intersection of the relevant IRF with the line indicating persistence of 0.5. The

Figure 2. Impulse-Response Function (IRF) of a Shock to the Real Effective Exchange Rate, Selected Countries, 1973:3–2002:3



Source: Authors' calculations.

impact of high-frequency noise can be seen in the first few lags of the hump-shaped IRFs of the United States and Nigeria. The hump-shaped IRFs indicate that real exchange rate shocks are initially magnified. In both cases, the magnification of the real exchange rate shock can last up to one and a half years, with the IRF taking about four years to return to unity.[9] As pointed out by Cheung and Lai (2000b), this nonmonotonicity in the response of the IRF to shocks results in a sig-

[9]Following Cheung and Lai (2000b), the effects of nonmonotonic reversion to parity can also be illustrated by calculating the half-life of real exchange rate shocks *after* the period of initial magnification. For both the United States and Nigeria, these modified half-life estimates are much shorter than the standard half-life of parity reversion, indicating that once the initial magnification period is completed, the IRFs dissipate rather rapidly. The duration of postmagnification half-lives can be calculated from Table 1 as the difference between the unbiased half-life (column 3) and the time to peak (column 4).

nificant extension of the process of parity reversion (shock dissipation). Clearly, it is important to control for this kind of serial correlation in real exchange rates; while the AR(p) regressions can do so, standard AR(1) regressions assume that shocks decay monotonically and so would drastically underestimate the half-life of real exchange rate shocks.

There may be biases that offset the mean bias emanating from least squares estimation of shock persistence. Taylor (2001) contends there are two sources of *upward* bias in the estimation of the duration of exchange rate deviations from parity. The first is *temporal aggregation bias,* in which the use of low-frequency data does not allow the identification of high-frequency adjustment processes. The second is the *linear AR(1) specification* of unit root models, which assume monotonic reversion to parity. The use of monthly data should reduce the risk of temporal aggregation bias, while the use of an AR(p) model should remove the source of the second bias (see Cashin and McDermott, 2003).[10]

Imbs and others (2005) argue that the PPP puzzle is largely an artifact of aggregation across sectors, whereby sectoral heterogeneity in the duration of deviations from the law of one price can induce an upward bias in the speed of parity reversion of the real exchange rate derived using aggregate price indices. The Imbs and others (2005) result is to some extent controversial (see Chen and Engel, 2004), but sectoral aggregation could be a source of upward bias in the half-lives derived from both conventional and median-unbiased estimates of AR models.

In addition, the use of real effective (rather than bilateral) exchange rate data imparts another potential aggregation bias at the international level, since trade-weighted price levels are used in the calculation of a country's overall effective exchange rate. Such international aggregation could also be a source of upward bias to the estimated speed of parity reversion. However, previous work has found that the persistence properties of shocks to the real effective and bilateral exchange rate series are very similar (see McDermott, 1996).

Finally, Chen and Rogoff (2003) and Cashin, Céspedes, and Sahay (2004) point out that there are important real factors (such as real commodity export prices) that might be expected to affect the equilibrium real exchange rate of both advanced and developing countries. Both studies find that the half-life of the reversion of the real exchange rate to its (constant) long-run average level is much longer than the half-life of the reversion of the real exchange rate to its (time-varying) long-run equilibrium with real commodity prices. In this view, the long-run reversion of real exchange rates to purchasing power parity is a first approximation only (see Taylor, 2003).

Cross-Country Heterogeneity of Half-Lives of Real Exchange Rate Shocks

The results of our estimation of the median-unbiased half-life of reversion to PPP are shown in the histograms of Figures 3A to 3D. In our sample of 90 countries,

[10]The next step in the research agenda attempting to solve the PPP puzzle could lie in the development and implementation of median-unbiased nonlinear estimators of the speed of parity reversion.

there are 50 countries for which the half-life estimates range from 1 to 14 years, and 40 countries that experience permanent real exchange rate shocks (half-lives of 29 years or more). Examining industrial countries alone, the average (median) half-life is about eight years, which is twice as long as the downward-biased estimates of previous studies (Rogoff, 1996; and Cheung and Lai, 2000a).[11] In contrast, the half-life estimates for developing countries appear to be evenly spread, with most of the estimated deviations from parity being permanent. Accordingly, median-unbiased estimates of half-lives tend to be much longer for developing countries than for industrial countries. This finding of relatively slower parity reversion for developing countries contradicts the results of Cheung and Lai (2000a). However, it is consistent with Froot and Rogoff (1995), who expected that the rapid economic growth often associated with low-income countries would induce such drastic changes in the relative price of tradables and nontradables that the likelihood of parity reversion holding in any given country would rise with the level of income. Finally, the hypothesis that the real exchange rate is mean-reverting in developing countries receives some support—for 32 of 70 developing countries, the half-life of parity reversion is finite (Table 1). In contrast, the hypothesis of mean reversion of real exchange rates for developed countries receives much stronger support: 18 of 20 developed countries have real exchange rates that typically experience finite shocks.

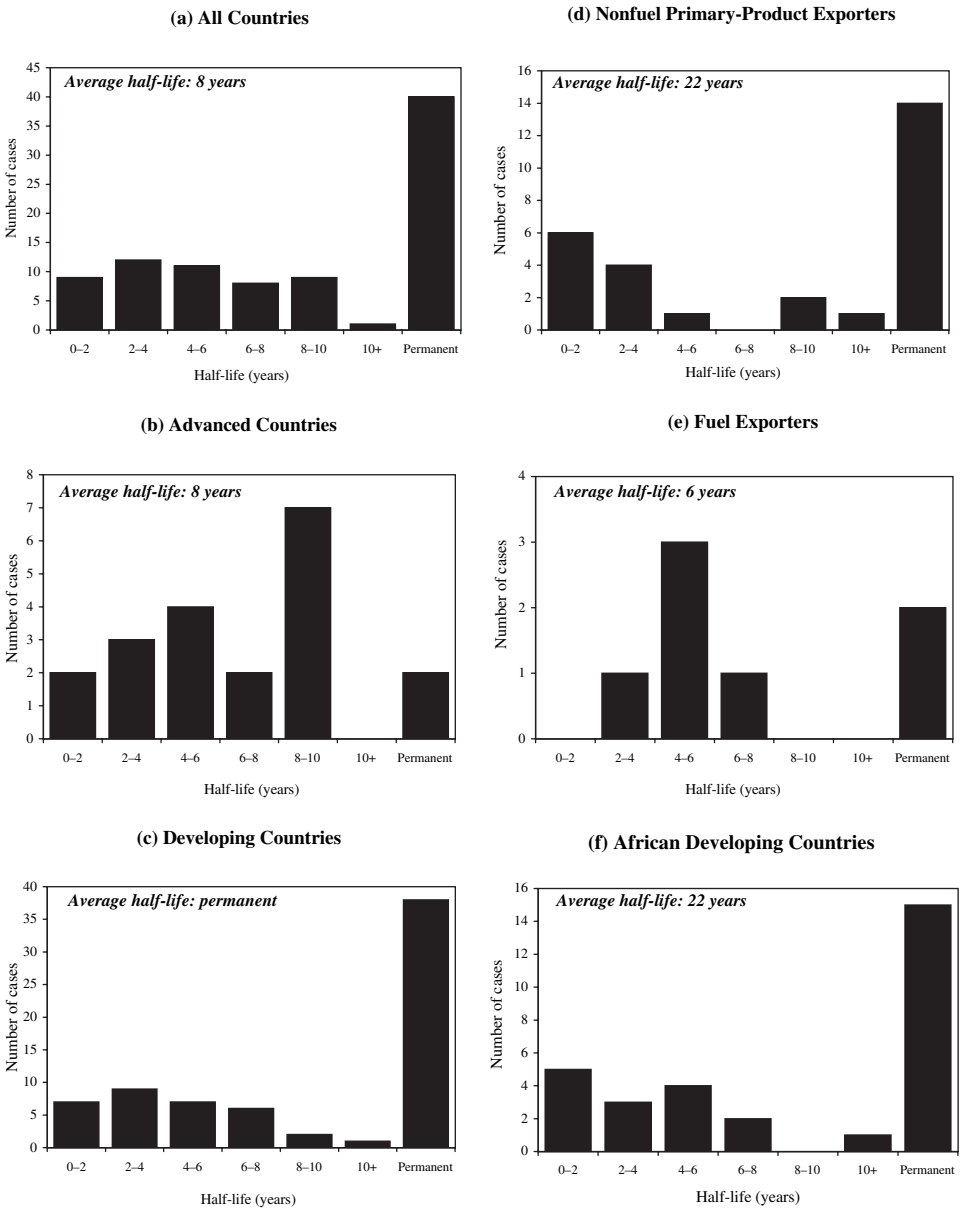
When countries are classified by their dominant exportable, the half-life estimates for (nonfuel) primary commodity exporters appear rather dispersed (Figure 3A). This is consistent with earlier findings on the heterogeneous persistence of terms of trade shocks affecting African commodity-exporting countries (see Cashin, McDermott, and Pattillo, 2004). The average half-life for nonfuel primary-product exporters is found to be about 22 years; while finite, this implies only a very slow speed of parity reversion.

The regional grouping of developing countries reveals that shock persistence of African and Western Hemisphere countries is evenly spread between finite and permanent shocks; in contrast, permanent shocks dominate in Asian countries (Figures 3A and 3B). The existing literature on parity reversion in developing countries has a sample-selection problem, as it has typically concentrated on analyses of high-inflation Latin American countries. Consistent with that literature, we find that parity deviations in the often-studied countries of the Western Hemisphere (especially Brazil, Mexico, Chile, and Argentina) are rather short-lived (all with half-lives of less than four years). Not surprisingly, the parity reversion half-lives for heavily indebted poor countries (HIPCs), which are mostly African, resemble the African results—the average half-life for HIPCs is found to be about 11 years.

Net debtor developing countries have also been grouped by source of external financing—for both private external financing and official external financing, real exchange rate shocks are typically permanent. Country groups of different income levels also exhibit a systematic pattern of differences in the persistence of real

[11]This result is also consistent with the median-unbiased half-lives calculated for industrial countries by Cashin and McDermott (2003).

Figure 3A. Frequency Distribution of Median-Unbiased Half-Lives of Deviations from Purchasing Power Parity, Country Groups, 1973–2002

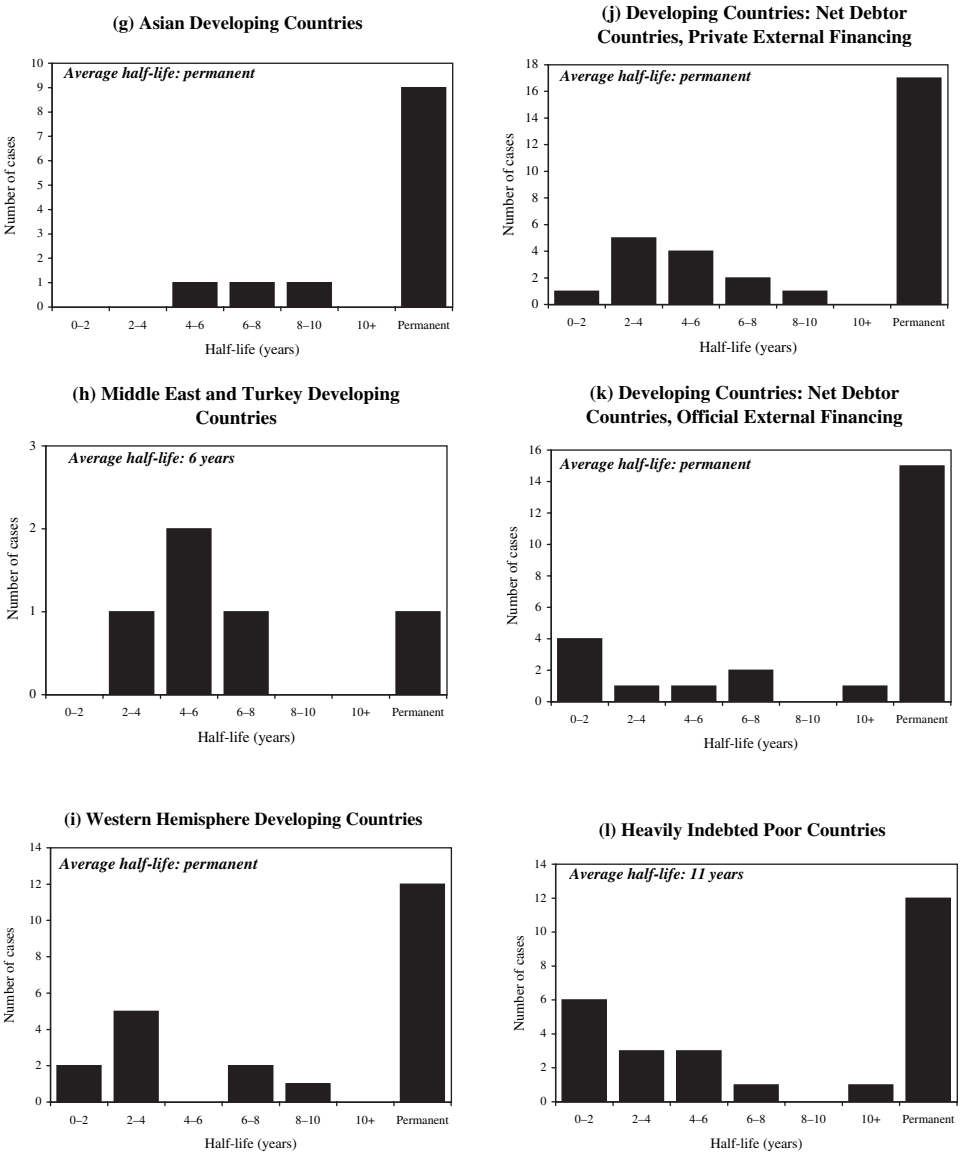


Source: Authors' calculations.

exchange rate shocks (Figure 3C)—our results tend to confirm previous work, finding an inverse relationship between income level and shock persistence (see Cheung and Lai, 2000a).

We also examine the persistence of parity deviations after classifying countries by type of nominal exchange rate regime. We classify country real exchange

Figure 3B. Frequency Distribution of Median-Unbiased Half-Lives of Deviations from Purchasing Power Parity, Country Groups, 1973–2002

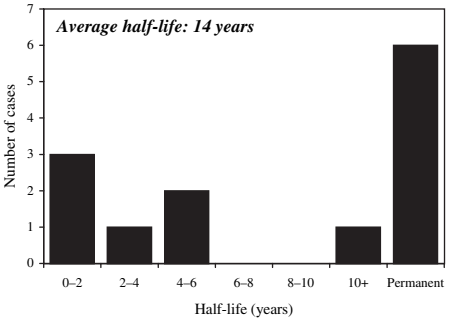


Source: Authors' calculations.

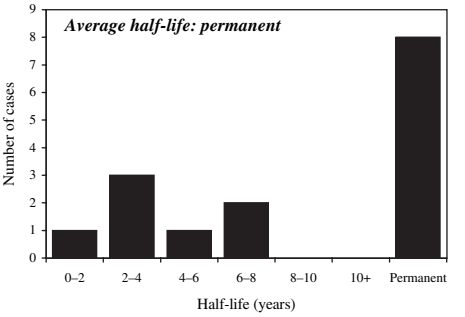
rates by type of nominal exchange rate regime, using (1) the IMF's (1998) de jure classification, which is based on the publicly stated commitment of the authorities of the country in question; and (2) the de facto classification of Reinhart and Rogoff (2002), which is based on the observed behavior of market-determined real exchange rates, including that of active parallel exchange rate markets. It is important to examine both de facto and de jure exchange rates because, through exchange

Figure 3C. Frequency Distribution of Median-Unbiased Half-Lives of Deviations from Purchasing Power Parity, Country Groups, 1973–2002

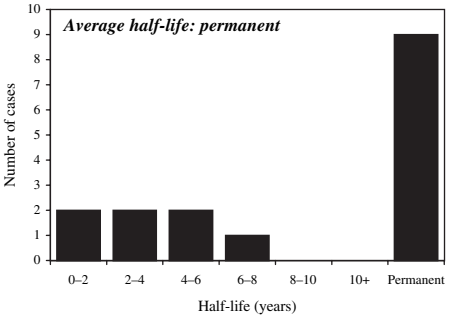
(m) Countries with Annual Income $102 – $280 (real GDP per capita, 1995 U.S. dollars)



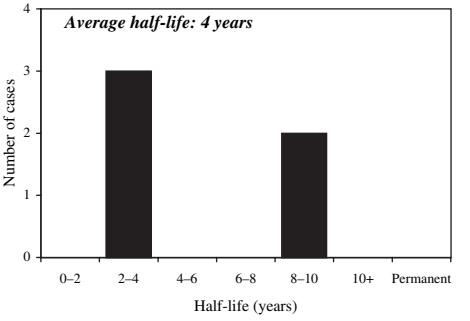
(p) Countries with Annual Income $2,112 – $5,792 (real GDP per capita, 1995 U.S. dollars)



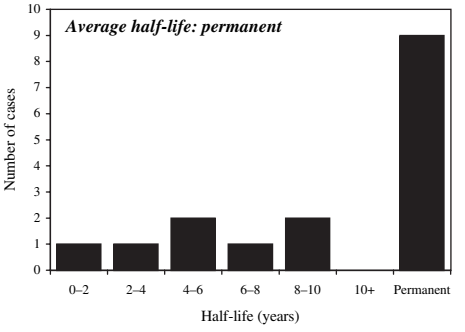
(n) Countries with Annual Income $281 – $769 (real GDP per capita, 1995 U.S. dollars)



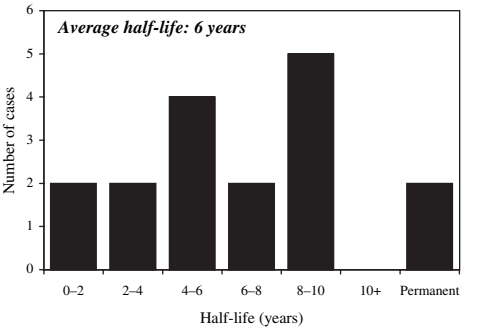
(q) Countries with Annual Income $5,793 – $15,891 (real GDP per capita, 1995 U.S. dollars)



(o) Countries with Annual Income $770 – $2,111 (real GDP per capita, 1995 U.S. dollars)



(r) Countries with Annual Income $15,892 – $43,600 (real GDP per capita, 1995 U.S. dollars)



Source: Authors' calculations.

market intervention or monetary policy, the authorities of a country can transform a de jure flexible exchange rate regime into a de facto pegged regime. Similarly, active parallel markets can transform de jure pegged official exchange rates into de facto flexible regimes.

When all 90 countries are categorized by the IMF's (1998) de jure exchange rate classification rules, it is clear that pegged exchange rate countries typically experience permanent deviations from parity, while more flexible exchange rate countries have much more dispersed half-lives of deviations from parity (Figure 3D). Very similar results are derived when exchange rates are classified using the Reinhart and Rogoff (2002) de facto exchange rate classification. The de facto classification reveals that while permanent real exchange rate shocks dominate the experience of pegged exchange rate countries, parity deviations for the majority of flexible exchange rate countries are finite and typically range between 2 and 10 years.

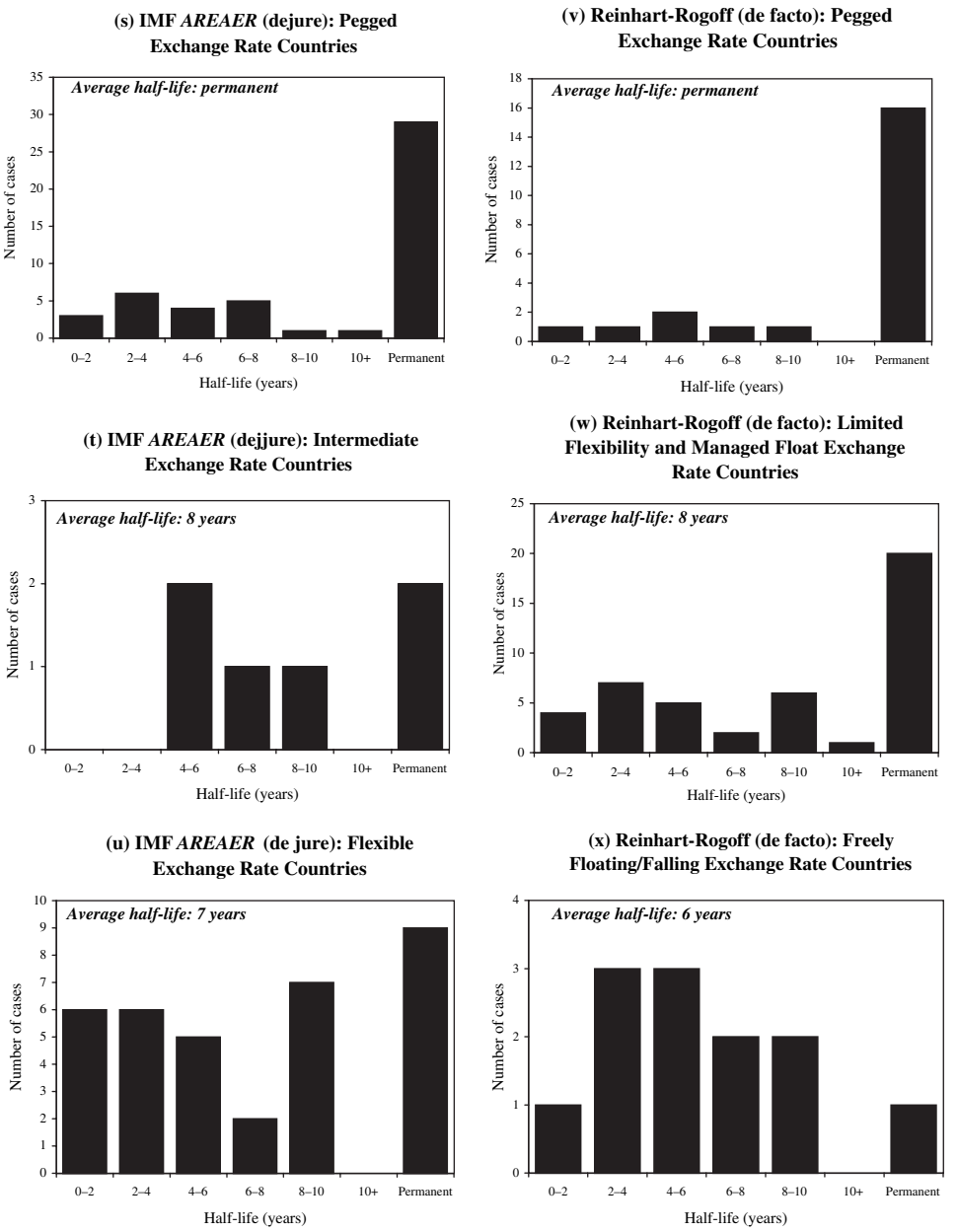
IV. What Is Causing the Heterogeneous Persistence of Deviations from Parity?

Underpinning the fundamental idea of long-run PPP is the notion that arbitrage in goods ensures that the parity condition is satisfied across a range of goods over a certain period. A key question is what might be causing the heterogeneous duration of exchange rate deviations from parity. In particular, in this section we will investigate the differences in the speed of parity reversion among advanced (industrial) and developing countries, and among countries with fixed and flexible nominal exchange rate regimes. Following Cheung and Lai (2000a), we examine whether the observed pattern of the persistence of real exchange rate shocks can be linked to any systematic differences in structural characteristics across countries.

Given that the usual distributional assumption of normality is not likely to hold for the distribution of either the half-lives of deviations from parity or the structural characteristics of countries, we use several nonparametric tests to examine the statistical significance of various hypotheses. These tests concern the equality of the average duration of parity deviations across country groups; the equality of the variability of the average duration of parity deviations across country groups; and whether the pattern of persistence of parity deviations is correlated with countries' structural characteristics.

First, we implement a nonparametric test of the equality of the median half-life of deviations from parity across country groups (the Wilcoxon-Mann-Whitney test). The results (listed in column 4 of Table 2) indicate that the null hypothesis of equal median half-lives is rejected, at the 5 percent significance level, for advanced versus developing countries. Further, the null hypothesis of equal median half-lives is rejected, at the 5 percent significance level, for pegged versus flexible exchange rate countries (classified under either the IMF (1998) or Reinhart and Rogoff (2002) classification scheme). Accordingly, we conclude that the average half-life of real exchange rate deviations from parity for advanced (industrial) countries is shorter in duration than the average for developing countries. In addition, the average half-life for countries with pegged nominal exchange rates is significantly longer in duration than that for countries with flexible nominal exchange rates.

Figure 3D. Frequency Distribution of Median-Unbiased Half-Lives of Deviations from Purchasing Power Parity, Country Groups, 1973–2002



Source: Authors' calculations.

When examining the persistence of real exchange rate shocks, the statistical significance of the difference of the group medians confirms the pattern shown in Figures 3A–3D—industrial countries display less persistence than developing

Table 2. Descriptive Statistics of Persistence of PPP Deviations, Country Groups

Country Group (1)	Average Persistence of Deviations from PPP (years) (2)	Number of Observations per Country Group (3)	H_0: Equality of Median Deviation from PPP (p-value) WMW test (4)	H_0: Equality of Variance of Deviation from PPP (p-value) BF test (5)
All countries	8.17	90		
Advanced	7.67	20		
Developing	30.00	70	2.18 (0.029)	6.11 (0.015)
IMF de jure classification:				
Peg	30.00	49		
Flexible	7.25	35	2.73 (0.006)	0.68 (0.413)
Reinhart-Rogoff de facto classification:				
Peg	30.00	22		
Free-floating/falling	5.58	12	3.22 (0.001)	0.67 (0.418)

Source: Authors' calculations.

Notes: Column (1): Country group; for definition and derivation, see Appendix II. Column (2): Persistence of deviations from PPP is calculated as the group average (median) half-life (in years) of deviations of the real exchange rate from parity. Column (3): Number of observations is the number of countries in each country group. Column (4): WMW is the Wilcoxon-Mann-Whitney test statistic of the null hypothesis of equality of the median half-life of deviations from PPP for each country group; the p-value is for the asymptotic normal approximation to the Wilcoxon t-statistic (see Conover, 1999). Column (5): BF is the Brown-Forsythe test statistic of the null hypothesis of equality of the variance of the median half-life of deviations from PPP for each country group. The F-statistic for the BF test has an approximate F-distribution with $G = 1$ numerator degrees of freedom and $N–G$ denominator degrees of freedom, under the null of equal variances in each group, where G is the number of groups and N the number of observations; the approximate p-value is given in parentheses.

countries, and countries with fixed nominal exchange rate regimes display more persistence than countries with flexible nominal exchange rate regimes.

Second, we examine whether the differences in the sample variances of the half-lives across country groups are statistically significant by implementing the Brown-Forsythe test. The results (listed in column 5 of Table 2) indicate that the null hypothesis of equal variances of the half-lives is rejected, at the 5 percent significance level, for the industrial countries in comparison with the developing countries. Interestingly, the null of equality of the variances of the half-lives across pegged and flexible exchange rate regimes cannot be rejected. Accordingly, there is evidence that the duration of half-lives of parity deviations for developing countries is more variable than the duration of half-lives of parity deviations for industrial countries.

Third, we use our cross-country data to examine whether each country's persistence of deviations from parity is correlated with its structural characteristics. As normality is unlikely to hold here, we again use a nonparametric test—the

Spearman rank correlation test (see Cheung and Lai (2000a) and Conover (1999) for details). The Spearman rank correlation test measures whether there is a significant relationship between the persistence of parity deviations and key structural characteristics (such as cross-country differentials in inflation and trade openness). The null hypothesis of the Spearman rank correlation test is that there is no rank correlation between the persistence of parity deviations and differentials in each country's structural characteristics. In explaining empirical cross-country heterogeneity in the duration of real exchange rate deviations from parity, we will evaluate several fundamental characteristics (see Appendix I for details), including cross-country differentials in inflation (INF), nominal exchange rate arrangements (VOFFER and VPARER), openness to trade (TGDP), productivity growth (PCGDP), and the share of government spending in the economy (GGDP).

Inflation

If national price movements were dominated by nominal (monetary) shocks, then parity reversion would be expected to be rather fast. Given the presence of nominal rigidities, a higher inflation rate may bring about more frequent adjustment of goods prices and shrink the duration of deviations from parity. Indeed, previous work has indicated that PPP typically holds for high-inflation countries (Frenkel, 1978; and McNown and Wallace, 1989). To analyze whether relative inflation is related to cross-country differences in persistence of parity deviations, we construct the average (median) inflation rate over the sample period (INF) for each country. The rank correlation coefficient between the half-life of the real exchange rate and the rate of inflation is –0.281, with an approximate *p*-value of 0.007 (Table 3). Using the critical values of Zar (1972), the null of no rank correlation is decisively rejected, with the rank correlation being negative and statistically significant at the 1 percent level. This indicates that, across advanced and developing countries, there is an inverse relationship between inflation and the persistence of real exchange rate shocks. That is, countries with higher inflation rates tend to have shorter lived deviations of their real exchange rates from PPP, which suggests that parity reversion is fast when price movements largely reflect monetary shocks.

Nominal Exchange Rate Volatility

Mussa (1986) stressed that exchange rates behaved differently under alternative exchange rate regimes, finding that the post–Bretton Woods float of major currencies had induced large real exchange rate variability in many industrial countries. Greater flexibility in nominal exchange rates would be expected to increase the speed of parity reversion of real exchange rates by encouraging more frequent adjustment of goods prices. However, as noted above, it is important to measure the variability of nominal exchange rates using both the official exchange rate (VOFFER) and the exchange rate determined in parallel markets (VPARER). In doing so, we use both the official and parallel exchange rate market data of Reinhart and Rogoff (2002) for the period 1973–98, with variability measured as the standard deviation of the (monthly) rate of change of the (log) series.

Table 3. Rank Correlation of Half-Life of Deviation from Purchasing Power Parity with Country Characteristics, All Countries

Country Characteristic (1)	Rank Correlation of Characteristic with Half-Life of Parity Deviations (2)	Approximate *p*-Value (number of observations) (3)
Rate of inflation	–0.281*	0.007 (90)
Volatility of the official nominal exchange rate	–0.379*	0.002 (70)
Volatility of the parallel-market nominal exchange rate	–0.282*	0.02 (70)
Productivity growth	–0.038	0.77 (90)
Government spending	–0.073	0.50 (88)
Trade openness	0.180	0.10 (89)

Source: Authors' calculations.

Notes: Column (1): Country characteristic; for definition and derivation, see Appendix I. Column (2): Spearman rank correlation coefficient—the null hypothesis of the Spearman test is that there is no correlation between the (tie-adjusted) rank of each country's bias-corrected half-life of parity deviation and the rank of its period-average value of the country characteristic. Persistence of deviations from PPP is calculated as the median-unbiased half-life (in years) of real exchange rate deviations from parity (listed in column 3 of Table 1). Column (3): The approximate *p*-value is taken from Zar (1972). An asterisk (*) denotes that the null hypothesis of no rank correlation is rejected at the 5 percent level of significance. The number of country observations is given in parentheses.

The rank correlation coefficient between the half-life of the real exchange rate and the variability of the official nominal exchange rate is –0.379, with an approximate *p*-value of 0.002 (Table 3). Similarly, the rank correlation coefficient between the half-life of the real exchange rate and the variability of the parallel-market nominal exchange rate is –0.282, with an approximate *p*-value of 0.02. Both correlations are negative and statistically significant at the 1 percent level. This indicates that, across advanced and developing countries, there is an inverse relationship between nominal exchange rate variability and the persistence of real exchange rate shocks; that is, countries with more variable nominal exchange rates tend to have shorter lived deviations of their real exchange rates from PPP. These results are broadly consistent with the conclusions of Goldfajn and Valdés (1999), who find that overvaluation of real exchange rates was typically corrected by changes in the nominal exchange rate rather than changes in inflation differentials.

Productivity Growth and Government Spending

Productivity growth is a supply-side factor that can affect the persistence of real exchange rate shocks. The Balassa-Samuelson effect has traditionally been the most popular explanation for the persistence of parity deviations. The Balassa-Samuelson hypothesis highlights the potential effects of differential productivity growth (favoring the traded goods sector) on the behavior of real exchange rates, which ultimately raises the relative price of nontraded goods. Importantly, this

traded-goods productivity bias is deemed to rise with the wealth of the country. Following Balassa (1964), we proxy for productivity growth by using the rate of growth of per capita real GDP; accordingly, we construct the average (median) per capita growth rate over the sample period for each country (PCGDP).

In the presence of capital and labor that are mobile across sectors in the long run but not the short run, government spending is a demand-side factor that is also hypothesized to affect the speed of parity reversion by producing a stronger home-goods bias. Froot and Rogoff (1991) note that government spending typically falls more heavily on nontraded goods, thereby bidding up their price relative to the price of tradables. Bergin and Feenstra (2001) suggest that real exchange rate persistence rises with the share of produced goods that are nontraded. Accordingly, we construct the average (median) for government spending as a share of GDP over the sample period for each country (GGDP).

The rank correlation coefficient between the half-life of the real exchange rate and the growth of per capita GDP is –0.038, with an approximate *p*-value of 0.77 (Table 3). Similarly, the rank correlation coefficient between the half-life of the real exchange rate and government spending as a share of GDP is –0.073, with an approximate *p*-value of 0.50. Neither rank correlation is significantly different from zero. We thus conclude that there is little evidence that either higher productivity growth or greater government spending can explain much of the observed pattern of deviation from PPP across countries. Our conclusions are consistent with those of Rogoff (1996).

Trade Openness

Rogoff, Froot, and Kim (2001) highlight the effect of goods market arbitrage, which, if operative, can accelerate the speed of parity reversion across goods. Greater trade flows should, in principle, promote goods market arbitrage, encourage more frequent price adjustment by firms, and reduce the persistence of real exchange rate shocks (Faruqee, 1995). As a measure of trade openness, we construct for each country over the sample period the average (median) ratio of exports plus imports to GDP (TGDP). The rank correlation coefficient between the half-life of the real exchange rate and external trade as a share of GDP is 0.180, with an approximate *p*-value of 0.10 (Table 3). This rank correlation differs insignificantly from zero. Accordingly, there is little evidence in favor of the goods arbitrage view of PPP—differences in trade openness are not associated with the persistence of parity deviations across countries.[12]

To summarize, we find that cross-country differences in inflation rates and nominal exchange rate variability have a strong and statistically significant inverse relationship with the observed pattern of the persistence of deviations from purchasing

[12]Cheung, Chinn, and Fujii (2001) also find a nonnegative relationship between openness and the persistence of sectoral real exchange rate deviations. They attribute this result to the inverse relationship between openness and inflation (see Romer, 1993). For our sample of countries, we also find a strong negative rank correlation between openness and inflation (–0.397). This suggests that greater openness is associated with lower inflation, which implies a slower speed of parity reversion.

power parity.[13] Indeed, to the extent that the existing literature on measuring parity reversion in developing countries has largely concentrated on the Latin American experience (which is dominated by high-inflation countries), this sample-selection bias can account for the erroneous received wisdom that parity reversion in developing countries is faster than that observed for developed countries. However, once a wider set of developing countries is analyzed, parity reversion in developing countries is typically slower than that observed for developed countries.

V. Conclusion

The validity of purchasing power parity (PPP)—the notion that prices in different countries move toward equality in common currency terms—is of interest to policymakers for two main reasons. First, PPP provides a long-term benchmark for the equilibrium value of exchange rates and, as such, is a criterion for evaluating the competitiveness of real exchange rates. Second, PPP has been adopted as a central building block of many theories of exchange rate determination; the quality of policy advice based on these theories may depend on the validity of PPP.

This paper has reexamined whether PPP holds during the post–Bretton Woods period by investigating the time series properties of the real effective exchange rate of 90 advanced and developing countries. The post-1973 revival of flexible exchange rates spawned a great interest in the empirical relevance of the PPP theory of real exchange rate determination. Previous studies of PPP reversion largely focused on developed countries, and univariate studies of the hypothesis of unit roots in real exchange rates yielded consensus estimates of the half-life of deviations of real exchange rates from PPP of about four years (Rogoff, 1996).

Using least squares estimation of unit root models, we replicate the consensus finding in the literature. However, using median-unbiased estimation techniques that remove the downward bias of least squares, we find that the half-lives of parity reversion are much longer than the consensus estimate, with a cross-country average of bias-corrected half-lives of about eight years. Our results confirm Rogoff's (1996) "PPP puzzle"—that while PPP holds for the majority of countries, the reversion of real exchange rates to parity is, in many cases, rather slow.

Using the median-unbiased estimates of the half-lives of deviations from parity and the Andrews (1993) unbiased model-selection rule, we conclude that 50 of the 90 countries in our sample have finite half-lives of parity reversion, which indicates that there is a better than even chance that shocks to their real exchange rates are transitory. For these 50 countries, we can conclude that there is reversion of real exchange rates to parity and that PPP holds in the post–Bretton Woods period. Conversely, there is little evidence of parity reversion for the remaining 40 countries in our sample, where shocks to the real exchange rate are best viewed as being permanent.

[13]As a robustness check on the unconditional correlation results, least squares regression of the half-life of parity deviation on the above-mentioned structural characteristics was also carried out. As in the bivariate correlations, inflation is the most significant variable for reducing the half-life of parity deviations.

Our analysis yields evidence of significant heterogeneity of parity reversion across countries and across groups of countries. In our view, the general relevance of parity reversion has been exaggerated by the predilection of studies to focus on the reversion experience of developed countries. In addition, the few studies of parity reversion for developing countries have suffered from sample-selection bias, in that they typically examined the speed of reversion in high-inflation Latin American countries. We find that parity reversion is more likely to be found for developed countries than for developing countries. Parity reversion is also more likely to be found for countries with flexible nominal exchange rate regimes than for countries with fixed nominal exchange rate regimes. Finally, when we examine the determinants of the observed cross-country heterogeneity in the persistence of reversion of real exchange rates to parity, we find that parity reversion tends to be faster in high-inflation countries than in low-inflation countries and slower in countries with less nominal exchange rate variability.

APPENDIX I

Description of the Data

The primary data sources are the IMF's *International Financial Statistics* (*IFS*) and Information Notice System (INS), Reinhart and Rogoff (2002), and the World Bank's *World Development Indicators* (2002). Below we provide a description of each series used in the paper. The 90 developing and developed countries in the full sample are listed in Table 1. Unless otherwise noted, the data are for the sample period 1973–2002.

REER: The real effective exchange rate data are of monthly frequency, for the period March 1973–March 2002; a total of 358 observations. REER is the trade-weighted measure of the seasonally adjusted, CPI-based real effective exchange rate (base 1990=100); obtained from the IMF's INS.

INF: The rate of change of consumer prices (percent a year); period average of annual data 1973–2002; obtained from the IMF's *IFS.*

GGDP: General government final consumption spending as a share of GDP; period average of annual data 1973–2002; obtained from World Bank (2002).

PCGDP: Growth of per capita GDP (percent per annum); period average of annual data, 1973–2002; obtained from World Bank (2002).

TGDP: Exports and imports of goods and services (valued in current U.S. dollars) as a share of GDP (valued in current U.S. dollars); period average of annual data, 1973–2002; obtained from World Bank (2002).

VPARER: Volatility of the parallel market exchange rate; measured as the standard deviation of the (monthly) rate of change of the parallel market exchange rate, with the exchange rate measured as the logarithm of the parallel nominal exchange rate (local currency per U.S. dollar), March 1973 to December 1998; obtained from Reinhart and Rogoff (2002).

VOFFER: Volatility of the official exchange rate; measured as the standard deviation of the (monthly) rate of change of the official exchange rate, with the exchange rate measured as the logarithm of the official nominal exchange rate (local currency per U.S. dollar), March 1973 to December 1998; obtained from Reinhart and Rogoff (2002).

Nominal Exchange Rate Regime

The IMF's de jure classification, used between 1975 and 1997 in its *Annual Report on Exchange Arrangements and Exchange Restrictions* (*AREAER*), consisted of 10 categories grouped as follows: regimes 1–5 are defined as fixed pegs; regimes 6–7 (limited flexibility with respect to a single currency, cooperative arrangements) are intermediate; and regimes 8–10 (including managed floating and independently floating) are flexible arrangements. For each country, the de jure exchange rate regime classification is the mode of the *AREAER* classification numbers over the period 1975–98. See Bubula and Otker-Robe (2002) for details.

The Reinhart and Rogoff (2002) de facto classification describes exchange rate regimes as: (1) de facto pegs (including no separate legal tender and currency boards), denoted as regime 1; (2) limited flexibility (including crawling pegs and narrow crawling bands), denoted as regime 2; (3) managed floating (including wider crawling bands), denoted as regime 3; (4) freely floating, denoted as regime 4; and (5) freely falling (where the annualized rate of inflation exceeds 40 percent), denoted as regime 5. For each country, the de facto exchange rate regime classification is the mode of the annual Reinhart and Rogoff (2002) classification numbers over the period 1973–98. The following countries had no Reinhart and Rogoff (2002) classification data: Barbados, the Democratic Republic of the Congo, Ethiopia, Fiji, Papua New Guinea, Rwanda, Samoa, Seychelles, Sierra Leone, Sudan, and Trinidad and Tobago.

APPENDIX II

Country Group Classifications

The 90 developed and developing countries in our sample (listed in Table 1) have been classified into various country groups in order to undertake cross-country comparisons of the persistence of parity deviations. The major sources of classification were the IMF's *World Economic Outlook* (2002, 2000); the IMF's *Annual Report on Exchange Arrangements and Exchange Restrictions* (*AREAER*), various issues; the World Bank's *World Development Indicators* database (2002); Andrews and others (1999); and Reinhart and Rogoff (2002). The country groups, along with the country members and classification rule, follow.

The IMF's *World Economic Outlook* (2002) classifies countries into groups on the basis of certain criteria. For nondeveloping countries, the groups are *advanced economies* and *countries in transition.* Developing countries are classified by their predominant export. *Primary product exporters* are those countries whose exports of agricultural and mineral primary products (Standard Industrial Trade Classification (SITC) 0, 1, 2, 4, 68) accounted for at least 50 percent of their total export earnings during 1994–98. *Fuel exporters* are those countries whose exports of fuel products (SITC 3) accounted for at least 50 percent of their total export earnings during 1994–98. *Net debtor countries* are defined as developing countries with negative external assets at the end of 1998. Net debtor countries are then differentiated by their main source of external financing: Those with official financing (including official grants) accounting for more than two-thirds of their total 1994–98 external financing are classified as *official external financing countries;* and those with private financing (including direct and portfolio investment) accounting for more than two-thirds of their total 1994–98 external financing are classified as *private external financing countries.*

Advanced economies: Australia, Austria, Belgium, Canada, Finland, Germany, Iceland, Ireland, Italy, Japan, Korea, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States (IMF (2002) classification).

Developing countries: Argentina, Barbados, Bolivia, Brazil, Burkina Faso, Cameroon, Central African Republic, Chad, Chile, Colombia, Democratic Republic of the Congo, Republic of

Congo, Côte d'Ivoire, Costa Rica, Dominican Republic, Ecuador, Egypt, El Salvador, Ethiopia, Fiji, Gabon, The Gambia, Ghana, Guatemala, Guyana, Haiti, Honduras, Hungary, India, Indonesia, Islamic Republic of Iran, Jamaica, Kenya, Lesotho, Madagascar, Malawi, Malaysia, Malta, Mauritania, Mauritius, Mexico, Morocco, Myanmar, Nepal, Niger, Nigeria, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Rwanda, Samoa, Senegal, Seychelles, Sierra Leone, Sri Lanka, Sudan, Syria, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, Uruguay, República Bolivariana de Venezuela, and Zambia (IMF (2002) classification).

Nonfuel primary-product exporting countries: Australia, Bolivia, Burkina Faso, Canada, Central African Republic, Chad, Chile, Democratic Republic of the Congo, Cote d'Ivoire, The Gambia, Ghana, Guyana, Honduras, Iceland, Madagascar, Malawi, Mauritania, Myanmar, New Zealand, Niger, Papua New Guinea, Paraguay, Peru, Sudan, Tanzania, Togo, and Zambia (IMF (2002) classification).

Fuel-exporting countries: Republic of Congo, Gabon, Islamic Republic of Iran, Nigeria, Norway, Trinidad and Tobago, and República Bolivariana de Venezuela (IMF (2002) classification).

African developing countries: Burkina Faso, Cameroon, Central African Republic, Chad, Democratic Republic of the Congo, Republic of Congo, Côte d'Ivoire, Ethiopia, Gabon, The Gambia, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mauritania, Mauritius, Morocco, Niger, Nigeria, Rwanda, Senegal, Seychelles, Sierra Leone, Sudan, Tanzania, Togo, Tunisia, Uganda, and Zambia (IMF (2002) classification).

Asian developing countries: Fiji, India, Indonesia, Malaysia, Myanmar, Nepal, Pakistan, Papua New Guinea, Philippines, Samoa, Sri Lanka, and Thailand (IMF (2002) classification).

Western Hemisphere developing countries: Argentina, Barbados, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Panama, Paraguay, Peru, Trinidad and Tobago, Uruguay, and República Bolivariana de Venezuela (IMF (2002) classification).

Middle East and Turkey developing countries: Egypt, Islamic Republic of Iran, Malta, Syria, and Turkey (IMF (2002) classification).

Heavily indebted poor countries (HIPCs): Comprises those countries (except Nigeria) considered by the IMF and World Bank for their HIPC debt initiative—Bolivia, Burkina Faso, Cameroon, Central African Republic, Chad, Democratic Republic of the Congo, Republic of Congo, Côte d'Ivoire, Ethiopia, Ghana, Guyana, Honduras, Kenya, Madagascar, Malawi, Mauritania, Myanmar, Niger, Rwanda, Senegal, Sierra Leone, Sudan, Tanzania, Togo, Uganda, and Zambia (IMF (2002) classification; see Andrews and others, (1999)).

Developing countries—net debtor countries, private external financing: Argentina, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Egypt, Fiji, Guatemala, India, Indonesia, Islamic Republic of Iran, Jamaica, Kenya, Lesotho, Malaysia, Malta, Mexico, Morocco, Myanmar, Panama, Paraguay, Peru, Seychelles, Sierra Leone, Thailand, Trinidad and Tobago, Turkey, and República Bolivariana de Venezuela (IMF (2000) classification).

Developing countries—net debtor countries, official external financing: Burkina Faso, Cameroon, Central African Republic, Chad, Democratic Republic of the Congo, Republic of Congo, Ethiopia, Gabon, The Gambia, Guatemala, Guyana, Haiti, Madagascar, Malawi, Mauritania, Nepal, Niger, Rwanda, Samoa, Senegal, Tanzania, Togo, Uganda, and Zambia (IMF (2000) classification).

Countries with annual income (real GDP per capita in U.S. dollars in 1995) of $102–$280: Burkina Faso, Chad, Democratic Republic of the Congo, Ethiopia, Madagascar, Malawi, Nepal, Niger, Nigeria, Rwanda, Sierra Leone, Sudan, and Tanzania (World Bank (2002) classification).

Countries with annual income (real GDP per capita in U.S. dollars in 1995) of $281–$769: Cameroon, Central African Republic, Côte d'Ivoire, The Gambia, Ghana, Haiti, Honduras, India, Kenya, Mauritania, Pakistan, Senegal, Sri Lanka, Togo, Uganda, and Zambia (World Bank (2002) classification).

Countries with annual income (real GDP per capita in U.S. dollars in 1995) of $770–$2,111: Bolivia, Republic of Congo, Dominican Republic, Ecuador, Egypt, El Salvador, Guatemala, Indonesia, Islamic Republic of Iran, Jamaica, Morocco, Papua New Guinea, Paraguay, Philippines, Syria, and Tunisia (World Bank (2002) classification).

Countries with annual income (real GDP per capita in U.S. dollars in 1995) of $2,112–$5,792: Brazil, Chile, Colombia, Costa Rica, Gabon, Hungary, Malaysia, Mauritius, Mexico, Peru, Thailand, Trinidad and Tobago, Turkey, Uruguay, and República Bolivariana de Venezuela (World Bank (2002) classification).

Countries with annual income (real GDP per capita in U.S. dollars in 1995) of $5,793–$15,891: Argentina, Barbados, Korea, Portugal, and Spain (World Bank (2002) classification).

Countries with annual income (real GDP per capita in U.S. dollars in 1995) of $15,892–$43,600: Australia, Austria, Belgium, Canada, Finland, Germany, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, and United States (World Bank (2002) classification).

IMF pegged exchange rate countries: Argentina, Austria, Barbados, Burkina Faso, Cameroon, Central African Republic, Chad, Republic of Congo, Côte d'Ivoire, Dominican Republic, Egypt, El Salvador, Ethiopia, Fiji, Finland, Gabon, Guatemala, Guyana, Haiti, Honduras, Hungary, Islamic Republic of Iran, Kenya, Lesotho, Malawi, Malaysia, Malta, Mauritania, Mauritius, Morocco, Myanmar, Niger, Norway, Panama, Papua New Guinea, Paraguay, Rwanda, Samoa, Senegal, Seychelles, Sudan, Sweden, Syria, Tanzania, Thailand, Togo, Trinidad and Tobago, República Bolivariana de Venezuela, and Zambia (*AREAER* classification; see Appendix I).

IMF intermediate exchange rate countries: Belgium, Germany, Ireland, Italy, Nepal, and Netherlands (*AREAER* classification; see Appendix I).

IMF flexible exchange rate countries: Australia, Bolivia, Brazil, Canada, Chile, Colombia, Democratic Republic of the Congo, Costa Rica, Ecuador, The Gambia, Ghana, Iceland, India, Indonesia, Jamaica, Japan, Korea, Madagascar, Mexico, New Zealand, Nigeria, Pakistan, Peru, Philippines, Portugal, Sierra Leone, Spain, Sri Lanka, Switzerland, Tunisia, Turkey, Uganda, United Kingdom, United States, and Uruguay (*AREAER* classification; see Appendix I).

Reinhart and Rogoff pegged exchange rate countries: Austria, Belgium, Burkina Faso, Cameroon, Central African Republic, Chad, Côte d'Ivoire, Ecuador, El Salvador, Gabon, Guatemala, Haiti, Kenya, Lesotho, Mexico, Netherlands, Niger, Panama, Senegal, Thailand, Togo, and República Bolivariana de Venezuela (Reinhart and Rogoff (2002) classification; see Appendix I).

Reinhart and Rogoff limited flexibility and managed float exchange rate countries: Bolivia, Canada, Chile, Colombia, Costa Rica, Dominican Republic, Egypt, Finland, The

Gambia, Guyana, Honduras, Hungary, Iceland, India, Indonesia, Islamic Republic of Iran, Ireland, Italy, Jamaica, Korea, Madagascar, Malawi, Malaysia, Malta, Mauritania, Mauritius, Morocco, Myanmar, Nepal, New Zealand, Nigeria, Norway, Pakistan, Paraguay, Philippines, Portugal, Spain, Sri Lanka, Sweden, Switzerland, Syria, Tanzania, Tunisia, Uganda, and United Kingdom (Reinhart and Rogoff (2002) classification; see Appendix I).

Reinhart and Rogoff freely floating/falling exchange rate countries: Argentina, Australia, Brazil, Republic of Congo, Germany, Ghana, Japan, Peru, Turkey, United States, Uruguay, and Zambia (Reinhart and Rogoff (2002) classification; see Appendix I).

REFERENCES

Andrews, David, Anthony Boote, Syed Rizavi, and Sukwinder Singh, 1999, *Debt Relief for Low-Income Countries: The Enhanced HIPC Initiative,* IMF Pamphlet Series, No. 51 (Washington: International Monetary Fund).

Andrews, Donald, 1993, "Exactly Median-Unbiased Estimation of First Order Autoregressive/ Unit Root Models," *Econometrica,* Vol. 61 (1), pp. 139–66.

———, and Hong-Yuan Chen, 1994, "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics,* Vol. 12 (April), pp. 187–204.

Bahmani-Oskooee, M., 1995, "Real and Nominal Effective Exchange Rates for 22 LDCs: 1971:1–1990:4," *Applied Economics,* Vol. 27 (July), pp. 591–604.

Balassa, Bela, 1964, "The Purchasing Power Parity Doctrine: A Reappraisal," *Journal of Political Economy,* Vol. 72, pp. 584–96.

Bergin, Paul, and Robert Feenstra, 2001, "Pricing-to-Market, Staggered Contracts, and Real Exchange Rate Persistence," *Journal of International Economics,* Vol. 54 (August), pp. 333–59.

Bubula, Andrea, and Inci Otker-Robe, 2002, "The Evolution of Exchange Rate Regimes Since 1990: Evidence from De Facto Policies," IMF Working Paper 02/155 (Washington: International Monetary Fund).

Calvo, Guillermo, Carmen Reinhart, and Carlos Végh, 1995, "Targeting the Real Exchange Rate: Theory and Evidence," *Journal of Development Economics,* Vol. 47 (June), pp. 97–133.

Cashin, Paul, Luis Céspedes, and Ratna Sahay, 2004, "Commodity Currencies and the Real Exchange Rate," *Journal of Development Economics,* Vol. 75 (October), pp. 239–68.

Cashin, Paul, Hong Liang, and C. John McDermott, 2000, "How Persistent Are Shocks to World Commodity Prices?" *IMF Staff Papers,* Vol. 47 (December), pp. 177–217.

Cashin, Paul, and C. John McDermott, 2003, "An Unbiased Appraisal of Purchasing Power Parity," *IMF Staff Papers,* Vol. 50 (September), pp. 321–51.

———, and Catherine Pattillo, 2004, "Terms of Trade Shocks in Africa: Are They Short-Lived or Long-Lived?" *Journal of Development Economics,* Vol. 73 (April), pp. 727–44.

Chen, Shiu-Sheng, and Charles Engel, 2004, "Does 'Aggregation Bias' Explain the PPP Puzzle?" NBER Working Paper No. 10304 (Cambridge, Massachusetts: National Bureau of Economic Research).

Chen, Yu-Chin, and Kenneth Rogoff, 2003, "Commodity Currencies," *Journal of International Economics,* Vol. 60 (May), pp. 133–60.

Cheung, Yin-Wong, Menzie Chinn, and Eiji Fujii, 2001, "Market Structure and the Persistence of Sectoral Real Exchange Rates," *International Journal of Finance and Economics,* Vol. 6 (April), pp. 95–114.

Cheung, Yin-Wong, and Kon S. Lai, 2000a, "On Cross-Country Differences in the Persistence of Real Exchange Rates," *Journal of International Economics,* Vol. 50 (April), pp. 375–97.

———, 2000b, "On the Purchasing Power Parity Puzzle," *Journal of International Economics,* Vol. 52 (December), pp. 321–30.

Conover, W. J., 1999, *Practical Nonparametric Statistics* (New York: Wiley, 3rd ed.).

Devereux, J., and M. Connolly, 1996, "Commercial Policy, the Terms of Trade and the Real Exchange Rate Revisited," *Journal of Development Economics,* Vol. 50 (June), pp. 81–99.

Edwards, Sebastian, and Miguel A. Savastano, 2000, "Exchange Rates in Emerging Economies: What Do We Know? What Do We Need to Know?" in *Economic Policy Reform: The Second Stage,* ed. by Anne O. Krueger (Chicago: University of Chicago Press), pp. 453–510.

Faruqee, Hamid, 1995, "Pricing to Market and the Real Exchange Rate," *IMF Staff Papers,* Vol. 42, No. 4, pp. 855–81.

Frenkel, Jacob, 1978, "Purchasing Power Parity: Doctrinal Perspective and Evidence from the 1920s," *Journal of International Economics,* Vol. 8 (May), pp. 169–91.

Froot, Kenneth, and Kenneth Rogoff, 1991, "The EMS, the EMU, and the Transition to a Common Currency," in *NBER Macroeconomics Annual 1991,* ed. by Olivier Blanchard and Stanley Fischer (Cambridge, Massachusetts: MIT Press), pp. 269–317.

———, 1995, "Perspectives on PPP and the Long-Run Real Exchange Rates," in *Handbook of International Economics, Vol. 3,* ed. by Gene Grossman and Kenneth Rogoff (Amsterdam: Elsevier), pp. 1647–88.

Goldfajn, Ilan, and Rodrigo Valdés, 1999, "The Aftermath of Appreciations," *Quarterly Journal of Economics,* Vol. 114 (February), pp. 229–62.

Imbs, Jean, Haroon Mumtaz, Morten Ravn, and Hélène Rey, 2005, "PPP Strikes Back: Aggregation and the Real Exchange Rate," *Quarterly Journal of Economics,* Vol. 120 (February), pp. 1–43.

International Monetary Fund, 1998, *Annual Report on Exchange Arrangements and Exchange Restrictions,* and earlier issues (Washington).

———, 2000, *World Economic Outlook: May 2000* (Washington).

———, 2002, *World Economic Outlook: September 2002* (Washington).

McDermott, C. John, 1996, "Estimation of the Near Unit Root Model of Real Exchange Rates," IMF Working Paper 96/50 (Washington: International Monetary Fund).

McNown, R., and M. Wallace, 1989, "National Price Levels, Purchasing Power Parity, and Cointegration: A Test of Four High-Inflation Economies," *Journal of International Money and Finance,* Vol. 8 (December), pp. 533–45.

Montiel, Peter, 1997, "Exchange Rate Policy and Macroeconomic Management in ASEAN Countries," in *Macroeconomic Issues Facing ASEAN Countries,* ed. by J. Hicklin, D. Robinson, and A. Singh (Washington: International Monetary Fund), pp. 253–98.

Murray, Christian, and David Papell, 2002, "The Purchasing Power Parity Persistence Paradigm," *Journal of International Economics,* Vol. 56 (January), pp. 1–19.

Mussa, Michael, 1986, "Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates: Evidence and Implications," Carnegie-Rochester Conference Series on Public Policy, Vol. 25, pp. 117–213.

Reinhart, Carmen, and Kenneth Rogoff, 2002, "The Modern History of Exchange Rate Arrangements: A Reinterpretation," NBER Working Paper No. 8963 (Cambridge, Massachusetts: National Bureau of Economic Research).

Rogoff, Kenneth, 1996, "The Purchasing Power Parity Puzzle," *Journal of Economic Literature,* Vol. 34 (June), pp. 647–68.

———, Kenneth Froot, and Michael Kim, 2001, "The Law of One Price Over 700 Years," IMF Working Paper 01/174 (Washington: International Monetary Fund).

Romer, David, 1993, "Openness and Inflation: Theory and Evidence," *Quarterly Journal of Economics,* Vol. 108 (November), pp. 869–903.

Sarno, Lucio, and Mark P. Taylor, 2002, "Purchasing Power Parity and the Real Exchange Rate," *IMF Staff Papers,* Vol. 49 (April), pp. 65–105.

Taylor, Alan, 2001, "Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price," *Econometrica,* Vol. 69 (2), pp. 473–98.

Taylor, Mark P., 2003, "Purchasing Power Parity," *Review of International Economics,* Vol. 11 (August), pp. 436–452.

———, and Lucio Sarno, 1998, "The Behavior of Real Exchange Rates During the Post-Bretton Woods Period," *Journal of International Economics,* Vol. 46 (December), pp. 281–312.

Wickham, Peter, 1993, "A Cautionary Note on the Use of Exchange Rate Indicators," IMF Paper on Policy Analysis and Assessment No. 93/5 (Washington: International Monetary Fund).

World Bank, 2002, *World Development Indicators* (Washington).

Zanello, Alessandro, and Dominique Desruelle, 1997, "A Primer on the IMF's Information Notice System," IMF Working Paper 97/71 (Washington: International Monetary Fund).

Zar, Jerrold, 1972, "Significance Testing of the Spearman Rank Correlation Coefficient," *Journal of the American Statistical Association,* Vol. 67 (September), pp. 578–80.

IMF Staff Papers
Vol. 53, No. 1

New Estimates of Government Net Capital Stocks for 22 OECD Countries, 1960–2001

CHRISTOPHE KAMPS*

The issue of whether government capital is productive has received a great deal of attention recently, yet empirical analyses of public capital productivity have generally been limited to the official capital stock estimates available in a small sample of countries. Alternatively, many researchers have investigated the output effects of public investment—recognizing that investment may be a poor proxy for the corresponding capital stock. This paper attempts to overcome the data shortage by providing internationally comparable capital stock estimates for 22 Organization for Economic Cooperation and Development (OECD) countries. [JEL C82, E22, E62, H54]

The issue of whether government capital is productive has received a great deal of recent attention. In his seminal contributions, Aschauer (1989a, 1989b) found large positive output effects of government capital in the United States. His results suggested that government capital was even more productive than private capital. However, the large body of empirical literature that developed after Aschauer's early studies challenged these results. This literature—surveyed, for example, in Gramlich (1994); Sturm, Kuper, and de Haan (1998); and Seitz (2001)—concludes

*Christophe Kamps is an Economist at the European Central Bank. He thanks the editor, Robert P. Flood, and an anonymous referee for helpful comments and suggestions. The first draft of this paper was written while Kamps was a summer intern in the Fiscal Affairs Department of the International Monetary Fund. He thanks the staff of the Fiscal Policy and Surveillance Division for their hospitality and cooperation. He is especially indebted to Teresa Dabán for many helpful and clarifying discussions. He thanks country desk economists for providing him with information on national data sources. Finally, he thanks Alfred Boss, Kai Carstensen, Joerg Doepke, Joachim Scheide, Jan-Egbert Sturm, Marcel Timmer, and Bart van Ark for helpful comments. The capital stock estimates are available on request from *IMF Staff Papers* (publications@imf.org) or via the Internet at: www.ifw-kiel.de/forschung/netcap/netcap.htm.

that although there is evidence for positive output effects of government capital, the magnitude of these effects is generally much smaller than Aschauer reported.

The lack of public capital stock data for a large number of Organization for Economic Cooperation and Development (OECD) member countries has forced most empirical studies to focus on the United States. Few studies have investigated the productivity of government capital for other OECD countries; examples are Ford and Poret (1991) and Evans and Karras (1994). Both studies drew their data from OECD (1997), which included capital stock series (provided by national authorities) for 12 countries over the period 1970–96. However, the OECD data were not internationally comparable, because estimation methods differed widely across countries.[1] This was one of the reasons the OECD suspended publication of the capital stock series after 1997 and cofounded the Canberra Group on Capital Stock Statistics, whose activities resulted in the publication of a manual on the measurement of capital (OECD, 2001). Thus far, only a few countries have adjusted their estimation methods, so internationally comparable capital stock data are still not available.

In spite of these constraints, the analysis of public capital productivity has continued to be an active area of research. Most recent studies use an approach based on vector autoregressive (VAR) models—for example, Mittnik and Neumann (2001), Voss (2002), and Kamps (2005)—that, unlike the earlier production function and cost function approaches, does not impose causal links among the variables under investigation. The main disadvantage of this approach is that it generally requires large data samples for conventional lag lengths. For this reason, most researchers employing the VAR approach have used data on public investment instead of data on the public capital stock. This choice has been dictated by the lack of capital stock data for a large number of countries and the fact that public investment data are usually available each quarter, whereas public capital stock data are available only once a year. One drawback of this choice is the implicit assumption that the effects of public investment are independent of the level of the corresponding capital stock. Economic theory suggests that this assumption is dubious. According to the law of diminishing returns, an increment to the public capital stock (that is, public investment corrected for fixed capital consumption) would have a small (large) output effect if the capital stock in the previous period were large (small). There is indeed evidence for a fast decline in the marginal productivity of public capital (Demetriades and Mamuneas, 2000, p. 702).

This paper provides internationally comparable annual capital stock estimates for 22 OECD countries for the period 1960–2001, calculated using the perpetual inventory method based on a geometric depreciation pattern. Capital stock data are estimated for three categories of investment: (1) private nonresidential gross fixed capital formation, (2) private residential gross fixed capital formation, and (3) government gross fixed capital formation. A sensitivity analysis explores the robust-

[1]Sturm, Kuper, and de Haan (1998, p. 382) draw attention to the problem that "most authors employ data in their analysis which are generally chosen on the ground of their availability, without analyzing whether their conclusions are sensitive not only to the concept of the public capital stock (narrow versus broad definition), but also to the way the capital stock has been constructed."

ness of the capital stock estimates by varying the main estimation assumptions for a reference country, the United States, for which public capital stock estimates from official sources are available. The results of this sensitivity analysis suggest that the benchmark capital stock estimates are robust. Finally, the paper reports and compares estimation results for the production function approach, using three alternative measures of the public capital stock: (1) the author's estimates, (2) estimates from national authorities, and (3) estimates from OECD (1997). The regression results confirm previous results in the literature based on the production function approach and show that the elasticity of output with respect to public capital is positive and statistically significant, but quite large for most countries. According to estimates based on a simple panel data model, the elasticity of output with respect to public capital is, on average, 0.2 in OECD countries.

I. Data

The basic ingredient for the estimation of capital stocks is historical data on gross investment. These data are taken from the June 2002 version of the OECD Analytical Database.[2] The estimation is carried out for the following 22 OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States.[3] The series retrieved from this database expressed in constant prices and in national currencies are (OECD code in parentheses) total gross fixed capital formation (ITV), total private gross fixed capital formation (IPV), private nonresidential gross fixed capital formation (IBV), private residential gross fixed capital formation (IHV), and government gross fixed capital formation (IGV).

This paper uses public investment data from the OECD Analytical Database because, in addition to providing long time series for a large panel of countries, the data are categorized by institutional sector and are more comparable internationally than data from national sources. Table 1 gives details about the institutional coverage of the public investment series. This information is important because the definition of the public sector underlying the investment series varies not only across sources for a given country but also across countries for a given source. Three definitions regarding the coverage of public investment are used by the countries under investigation:[4] (1) public investment of the general government; (2) public

[2]This database includes not only investment data but also a large set of other macroeconomic variables. In addition to investment, real gross domestic product (GDPV) and employment (ET) are retrieved. Most of these data are available via the Internet at: www.sourceoecd.org.

[3]For most of these countries, the data are available for the period 1960–2001. The following OECD countries are not included in the analysis because long investment series are not available: the Czech Republic, Hungary, the Republic of Korea, Luxembourg, Mexico, Poland, the Slovak Republic, and Turkey.

[4]In OECD (1997), public entities are referred to as "producers of government services." This category, in most cases, corresponds to the definition of public activities underlying the investment series from national sources. An important difference, however, is that capital stocks in OECD (1997) were based on classifications of public activities according to the 1968 System of National Accounts, whereas recent national data are based on classifications according to the 1993 System of National Accounts.

Table 1. Coverage of Government Gross Fixed Capital Formation

Country	OECD ADB	National Data	OECD FSFC
Australia	General government	General government	PGS[4]
Austria	General government	General government	—
Belgium	General government	Public administration[1]	PGS[4]
Canada	General government	Public administration[2]	PGS[4]
Denmark	General government	General government	PGS[4]
Finland	General government	General government	PGS[4]
France	General government	—	—
Germany	General government	General government	PGS[4]
Greece	General government	—	General government
Iceland	General government	—	—
Ireland	General government	—	—
Italy	General government	Public administration[1]	PGS[4]
Japan	Public sector	General government	—
Netherlands	General government	—	—
New Zealand	Public sector	Public sector	—
Norway	General government	General government	PGS[4]
Portugal	General government	—	—
Spain	General government	Public administration[3]	—
Sweden	General government	—	PGS[4]
Switzerland	General government	—	—
United Kingdom	General government	General government	PGS[4]
United States	Public sector	Public sector	—

Sources: OECD ADB = OECD Analytical Database, which provides the investment series used in the calculation of the author's capital stock estimates. National data = government capital stock data available from national authorities (see Kamps, 2004, Appendix I for details). OECD FSFC = OECD (1997) publication *Flows and Stocks of Fixed Capital 1970–1996.*

Notes: The general government sector comprises the central, local, and state government subsectors, including social security funds. The public sector comprises the general government and nonfinancial public corporations. See IMF (2001) for details.

[1]Public administration and defense services, compulsory social security services (ISIC (International Standard Industrial Classification) category).

[2]Public administration, educational services, and health care and social assistance (ISIC category).

[3]Public administration, educational services, and health care and social assistance provided by the government (no private provision included) plus investment in infrastructure provided by public corporations.

[4]Producers of government services (PGS). The coverage of PGS is somewhat smaller than that of the general government, since the latter also includes departmental enterprises (Florio, 2001, p. 185). Moreover, capital stocks in OECD (1997) were based on classifications of government activities according to the 1968 System of National Accounts, whereas recent national data are based on classifications according to the 1993 System of National Accounts.

investment of the nonfinancial public sector (general government plus nonfinancial public enterprises (NFPEs)); and (3) investment carried out by economic units engaged in activities labeled "Public administration and defense services, compulsory social security services" (narrow definition) and "Public administration, educational services, and health care and social assistance" (broad definition) according to the International Standard Industrial Classification (ISIC). Whereas

the first two definitions categorize by institutional sector, the latter two categorize by type of economic activity. The ISIC category "Public administration, educational services, and health care and social assistance" comes close to the definition of the nonfinancial public sector, but it also includes private spending on education, which may be important in some countries, such as Canada.

Figure 1 displays the evolution of the public investment–GDP ratio for the 22 countries over the period 1960–2001. The subpanels of Figure 1 for the individual countries show the public investment–GDP ratio for the data from the OECD Analytical Database as well as (if available) for data from national sources[5] and from OECD (1997). The public investment–GDP ratio has, in general, declined over the sample period. In some countries, such as Austria and Ireland, the decline has been particularly pronounced. Notable exceptions to the general pattern are Greece, Portugal, and Spain, where public investment–GDP ratios have increased considerably since their accession to the European Community in the 1980s. A comparison of data from the three sources reveals that the ratios calculated for data from the OECD Analytical Database match the ratios calculated with data from national sources except for Belgium, Canada, Italy, Japan, and Spain. The differences between data from the Analytical Database and national sources and data from OECD (1997) are quite large in most cases, partly reflecting the fact that OECD (1997) data are based on the 1968 System of National Accounts, while the other two sources rely on the 1993 System of National Accounts.[6] To a certain extent, the difference is also due to data revisions since the publication of OECD (1997). However, the most important reason for the data discrepancy seems to be different coverage of public investment.

II. Methodology

The methodology applied here in the estimation of capital stock data draws in large part on OECD (2001) and the U.S. Bureau of Economic Analysis (1999). The estimation exercise is comparable to that performed by Jacob, Sharma, and Grabowski (1997), who estimate capital stocks by industrial activity according to the ISIC. Here, however, the aim is to obtain estimates by institutional sector—with a special focus on the public sector—rather than by industrial activity. Other precursors of the present study include Berndt and Hansson (1991), who estimate the public capital stock for Sweden; Sturm and de Haan (1995), who estimate the public capital stock for the Netherlands; as well as Boskin, Robinson, and Huber (1987); Munnell (1990); and Holtz-Eakin (1993), who estimate the capital stock for local and state governments in the United States. This paper employs the same estimation approach as those studies: the perpetual inventory method. The net capital stock is estimated for the three subcategories of gross investment available from the OECD Analytical Database: (1) private nonresidential gross fixed capital

[5]See Kamps (2004, Appendix I) for a description of data from national sources.

[6]For details of the 1993 System of National Accounts, see European Commission and others (1993).

Figure 1. Real Government Gross Fixed Capital Formation in 22 OECD Countries, 1960–2001
(*As a percentage of real GDP*)

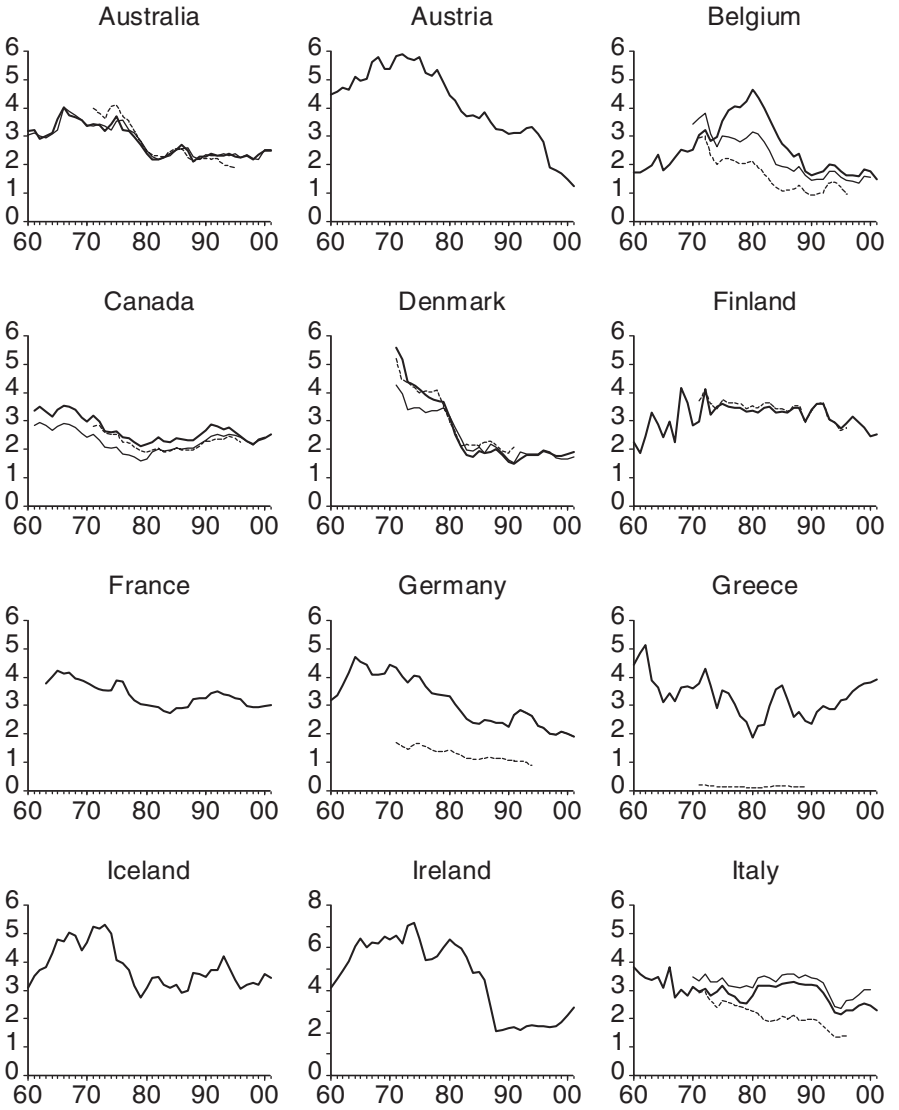
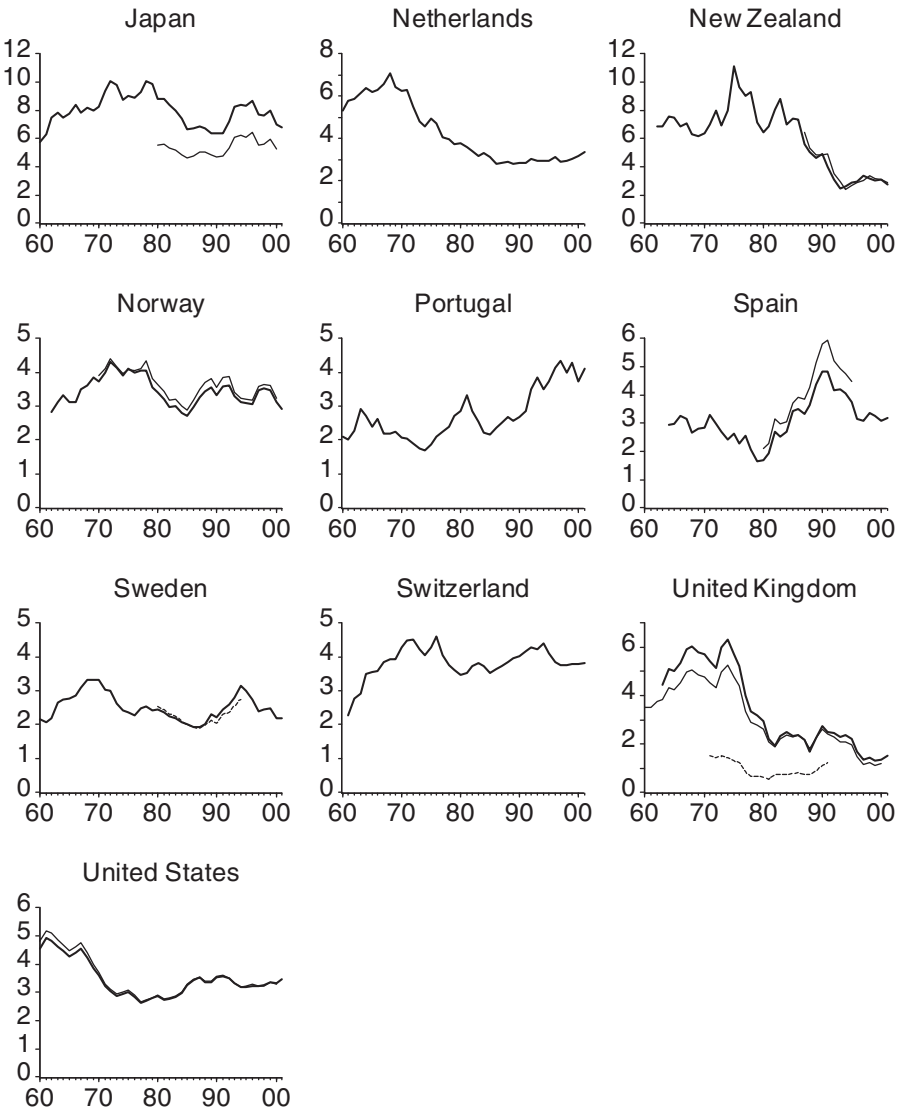


Figure 1. *(concluded)*



Sources: Series labeled "OECD ADB" are taken from the OECD Analytical Database, Version June 2002. Series labeled "National source" are estimates from national authorities. Series labeled "OECD FSFC" are calculated as follows: The gross investment series are taken from OECD (1997) and divided by GDP series taken from OECD (1998).

formation, (2) private residential gross fixed capital formation, and (3) government gross fixed capital formation.[7]

The basic idea of the perpetual inventory method is that the net capital stock at the beginning of the following period, K_{t+1}, can be expressed as a function of the net capital stock at the beginning of the current period, K_t, of gross investment in the current period, I_t, and of depreciation[8] in the current period, D_t:

$$K_{t+1} = K_t + I_t - D_t. \tag{1}$$

If one further assumes geometric depreciation (that is, the capital stock depreciates at a constant rate, δ), then the capital accumulation equation can be rewritten as

$$K_{t+1} = (1-\delta) K_t + I_t. \tag{2}$$

The method is called "perpetual" because all assets are forever part of the inventory of capital stocks. Of course, the quantity of services provided by an asset declines as it ages, but it never reaches zero. This can be seen by repeatedly substituting equation (2) for the capital stock at the beginning of period *t:*

$$K_{t+1} = \sum_{i=0}^{\infty} (1-\delta)^i I_{t-i}. \tag{3}$$

This expression shows that the capital stock at the beginning of period $t + 1$ is a weighted sum of past investment where the weights are a decreasing function of the distance between the current period and the investment period. In practice, an infinite number of past investment flows is not available, so equation (3) is replaced by the following expression:

$$K_{t+1} = (1-\delta)^t K_1 + \sum_{i=0}^{t-1} (1-\delta)^i I_{t-i}, \tag{4}$$

where K_1 is the initial capital stock at the beginning of period 1.

According to equation (4), the application of the perpetual inventory method requires the following inputs. First, a time series on gross investment flows is needed. The estimations in this paper rely on investment data from the OECD Analytical Database. Second, the estimations necessitate an initial capital stock; in our case, the capital stock at the beginning of 1960. Third, an assumption must be

[7]All series are expressed in the constant prices of 1995. For countries with a different base year (Australia 1999/2000; Canada, 1997; Iceland, 1990; Norway, 1997; Switzerland, 1990; and the United States, 1996), the series were rebased to 1995.

[8]The terms "depreciation" and "consumption of fixed capital" are used interchangeably in this paper. This is common in economic literature. Note, however, that "depreciation" as used here differs considerably from its use in company accounts, where it is calculated on the basis of historic costs rather than market prices.

made of the size and the time profile of the depreciation. Finally, a depreciation method must be chosen. This study relies on geometric depreciation.

There is no official information on the magnitude of the initial capital stock for any country except the United States. This paper follows an approach similar to that used by Jacob, Sharma, and Grabowski (1997, p. 567) to estimate the initial capital stock. For that purpose, an artificial investment series for the period 1860–1959 is constructed for each country by assuming that investment increased by 4 percent a year during this period, finally reaching its observed level in 1960.[9] The rationale for this assumption is that total gross investment in the 22 OECD countries under consideration grew by 4 percent a year on average during the period 1960–2001. It is, of course, highly improbable that investment in the earlier period grew at the same rate or that investment growth was the same for all countries. However, as historical information—especially on public investment—is not available for most of the countries under consideration, this study opts for an equal treatment of all countries.[10] The results of a sensitivity analysis described in Section IV suggest that the assumption on the initial capital stock does not affect the dynamics of the resulting capital stock series to a great extent. Also, the importance of the initial capital stock to the level of the capital stock series fades over time. Its contribution to the level of the capital stock at the end of the sample is less than 10 percent for the average OECD country.

This assumption implies that investment and the capital stock grew smoothly over the period 1860–1960, a simplification that needs to be justified, especially since some countries in the sample experienced severe damage during World War II.[11] The question arises as to whether the effects of war damage on the capital stock were persistent enough to affect our estimate of the initial capital stock at the beginning of 1960. There are at least two reasons why war damage might have had little effect by 1960. First, the countries that experienced the most war

[9]An exception to this general rule applies in the case of the United States. The U.S. Bureau of Economic Analysis (BEA) provides investment series starting in 1914. This information is used by chaining the OECD data, which are available for 1960–2001, with the BEA growth rates for 1914–60.

[10]Maddison (1995) estimates gross capital stocks for six OECD countries based on investment series that in some cases start in the 19th century. However, he acknowledges that the assembly of the investment series is a major problem, because historical series for different periods have to be linked and because these series rely on different weighting bases. Also, there are generally breaks in the historical investment series. However, most important for our purpose is that historical series on investment in general do not include a measure of government investment but only measures of private investment. An exception relates to the influential works of Feinstein (for example, 1972), who provides investment series for the public sector in the United Kingdom starting in 1856. However, the definition of the public sector in his studies differs considerably from that underlying the OECD series.

[11]In his estimation of nonresidential capital stocks for six OECD countries, Maddison (1991, pp. 284–92) assumed that the loss in capital stock caused by war damage amounted to 3 percent in the United Kingdom, 8 percent in France, 10 percent in the Netherlands, 16 percent in Germany, and 25.7 percent in Japan. These figures are subject to a large margin of uncertainty, though, and the war damage to productive capacities may well have been lower. For example, Ritschl (2003, p. 414) reports that the industrial capital stock in Germany in 1945 exceeded prewar levels by a third. This capital stock was often composed of multipurpose machinery and, thus, was available for civil production. Likewise, Giersch, Paqué, and Schmieding (1992, p. 17) and Eichengreen and Ritschl (1998, p. 8) note that war damage to the capital stock in Germany was quite limited.

damage grew much faster after World War II than the countries that suffered little or no war damage.[12] Thus, real GDP in countries subject to severe war damage might well have returned to its long-run trend by 1960.[13] Second, the strong rise in GDP after World War II in these countries was accompanied by fast increases in investment and in the capital stock.[14] As a consequence, the capital stock might also have returned to its long-run trend by 1960.[15] All in all, there are hints that the effects of war damage on the capital stock in countries that were most severely affected had disappeared by 1960, the starting year for our capital stock estimates. The strong growth performance in the quarter century following World War II (often referred to as a "growth miracle") was distributed unevenly among OECD countries. As Broadberry (1988, p. 26) notes, the countries most hit by war damage, such as Japan and Germany, grew much faster than countries that were hardly affected by war damage, such as the United Kingdom and the United States. The evidence presented in Maddison (1982, p. 55) and Eichengreen and Ritschl (1998, p. 32) further suggests that by 1960 real GDP in OECD countries was close to the level implied by its long-run growth path. Moreover, the analysis in Eichengreen and Ritschl (1998, p. 8) shows that the strong growth performance in Germany was not impaired by a lack of capital resulting from war damage. On the basis of these considerations, we chose not to correct the capital stock estimates at the beginning of 1960 for war damage.

The next assumption in the estimation relates to the size and the time profile of the depreciation rate. In this study, it is assumed that the depreciation rate is time-varying for the public capital stock and the private nonresidential capital stock and constant for the private residential capital stock.[16] This assumption allows us to

[12]Maddison (1987, p. 650) reports that between 1950 and 1973, real GDP grew by 9.4 percent a year on average in Japan and by 5.9 percent in Germany, whereas the average annual growth rate was only 3 percent for the United Kingdom and 3.7 percent for the United States.

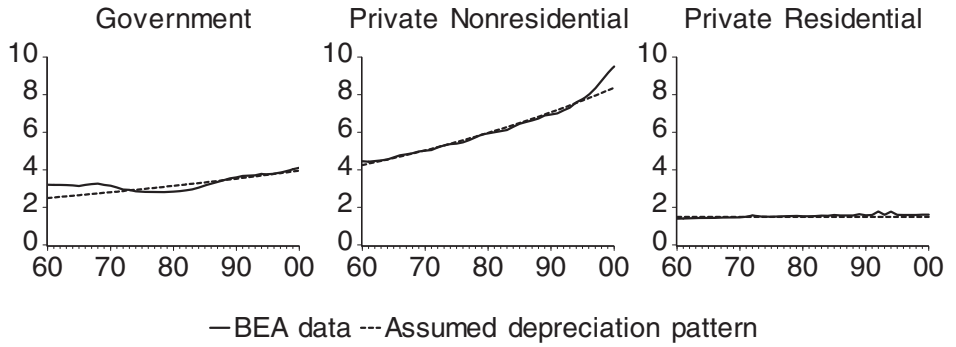
[13]For example, long-run growth in real GDP was remarkably stable in West Germany, despite the disruptions caused by World War II. Calculations based on historical GDP figures drawn from Ritschl and Spoerer (1997) reveal that real GDP grew by 2.9 percent a year on average between 1938 and 1960 and by 3.1 percent between 1960 and 1990. Of course, real GDP did not grow smoothly in the first subperiod. Real GDP in 1946 was lower than in 1938 by roughly 60 percent. However, real GDP growth over the period 1946–60 was extremely strong, averaging almost 12 percent a year.

[14]Maddison (1987, p. 657) reports that between 1950 and 1973, the private capital stock on average grew in Japan by 9.5 percent a year and in Germany by 7.2 percent. Calculations based on Luetzel (1977, p. 66) show that the total net capital stock in Germany grew by 7.9 percent a year on average between 1950 and 1960. Calculations based on Mitchell (1975) show that net investment in Germany grew by 9.7 percent a year on average between 1950 and 1960.

[15]The fall in capital stocks as a result of wartime disruptions was less pronounced than the fall in output (Maddison, 1982, p. 55), implying that the gap between capital stock levels at the end of World War II and their long-run trend was lower than was the case for output.

[16]This paper assumes that the time profile of the depreciation rates is the same across countries. Official estimates of capital stocks for different countries are, in general, based on different assumptions about depreciation rates. This is appropriate insofar as country-specific factors influence service lives. However, only a few countries have investigated service lives with particular care, among them the United States (OECD, 2001, p. 99). Therefore, it seems preferable to assume identical depreciation rates across countries for the purpose of international comparisons. Such a standardized approach is also adopted by Maddison (1995) and O'Mahony (1996).

Figure 2. Implicit Scrapping Rates for Net Capital Stocks in the United States, 1960–2000
(*In percentage*)



Sources: U.S. Bureau of Economic Analysis (BEA) and author's calculations.

take into account the empirically observed pattern of aggregate depreciation rates.[17] Figure 2 shows the implicit scrapping rate for the real government net capital stock, the real private nonresidential net capital stock, and the real private residential capital stock in the United States for the period 1960–2000. The implicit scrapping rate is calculated as the quotient of depreciation in period t and the net capital stock at the beginning of period t (data from the BEA):

$$s_t = \frac{D_t}{K_t} 100. \tag{5}$$

Figure 2 reveals that the implicit scrapping rates differ considerably across the three types of capital. The implicit scrapping rate is highest for private nonresidential capital and lowest for private residential capital. Whereas the scrapping rate for private residential capital has remained roughly constant over the period 1960–2000, the scrapping rates for private nonresidential capital and government capital have tended to rise over time. The increase is especially pronounced in the case of the private nonresidential capital stock, its scrapping rate having risen from about 4.5 percent in 1960 to about 9.5 percent in 2000.

Two developments may partly explain the rise in the scrapping rate. First, it may reflect an increasing weight of assets with relatively short asset lives, and second, it may reflect to a certain extent a shortening of asset lives. Both developments are characteristic of ICT-related assets that are part of the private nonresidential

[17]National authorities usually estimate the contribution of investment to the net capital stock for a large number of individual assets (BEA, 1999). For most of these assets, national authorities assume constant depreciation rates, except for assets related to information and communication technology (ICT). At the same time, they assume different depreciation rates for different types of assets. As the relative importance of different assets changes with time, so does the average depreciation rate. The latter will increase over time if assets with relatively short asset lives gain in importance. This paper tries to capture this phenomenon by assuming a time-varying aggregate depreciation rate.

capital stock and—to a lesser extent—of the government capital stock. A similar pattern for implicit scrapping rates can be observed for other countries. For example, Canadian and Australian data reveal that implicit scrapping rates of private nonresidential and government capital have also risen since the 1960s.

On the basis of this evidence, this paper makes the following assumptions about the time profile of the depreciation rates. For the period 1860–1960, the depreciation rate is assumed to be 2.5 percent for government assets, 4.25 percent for private nonresidential assets, and 1.5 percent for residential assets. For the period 1960–2001, it is assumed to increase gradually from 2.5 percent to 4 percent for government assets and from 4.25 percent to 8.5 percent for private nonresidential assets,[18] and to be a constant 1.5 percent for private residential assets (see Figure 2). Equation (6) formalizes the time profile of the depreciation rates:

$$\delta_t^j = \delta_{\min}^j \left(\left(\frac{\delta_{\max}^j}{\delta_{\min}^j} \right)^{\frac{1}{41}} \right)^{t-2001+41} \qquad \text{for all } t = 1960, 1961, \ldots, 2001. \tag{6}$$

where $\delta_{\min}^j$ is the depreciation rate of asset type j in 1960 and $\delta_{\max}^j$ is its depreciation rate in 2001.

Finally, the real net capital stock at the beginning of period $t + 1$ for investment category j can be expressed (see also BEA, 1999, p. M-5) as

$$\begin{aligned} K_{t+1}^j &= \sum_{i=1860}^{t-1} \left[\prod_{k=i+1}^{t} \left(1-\delta_k^j\right) \right] \left(1-\frac{\delta_i^j}{2}\right) I_i^j + \left(1-\frac{\delta_t^j}{2}\right) I_t^j \\ &= \left(1-\delta_t^j\right) K_t^j + \left(1-\frac{\delta_t^j}{2}\right) I_t^j \qquad \text{for all } t = 1860, 1861, \ldots, 2001. \end{aligned} \tag{7}$$

Note that the capital stock at the beginning of the initial period, K_{1860}, is set equal to zero. Equation (7) differs from equation (4) in two respects in order to increase the realism of the estimates. First, as discussed earlier, the equation incorporates time-varying depreciation rates. Second, new investment is assumed to be placed in service at midyear instead of at the end of the year as implied by equation (4). Investment typically occurs throughout the year, not only at the end of the year. Equation (7), in conjunction with the assumptions about the depreciation patterns, serves to estimate net capital stocks for government assets and for private nonresidential and private residential assets.

To sum up, the application of the perpetual inventory method requires three important assumptions. First, an initial capital stock must be constructed; in our

[18]Figure 2 shows that the implicit scrapping rate calculated for BEA data sharply accelerated after 1995. To some extent, this probably reflects the growing importance of ICT assets characterized by asset lives that are much shorter than those of other assets. Because the importance of the ICT sector is considerably lower in most other industrial countries than in the United States (OECD, 2002a), we chose a flatter depreciation profile for the years 1995–2001 than that implicit in U.S. data.

case, for the beginning of 1960. This study uses artificial investment series for the period 1860–1959 as an input for the estimation of the initial capital stock. Second, an assumption must be made on the level and the time profile of the depreciation rate. This paper assumes that the depreciation rate is time-varying and different across the three types of investment considered, but the same for all countries. Third, a depreciation method must be chosen. This study relies on geometric depreciation. Section IV analyzes in detail the importance of each of these assumptions for the resulting capital stock estimates.

III. Capital Stock Estimates for 22 OECD Countries, 1960–2001

Figure 3 displays the evolution of the ratio of the public net capital stock and GDP, both at 1995 prices, for the 22 OECD countries over the period 1960–2001.[19] The graphs for the individual countries plot not only the author's own estimates of the public capital stock but also—where available—estimates from national authorities and from OECD (1997). For seven countries, there is no benchmark against which the author's estimates could be assessed.[20] The same holds for Greece, in view of the very low capital stock estimate reported by OECD (1997) owing to a narrow definition of public investment. For the countries for which alternative estimates are available, the general picture is that in most cases the dynamics of the author's estimates resemble those of the alternative estimates. In some cases, the level of the public capital–GDP ratios is also similar, but in general there are significant differences. This, in part, reflects the fact that the initial level of the author's estimates depends on the artificial investment series assumed for the period 1860–1959. Yet, there are two other reasons for the differences: (1) for some countries, the investment series shown in Figure 1 differ sharply from each other (Greece); and (2) many different methods are used by national authorities to construct capital stock estimates. An attractive feature of the author's estimates is that they rely on the same methodology and homogenous investment data across countries, a condition that is not satisfied for the alternative estimates.

Table 2 shows the value of the public capital–GDP ratio for three reference years: 1980, 1990, and 2000. The average ratio for the 22 OECD countries has declined by six percentage points over the period 1980–2000. Public capital as a share of GDP has declined in 13 countries since the early 1980s; it has slightly increased in 4 countries and strongly risen in Greece, Japan, Portugal, Spain, and Switzerland. It does not come as a surprise that Greece, Portugal, and Spain are in this group, because these countries are known to have made substantial efforts to

[19]A special problem in the estimation of capital stocks relates to German reunification. The OECD investment and GDP series cover only West Germany for the period 1960–90, but they include East Germany from 1991 on. Since there is no information on the magnitude of the East German capital stock at the beginning of 1991, this paper assumes that the ratio of the East German capital stock to the West German capital stock equaled the ratio of East German and West German GDP in 1991 (8 percent). In the estimation, the German capital stock at the beginning of 1991 is thus increased by 8 percent for the three asset types considered. From 1991 on, this additional capital stock depreciates at the same rate as the other assets.

[20]Austria, France, Iceland, Ireland, the Netherlands, Portugal, and Switzerland.

Figure 3. Real Government Net Capital Stock
in 22 OECD Countries, 1960–2001
(*As a percentage of real GDP*)

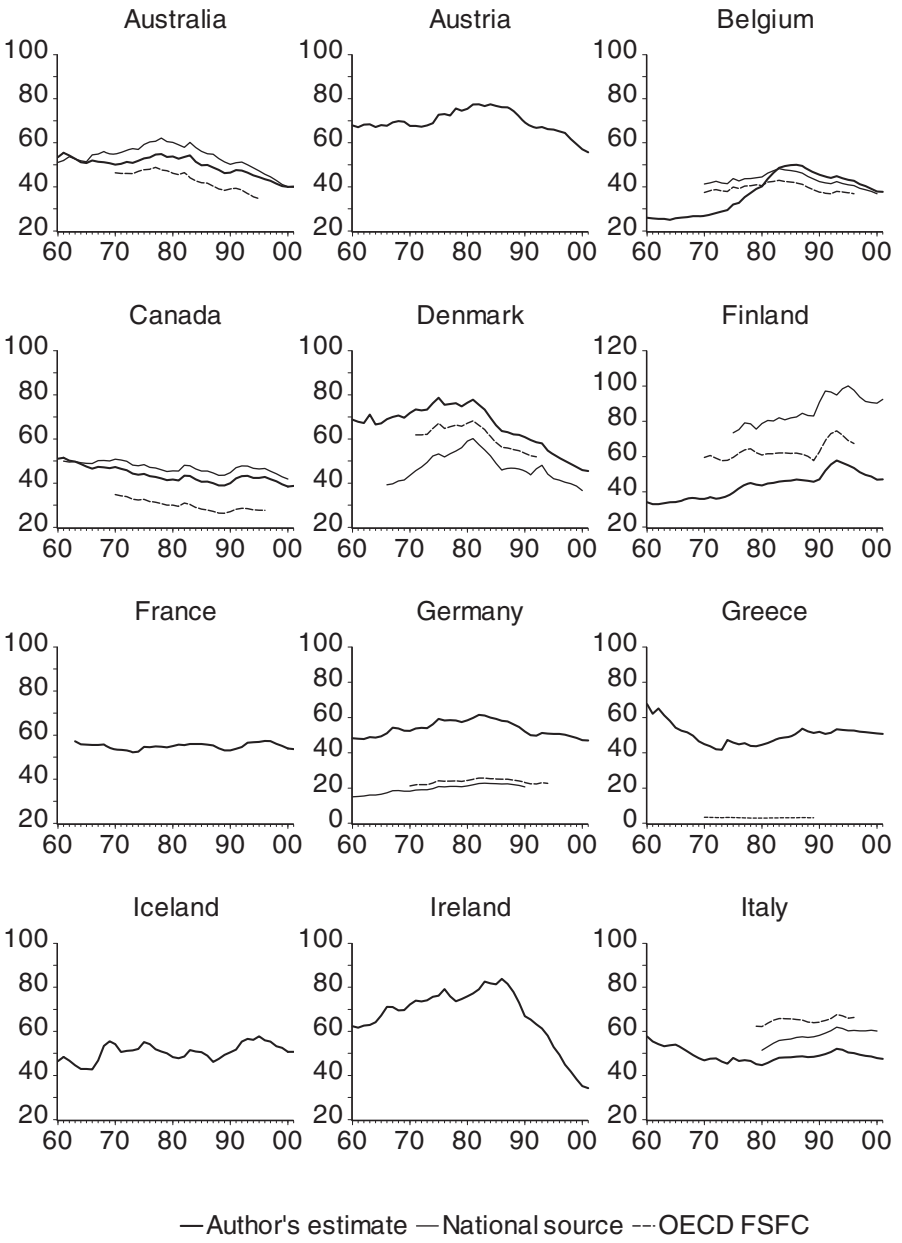
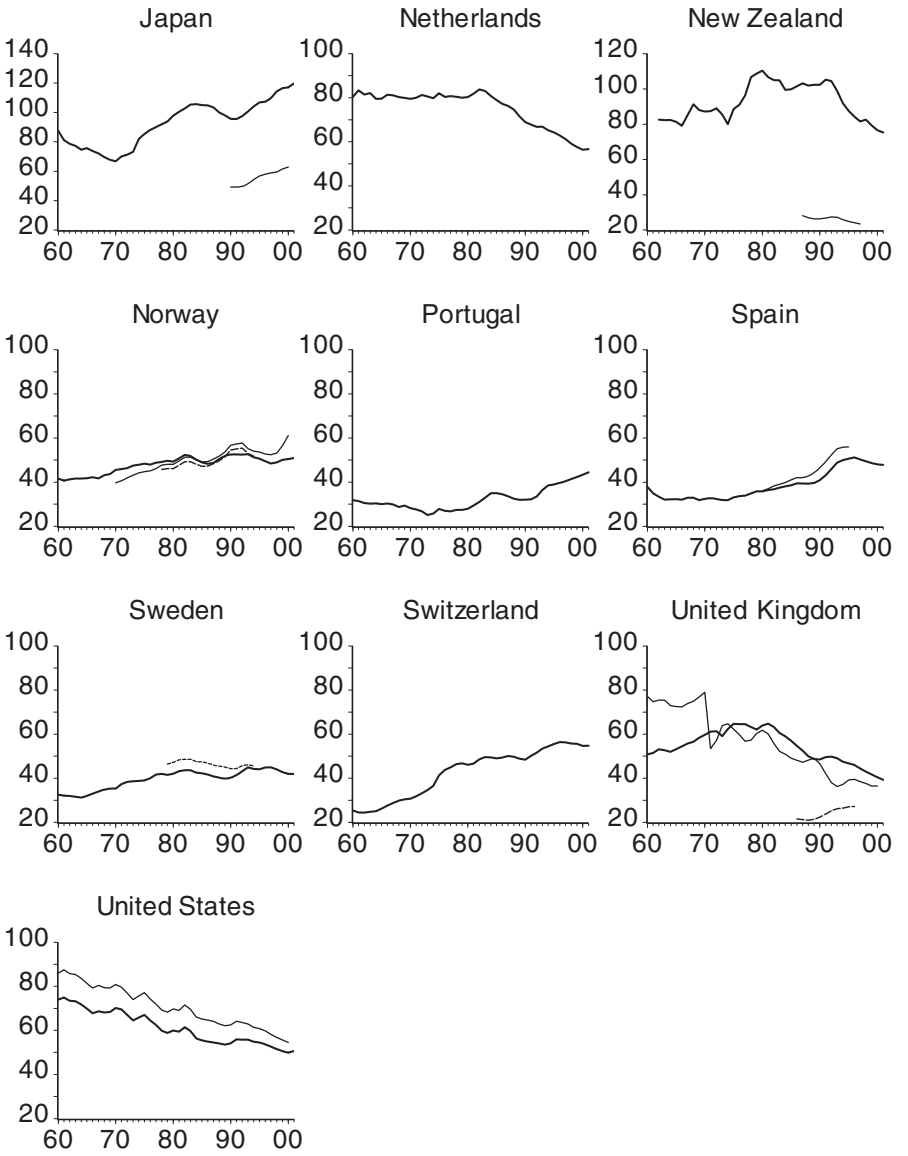


Figure 3 *(concluded)*



Sources: Series labeled "National source" are estimates from national authorities. Series labeled "OECD FSFC" are calculated as follows: the capital stock series are taken from OECD (1997), and divided by GDP series taken from OECD (1998).

Table 2. Government Net Capital Stock in 22 OECD Countries
(As a percentage of GDP, both at 1995 prices)

	1980		1990		2000		
Country	Ratio[1] (1)	Rank[2] (2)	Ratio[1] (3)	Rank[2] (4)	Ratio[1] (5)	Rank[2] (6)	Reliability[3] (7)
Australia	53.8	11	46.5	17	40.0	19	91.5
Austria	75.4	6	69.3	3	57.0	3	90.8
Belgium	40.2	20	45.5	18	37.9	21	94.4
Canada	41.6	19	40.0	21	38.4	20	91.5
Denmark	76.4	4	60.8	6	45.9	15	85.7
Finland	43.7	17	47.1	16	46.9	14	94.6
France	55.0	10	53.0	8	54.0	6	91.3
Germany	58.4	9	52.0	10	47.1	13	91.4
Greece	44.4	16	51.9	11	51.0	7	92.3
Iceland	48.4	13	50.5	12	50.7	8	94.8
Ireland	75.9	5	66.8	5	35.2	22	92.9
Italy	44.7	15	49.0	13	47.9	12	90.8
Japan	97.7	2	95.7	2	117.1	1	97.0
Netherlands	80.2	3	68.9	4	56.4	4	89.0
New Zealand	110.3	1	102.4	1	76.6	2	89.1
Norway	49.3	12	52.5	9	50.5	9	94.6
Portugal	27.9	22	32.0	22	43.3	16	96.2
Spain	35.8	21	40.9	19	48.0	11	95.7
Sweden	42.1	18	40.2	20	42.0	17	92.4
Switzerland	46.1	14	48.4	15	54.7	5	94.7
United Kingdom	63.9	7	48.5	14	40.3	18	86.8
United States	59.9	8	54.1	7	50.0	10	89.4
Average[4]	57.8		55.3		51.4		92.1
Standard deviation	20.6		17.0		17.1		3.0

Source: Author's estimates and calculations.

[1]The columns labeled "Ratio" give the ratio of the government capital stock and GDP (in percent).

[2]The columns labeled "Rank" give the ranking of the countries according to the size of their capital to GDP ratio; the country with the highest ratio ranks first.

[3]The column labeled "Reliability" gives the contribution of government investment in the period 1960–99 to the net capital stock at the beginning of 2000 (in percent). The difference between 100 percent and this figure reflects the contribution of government investment in the period 1860–1959, for which official data are not available.

[4]Unweighted average.

improve their infrastructure after joining the European Community. Furthermore, it is well known that during the 1990s the Japanese government repeatedly attempted (in vain) to reinvigorate the country's sluggish economy with the help of large spending programs focusing on construction. According to the estimates, in 2000 Japan had by far the largest public capital–GDP ratio among the OECD countries considered in this study, while Ireland had the lowest. The large decline in the public capital–GDP ratio in Ireland during the 1990s mirrors the strong fall of public investment as a share of GDP during this period. Fitz Gerald, Kearney, and

Morgenroth (1999) state that the Irish government viewed the lack of infrastructure as one of the major impediments to growth.

The last column of Table 2 shows the contribution of government investment in the period 1960–2000 to the government net capital stock in the year 2000. As explained in the previous section, assumed investment data for the period 1860–1959 are used to estimate the initial capital stock at the beginning of the year 1960. The contribution of the assumed investment data to the initial capital stock is 100 percent, but their influence gradually fades and is quite small at the end of the sample. The capital stock estimates for the year 2000 are largely unaffected by the assumption on investment during the period 1860–1959, for which no data from official sources are available. Investment from 1960 onward contributes an average 92 percent to the net capital stock in the year 2000, implying that the average contribution of the assumed investment data is only 8 percent.[21]

Another way to compare the government capital stock across countries is to look at its absolute value for each person. For an international comparison of real capital stock, three conditions must be met (OECD, 2002c, p. 8). First, the definition of the capital stock must be the same. This condition is fulfilled, since for all countries the investment data used in the capital stock estimations are compiled according to the 1993 System of National Accounts. Second, the capital stock must be expressed in the same currency; and third, the price level at which the capital stock is valued must be the same. The second and third conditions are not fulfilled, since the capital stock estimates are expressed in national currency and valued at the national price. These data can be converted to a common currency and revalued at a common set of prices using so-called purchasing power parities (PPPs). Unfortunately, PPPs for the public capital stock are not available. Instead, we must use a proxy for this measure. The OECD (2002c) provides PPPs for GDP and for total gross fixed capital formation. In the following paragraph, PPPs for gross fixed capital formation are used to convert the public capital stock estimates to U.S. dollars.[22]

Table 3 gives the real per capita public net capital stock expressed in 1999 PPPs for gross fixed capital formation in U.S. dollars for the years 1980, 1990, and 2000. The average capital stock for each person for the 22 OECD countries has increased by 32.2 percent over the period 1980–2000 (that is, by 1.4 percent a year). This growth rate has been lower than that of real GDP, implying that the public capital–GDP ratio has declined on average during this period in the OECD, as shown in Table 2. A comparison across countries reveals that the public capital stock per capita is by far the highest in Japan, exceeding the average by around 180 percent in the year 2000. Regarding the ranking of the countries, the United

[21]The contribution of the initial capital stock varies between 14.3 percent in the case of Denmark and 3.0 percent in the case of Japan. The differences in contributions across countries are mainly caused by differences in the level of public investment–GDP ratios over the sample period. For instance, the contribution of the initial capital stock is lowest in Japan because the public investment–GDP ratio there was the highest among the considered countries over the period 1960–2001.

[22]The most important qualitative results reported in the following paragraph are unaffected if PPPs for GDP are used instead. See Kamps (2004, Table 3) for results based on PPPs for GDP.

Table 3. Government Net Capital Stock Per Capita in 22 OECD Countries
(At 1999 PPPs in U.S. dollars)

	1980		1990		2000	
Country	Capital[1]	Rank[2]	Capital[1]	Rank[2]	Capital[1]	Rank[2]
Australia	8,003.7	11	8,176.0	15	8,817.5	16
Austria	10,856.7	7	12,359.1	5	12,188.5	4
Belgium	5,866.2	19	8,059.0	16	8,050.2	17
Canada	8,166.6	10	9,124.8	12	10,387.4	10
Denmark	11,809.9	4	11,343.6	7	10,361.4	11
Finland	7,004.6	15	9,809.3	10	11,629.0	8
France	7,615.3	13	8,805.9	13	10,317.5	12
Germany	10,415.8	8	11,273.6	8	10,242.9	13
Greece	3,828.0	20	4,562.3	21	5,412.1	21
Iceland	8,344.7	9	10,224.3	9	11,913.5	5
Ireland	6,219.7	18	7,589.3	18	7,449.7	19
Italy	6,259.3	17	8,531.9	14	9,566.1	14
Japan	16,396.7	1	22,717.8	1	31,147.2	1
Netherlands	11,280.2	6	11,387.6	6	11,623.0	9
New Zealand	12,008.5	3	13,403.4	4	11,697.7	7
Norway	7,170.4	14	9,329.5	11	11,844.4	6
Portugal	1,909.5	22	3,001.8	22	5,254.9	22
Spain	3,320.0	21	4,891.5	20	7,334.8	20
Sweden	6,742.0	16	7,766.6	17	9,322.4	15
Switzerland	11,798.3	5	14,434.3	2	16,591.4	2
United Kingdom	7,702.0	12	7,451.6	19	7,498.1	18
United States	12,640.2	2	14,230.5	3	16,397.4	3
Average[3]	8,425.4		9,930.6		11,138.5	
Standard deviation	3,440.8		4,115.1		5,313.1	

Source: Author's estimates and calculations.

[1]The columns labeled "Capital" give the level of the government net capital stock per capita at 1999 purchasing power parities (PPPs) for gross fixed capital formation in U.S. dollars. PPPs for gross fixed capital formation in national currency per U.S. dollar are taken from OECD (2002c). Population figures are taken from the OECD Analytical Database, Version June 2002.

[2]The columns labeled "Rank" give the ranking of the countries according to the level of public capital per capita; the country with the highest level ranks first.

[3]Unweighted average.

States ranks considerably higher than in Table 2, whereas Greece, Portugal, and Spain rank much lower. This discrepancy can be explained by the level of and change in real GDP. Output growth in the United States was very fast in the second half of the 1990s, implying a decreasing public capital–GDP ratio during the 1990s, even though public capital expanded substantially at the same time. In Greece, Portugal, and Spain, both public capital per capita and real GDP per capita remain low in the international comparison.[23] However, public capital per capita

[23]In 2000, real GDP per capita amounted to 66.3 percent of the OECD average in Greece, 72.4 percent in Portugal, and 79.5 percent in Spain (OECD, 2002b, p. 339).

has increased on average much faster in these countries than in OECD countries over the period 1980–2000, reflecting strong efforts to enhance public infrastructure in connection with accession to the European Community.

IV. Robustness of the Public Capital Stock Estimates

The following are the key assumptions made in the construction of the public capital stock estimates: (1) investment in the period 1860–1959 is assumed to have grown by 4 percent a year to finally reach its observed level in 1960, (2) the depreciation rate is time-varying and assumed to increase gradually for all asset types except residential assets, and (3) depreciation is assumed to be geometric. This section assesses the effect of the first two assumptions.[24] For that purpose, alternative public capital stock series are estimated for a reference country, the United States, modifying in turn each of the two assumptions. These alternative estimates are then compared with the benchmark estimates and with estimates from official sources. The results of this exercise are shown in Figure 4.

Figure 4a shows public capital stock estimates for alternative assumptions on the initial capital stock in 1960 for the United States, as well as the official capital stock estimate from the U.S. Bureau of Economic Analysis. The benchmark public capital stock estimate relies on the public investment series provided by the BEA for the period 1914–60 and assumes that public investment grew by 4 percent in the period 1860–1914.[25] The figure also shows an estimate that relies on the assumption that (as is assumed for all other countries) public investment in the United States grew by 4 percent in the period 1860–1959. In addition, Figure 4 shows an estimate that differs in the way the initial capital stock is constructed. This estimate of the initial public capital stock is constructed directly following a "steady-state approach" (Fuente and Doménech, 2002, p. 47). It follows from equation (2) that the growth rate of the capital stock can be expressed as

$$g_t^K = \frac{K_{t+1} - K_t}{K_t} = -\delta + \frac{I_t}{K_t}. \tag{8}$$

Thus, the capital stock at the beginning of period t can be calculated as investment in period t divided by the sum of the depreciation rate and the growth rate of the capital stock in period t:

$$K_t = \frac{I_t}{\delta + g_t^K}. \tag{9}$$

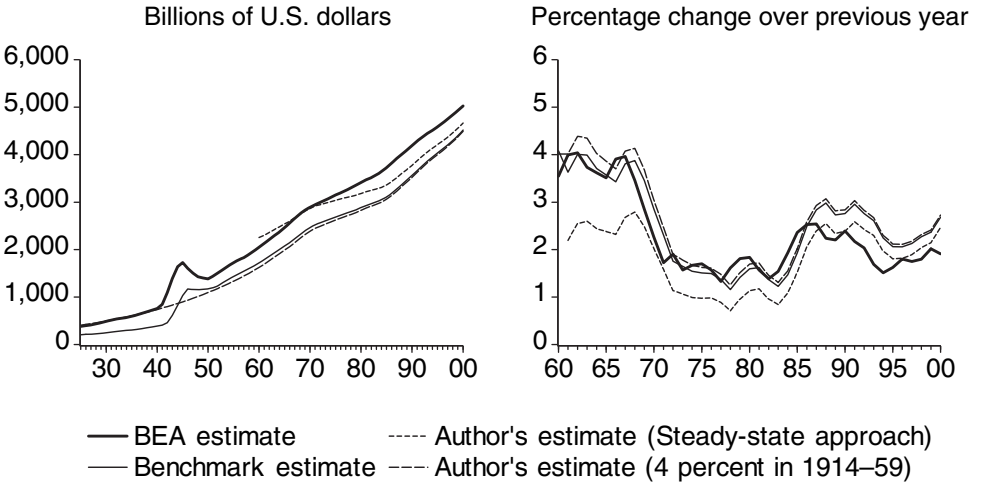
As the growth rate of the capital stock is unknown a priori, an assumption about its magnitude is needed. Neoclassical growth theory suggests that invest-

[24]See Kamps (2004, Section V) for an illustration of the influence of alternative depreciation methods for Canadian data, based on official estimates provided by Statistics Canada.

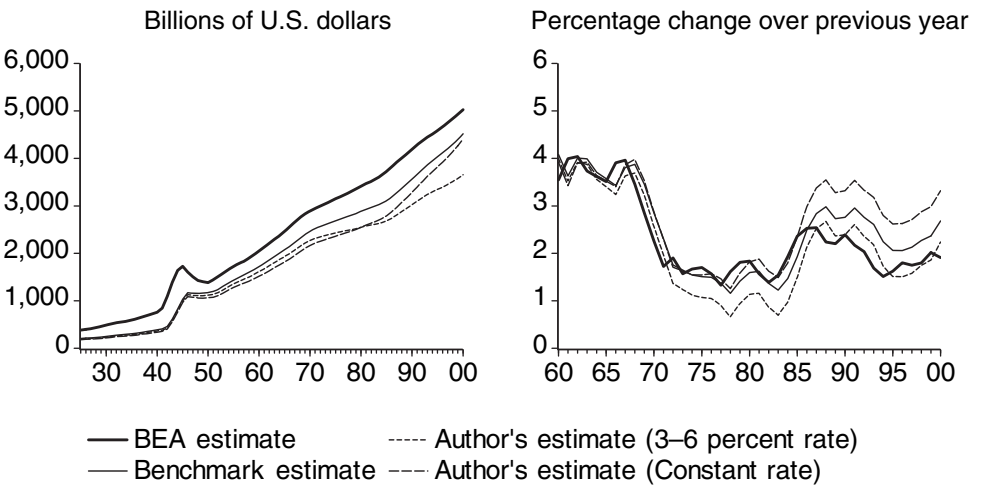
[25]The official growth rates provided by the BEA for the period 1914–60 are chained with the OECD investment series available for 1960–2001.

Figure 4. Estimates of Government Net Capital Stocks for the United States for Alternative Estimation Assumptions

(a) Alternative assumptions on the initial capital stock in 1960



(b) Alternative assumptions on depreciation rates



Source: Series labeled "BEA estimate" are from the U.S. Bureau of Economic Analysis (BEA).

ment and capital grow at the same rate in the steady state. Thus, the growth rate of capital can be approximated by the growth rate of investment. As it is improbable that the U.S. economy was in a steady state in 1960, the growth rate of capital is instead approximated by the average growth rate of investment over the period 1960–2001. This growth rate was 2.7 percent for government investment, 5.4 percent for private nonresidential investment, and 2.5 percent for private residential

investment. Finally, the investment series is filtered using a Hodrick-Prescott filter to remove its cyclical component.

Figure 4a shows that assuming a constant 4 percent growth rate of public investment over the period 1914–59 instead of using the historical BEA investment growth rates does not significantly affect the resulting public capital stock series. For the period of interest, 1960–2000, the levels and dynamics of both series are almost indistinguishable. Pronounced differences occur over the period 1940–50 only because the public capital stock based on BEA investment takes into account the large increase in public investment during World War II, whereas the alternative estimate does not. However, the effect of this difference is short-lived. The figure also reveals that although the level of the public capital stock estimates is lower than the capital stock series reported by the BEA, the dynamics are very similar, as witnessed by the strong co-movement of the growth rates of the public capital stock series.

Furthermore, formal tests for equality between the means and the variances of the author's estimates and the BEA series fail to reject the null hypothesis of equality.[26] Figure 4a also shows the author's estimate of the public capital stock relying on the steady-state approach. The time profile of this estimate differs considerably from that of the other series shown in the figure. While the level of the initial capital stock at the beginning of 1960 is higher than that of the BEA reference series, its level at the beginning of 2000 is comparable to that of the author's estimate and thus lower than that of the BEA series. Furthermore, the dynamics of this estimate are significantly different from that of the other series, as can be seen in the figure. This visual impression is confirmed by a formal test for equality between the means and variances of the growth rates of this estimate and the BEA series. These results suggest that the benchmark assumption on the initial capital stock made in this paper is a reasonable way to proceed.

Figure 4b shows public capital stock estimates for alternative assumptions on the time profile of the depreciation rate and contrasts these estimates with the BEA estimate. The benchmark estimate assumes a time-varying depreciation rate increasing from 2.5 percent to 4 percent, as explained in Section II. Figure 4b shows the public capital stock for a depreciation rate that—based on implicit scrapping rates for the total capital stock—gradually increases from 3 percent to 6 percent over the period 1960–2001. It also shows the public capital stock for a depreciation rate that is constant over time and equal to the average implicit scrapping rate empirically observed for the United States (3.5 percent for government assets).

The figure reveals that these two estimates exhibit dynamics that differ significantly from those of the BEA series and of the benchmark estimate. The reason is that the estimate assuming a constant depreciation rate considerably overestimates the growth of public capital in the 1990s, whereas the estimate relying on a steeper time profile of the depreciation rate underestimates the growth of public capital in the 1970s and early 1980s. All in all, the benchmark assumption on the depreciation rates based on empirically observed implicit scrapping rates for the United States seems to be the most sensible approach, given the alternatives.

[26]Detailed results are available upon request. These tests are carried out for the growth rates of the series, because unit root tests indicate that the public capital stock series are nonstationary.

This section has shown the effect of various assumptions on public capital stock estimates. Visual inspection and tests for equality between the benchmark estimates and reference series reported by the BEA reveal that the benchmark estimates of public capital seem quite plausible. The properties of the capital stock estimates based on the various assumptions are further explored in the next section, where these estimates are used as explanatory variables in regressions according to the so-called production function approach.

V. Evidence for the Production Function Approach

This section uses alternative capital stock measures to estimate the elasticity of output with respect to public capital, applying the production function approach. The first subsection presents regression results for the various capital stock estimates for the United States and, thus, supplements the sensitivity analysis presented in the previous section. The second subsection presents regression results for 22 OECD countries for capital stock estimates from different sources (i.e., the author's benchmark estimates, estimates from national sources, and estimates from OECD, 1997).

The production function approach is one of three main approaches that have been employed in the empirical literature on the macroeconomic effects of public capital. The other approaches are the cost function approach and the vector autoregressive (VAR) approach.[27] This study concentrates on the production function approach because it is the simplest of the three and because it remains widely applied in the empirical literature. Even recent studies relying on the less restrictive VAR approach often provide additional estimates according to the production function approach as a benchmark against which to evaluate the VAR results (Ligthart, 2002). Thus, there is ample evidence against which the regression results of this study can be evaluated.

The starting point of empirical applications relying on the production function approach is an aggregate production function of the form $Y = AF(L, K^P)$, where Y is output, L is labor input, K^P is the private capital stock, and A is an index representing multifactor productivity. This expression is extended by including the public capital stock, K^G, so that the aggregate production function can be written as $Y = AF(L, K^P, K^G)$. Using the Cobb-Douglas form, the production function can be expressed as $Y = AL^{\alpha}(K^P)^{\beta}(K^G)^{\gamma}$, where the parameters α, β, and γ can be interpreted as elasticities of output with respect to the factors of production. The empirical literature in general investigates the production function in intensive form, imposing constant returns to scale in private inputs, in order to avoid multicollinearity problems.[28] Dividing both sides by K^P yields the production function

[27]See Sturm, Kuper, and de Haan (1998) for a discussion of the three approaches and an overview of empirical studies.

[28]Multicollinearity among the regressors is frequently cited as a problem in the empirical literature estimating production functions and cost functions for individual countries (see the survey by Sturm, Kuper, and de Haan, 1998). An alternative way to deal with this problem is to exploit the cross-sectional dimension of the data and estimate panel data models instead of carrying out individual-country regressions. This is done in the second subsection.

in intensive form, $Y/K^P = A(L/K^P)^{\alpha}(K^P)^{\alpha+\beta-1}(K^G)^{\gamma}$. Finally, imposing constant returns to scale in private inputs ($\alpha + \beta = 1$) and taking the natural logarithm yields the regression equation that is typically employed in empirical studies relying on the production function approach:

$$\left(y_t - k_t^P\right) = a_0 + a_1 t + \alpha\left(l_t - k_t^P\right) + \gamma k_t^G + \varepsilon_t, \tag{10}$$

where y_t is the log of GDP, k_t^P is the log of private capital, l_t is the log of labor input, k_t^G is the log of public capital, and ε_t is a disturbance term. As in Aschauer (1989b), a constant, a_0, and a time trend, t, are introduced as a proxy for multifactor productivity.

An important issue raised by this specification is that of potential nonstationarity of the variables included in the regression. A large number of alternative unit root tests are designed to discriminate between nonstationary and stationary processes. We use the popular Dickey and Fuller (1979, 1981) test to determine whether the model variables are nonstationary. This test is applied to a total of 188 variables that are used in the individual-country regressions reported in the two subsections.[29] The test results indicate that the vast majority of variables is nonstationary; only 13 variables appear to be stationary in level. Among the variables that are nonstationary, many seem to be integrated of order 2 rather than of order 1, according to the test results. Yet it is well known that unit root tests have low power to discriminate between unit root and near unit root processes. This problem is especially pronounced for small samples. One way to alleviate the problem is to make use of the cross-sectional dimension of the data and apply panel unit root tests to the variables. The results of two popular panel unit root tests suggest that the variables are integrated to the first order.[30] The individual-country regressions are also based on this finding.

Since the variables are nonstationary, estimation of equation (10) makes sense only if the variables are co-integrated. There are many alternative methods that can be used to test for co-integration. In this paper, the two-step procedure initially proposed by Engle and Granger (1987) is applied. In the first step, equation (10) is estimated by ordinary least squares (OLS); in the second step, an augmented Dickey-Fuller test is performed on the residual sequence $\{\hat{\varepsilon}_t\}$ to determine whether it has a unit root. The null hypothesis of the test is that the residual sequence has a unit root or, in other words, that the variables are not co-integrated. Note that the critical values of the Engle-Granger test are larger in absolute value than those of

[29]Detailed results are available upon request. We use the so-called augmented Dickey-Fuller test, which is asymptotically valid in the presence of serial correlation in the errors, including two additional lags of the respective variable. The test is first carried out for the variables in levels. In this case, the test equation includes a constant and a linear time trend. If the null hypothesis of a unit root cannot be rejected at the 5 percent significance level, the test is also carried out for the variables in first differences. In this case, the test equation includes a constant. If the null hypothesis of a unit root still cannot be rejected, the test is also carried out for the variables in second differences. Small-sample critical values are derived from MacKinnon (1991).

[30]We use the panel unit root tests proposed by Levin, Lin, and Chu (2002) and Im, Pesaran, and Shin (2003). Detailed results are available upon request.

the standard Dickey-Fuller test, because the sequence $\{\hat{\varepsilon}_t\}$ is generated from a regression equation. The appropriate small-sample critical values can be derived from MacKinnon (1991). If the variables are co-integrated, the OLS estimates from regression (10) will be superconsistent. However, if the variables are not co-integrated, estimating equation (10) in levels might give rise to a spurious-regression problem. Instead, it is appropriate to difference the variables in order to induce stationarity. In this case, the regression equation takes the form

$$\left(\Delta y_t - \Delta k_t^P\right) = a_1 + \alpha\left(\Delta l_t - \Delta k_t^P\right) + \gamma \Delta k_t^G + \upsilon_t, \tag{11}$$

where Δ is the first-difference operator and υ_t is a disturbance term.

Estimates of Public Capital Productivity for Alternative Capital Measures

This subsection supplements the sensitivity analysis and reports estimates of the elasticity of output with respect to public capital for the United States, using the various public capital stock estimates discussed in Section IV. Real GDP and employment data are taken from the OECD Analytical Database; public and private capital stocks are estimates calculated using the alternative assumptions on the initial capital stock and the depreciation rate considered in Section IV. The private capital stock is the sum of private nonresidential and private residential capital. The private capital stocks are estimated using exactly the same method as that used to estimate government capital. The benchmark estimation assumes that the depreciation rate increases from 4.25 percent to 8.5 percent for private nonresidential assets, although it is assumed to be constant for private residential assets (1.5 percent). Two alternatives are considered regarding the time profile of the depreciation rate. The first alternative assumes that the depreciation rate increases from 3 percent to 6 percent over the period 1960–2001 for all assets. The second alternative assumes that the depreciation rate is time-invariant and—based on average implicit scrapping rates for the United States—is equal to 6 percent for private nonresidential capital and 1.5 percent for private residential capital.

Table 4 reports estimates of the elasticity of output with respect to public capital for the United States using the various public capital stock estimates. Figures in bold indicate whether the elasticity is taken from the model in levels or first differences, depending on the results of the test for co-integration. Table 4 shows that the Engle-Granger test statistic is lower in absolute value than the 5 percent critical value in all cases. The test results, thus, suggest that there is no co-integration among the variables and, by implication, that the production function as formalized in equation (10) does not constitute a stable long-run relationship. As a consequence, all estimates are based on the model in first differences.[31] The estimated

[31]As the usual inference procedures are inappropriate in the presence of nonspherical disturbances, the *t*-statistics reported in Tables 4 and 5 are based on the Newey and West (1987) heteroskedasticity and autocorrelation consistent covariance estimator.

Table 4. Elasticities of Output with Respect to Public Capital for the United States for Alternative Capital Stock Measures[1]

Model	Sample	Elasticity		Cointegration Test[2]
		Levels	Differences	Test statistic
(a) United States: Alternative assumptions on the initial capital stock				
Benchmark	1960–2001	0.839	**0.788***	–2.792
		(7.144)	(2.885)	
Steady-state approach	1960–2001	0.878	**0.850***	–3.558
		(5.060)	(2.354)	
4% growth rate[3]	1960–2001	0.795	**0.747***	–2.565
		(7.681)	(2.996)	
BEA capital stock	1960–2001	0.991	**0.830***	–1.619
		(6.486)	(3.083)	
(b) United States: Alternative assumptions on depreciation rates				
3%–6% for all assets	1960–2001	0.736	**0.372**	–2.552
		(7.162)	(1.112)	
Constant[4]	1960–2001	0.723	**0.723***	–3.518
		(5.355)	(2.672)	

Source: Author's estimates and calculations.

[1]*t*-values based on the Newey and West (1987) heteroscedasticity and autocorrelation consistent covariance estimator in parentheses. *denotes statistical significance at the 5 percent level. Elasticities in the column labeled "Levels" are set in bold if the variables are co-integrated; otherwise, the elasticity for the corresponding model estimated in first differences reported in the column labeled "Differences" is set in bold.

[2]This paper uses the so-called augmented Engle-Granger test, performing an augmented Dickey-Fuller test on the residuals of the model in levels (equation (10)). The MacKinnon (1991) 5 percent critical value for rejection of the null hypothesis of no co-integration is –4.795 for the sample 1960–2001.

[3]The growth rate of investment in the period 1914–59 is assumed to be 4 percent a year. The benchmark estimates, instead, rely on historical growth rates of investment reported by the U.S. Bureau of Economic Analysis (BEA).

[4]The depreciation rate is assumed to be constant over time. Based on the average implicit scrapping rate over the period 1960–2000, the depreciation rate is assumed to be 3.5 percent for government capital, 6 percent for private nonresidential capital, and 1.5 percent for private residential capital.

elasticities of output with respect to public capital are statistically significant in all cases but one and have the expected (positive) sign. The estimates are, however, very high, ranging from 0.37 to 0.85 and thus implying what Aschauer (1995, p. 91) has called "supernormal" returns to public capital. However, this is a common finding in the literature. For example, Sturm and de Haan (1995, p. 64) report elasticities of output with respect to public capital of 0.67 and 0.71 for the United States for a model estimated in first differences similar to equation (11).

The regressions reveal another problem associated with the production function approach: the difficulty of interpreting the regression results as representing the input elasticities of output. The estimates of the coefficient on labor input are larger than 1 in all regressions, implying a negative elasticity of output with respect

to private capital.[32] Taken at face value, these results suggest not only that public capital is much more productive than private capital but also that private capital is not productive at all. Again, such a finding is not uncommon in the literature. For example, Sturm and de Haan (1995, p. 64) report that their estimates for the United States imply a large negative elasticity of output with respect to private capital. There are several reasons to be cautious in interpreting this finding: (1) the assumed functional form of the production function may not be an appropriate description of the data, (2) the exogeneity assumptions underlying the production function approach may not be satisfied, and (3) the regressions may suffer from a small-sample problem. One way to deal with this problem is to exploit the cross-sectional dimension of the data and estimate panel data models instead of carrying out individual-country regressions. The panel estimation reported in the next subsection yields a more plausible value for the elasticity of output with respect to private capital.

Estimates of Public Capital Productivity for 22 OECD Countries

Table 5 reports estimates of the elasticity of output with respect to public capital for the 22 OECD countries for equation (10) using public capital stock estimates from three alternative sources: first, the author's estimates; second, estimates from national sources; and third, estimates from OECD (1997). The table reveals the main advantage of the author's capital stock estimates, which is that equation (10) can be estimated for 22 countries. The alternative sources allow for only 10 and 11 individual-country regressions, respectively. The estimates of the elasticity of output with respect to public capital are based on the model in first differences, as the augmented Engle-Granger test fails to reject the null hypothesis of no co-integration in all cases. Most of the coefficients are statistically insignificant for the estimations based on data from other sources. This result is similar to that of Ford and Poret (1991), who estimated models based on the production function approach for 11 OECD countries using capital stock data from an earlier volume of OECD (1997). In contrast, most of the coefficients stemming from estimations based on the author's capital stock estimates are statistically significant. However, in those cases where the elasticity of output with respect to public capital is significant, it is again quite large. Moreover, for some countries, the coefficient on labor input—not reported in the table—is larger than 1, making it difficult to interpret the estimated coefficients as parameters of a Cobb-Douglas production function.

The last row of Table 5 reports the elasticity of output with respect to public capital for a panel consisting of all 22 OECD countries. This estimate is based on the pooled regression equation

$$\Delta y_{it} = a_1 + \alpha \Delta l_{it} + \beta \Delta k_{it}^{P} + \gamma \Delta k_{it}^{G} + \upsilon_{it}, \tag{12}$$

[32]These estimates are not reported in Table 4. Detailed results are available upon request. The coefficient of labor input is statistically significant in all cases.

Table 5. Elasticities of Output with Respect to Public Capital for 22 OECD Countries for Capital Stock Estimates from Different Sources[1]

	Author's Estimates		National Source		OECD (1997)	
Country	Sample	Elasticity	Sample	Elasticity	Sample	Elasticity
Australia	1960–2001	0.270	1960–2001	0.115	1970–95	–0.280
		(1.836)		(1.034)		(–0.949)
Austria	1965–2001	0.437*				
		(2.705)				
Belgium	1960–2001	0.224*	1970–2000	0.471*	1970–95	0.601*
		(2.477)		(3.365)		(3.043)
Canada	1960–2001	0.478*	1961–2001	0.396*	1970–95	0.381
		(2.350)		(3.594)		(0.988)
Denmark	1960–2001	0.478*			1971–92	0.321
		(2.894)				(1.707)
Finland	1960–2001	0.313			1970–96	–0.330
		(1.074)				(–1.433)
France	1965–2001	1.106*				
		(3.872)				
Germany	1960–2001	0.028	1960–90	0.786*	1970–94	–0.311
		(0.202)		(4.920)		(–1.341)
Greece	1961–2001	0.167			1970–89	0.759*
		(0.283)				(4.421)
Iceland	1967–2001	–0.014				
		(–0.035)				
Ireland	1960–2001	–0.116				
		(–0.414)				
Italy	1960–2001	0.190	1980–2000	–0.206	1979–96	–0.375
		(0.791)		(–0.591)		(–0.693)
Japan	1960–2001	0.835*	1990–2000	–0.091		
		(2.710)		(–0.271)		
Netherlands	1969–2001	0.535*				
		(2.524)				
New Zealand	1962–2001	–0.050	1987–97	–0.849		
		(–0.199)		(–0.761)		
Norway	1960–2001	0.024	1970–2000	0.017	1978–94	0.054
		(0.128)		(0.165)		(0.496)
Portugal	1960–2001	–0.492				
		(–1.131)				
Spain	1961–2001	0.388*	1980–95	–0.087		
		(2.088)		(–0.820)		
Sweden	1960–2001	0.293*			1979–94	0.019
		(2.378)				(0.037)
Switzerland	1960–2001	0.503*				
		(3.809)				
United Kingdom	1960–2001	0.175			1986–96	0.407
		(1.264)				(0.981)
United States	1960–2001	0.788*	1960–2000	0.830*		
		(2.885)		(3.082)		
OECD (Panel)	1960–2001	0.223*				
		(4.835)				

Source: Author's estimates and calculations.

[1]*t*-values based on the Newey and West (1987) covariance estimator in parentheses. *denotes statistical significance at the 5 percent level. The model is estimated in first differences in all cases, based on the results of Engle-Granger tests for cointegration.

where the index *i* represents the countries considered. Compared with equation (11), the pooled regression does not impose the restriction of constant returns to scale a priori. The panel analysis, unlike the individual-country regressions, should not be subject to multicollinearity problems, because it exploits the cross-sectional variation in the data. Another obvious advantage of the panel regression over the individual-country regressions is the much larger number of observations: While most individual-country regressions have 37 degrees of freedom, the panel regression has more than 800 degrees of freedom. Consequently, the parameters of interest can be estimated much more precisely.

The estimation results for the panel model indicate that the elasticities of output with respect to public capital (0.22), private capital (0.26), and labor input (0.57) take on sensible values.[33] According to these estimates, private and public capital are roughly equally productive. Finally, based on the estimation results, it is not possible to determine whether the production function exhibits constant returns to scale in private inputs only or in all inputs. A test for constant returns to scale in private inputs and a test for constant returns to scale in all inputs both fail to reject the null hypothesis.

VI. Conclusion

This paper provides new estimates of the government net capital stock for 22 OECD countries for the period 1960–2001. This data set has several attractive features compared with existing alternatives. First, the same estimation approach is used across all countries, ensuring a maximum degree of international comparability. Second, the estimates are based on investment data (compiled by the OECD) that are homogenous across countries. Third, these investment data have been compiled according to the 1993 System of National Accounts, whereas the data reported in OECD (1997) were compiled according to the old 1968 System of National Accounts. Finally, the data set covers 22 countries and 42 years for each country and is, thus, much larger than any existing alternative.

The public capital stock estimates reveal that public capital–GDP ratios have tended to decline in most OECD countries since the late 1970s. The estimates further show that a considerable disparity exists in the public capital endowment of OECD countries, even though some convergence has taken place in the past two decades. A sensitivity analysis regarding the main estimation assumptions suggests that the benchmark estimates of public capital stocks are reasonable when they are compared with estimates from official sources for the United States. Regression results based on the production function approach indicate that the elasticity of output with respect to public capital is positive and statistically significant in many countries. This is confirmed by the results of a simple panel regression showing that public capital is productive, on average, in OECD countries.

The capital stock data may prove useful for applied research on the macroeconomic effects of public capital in several respects. First, they may allow an assessment of public capital productivity for countries for which such estimates

[33]The elasticities of output with respect to private capital and labor input are not reported in the table. All elasticities are significant at the 5 percent level. Detailed results are available upon request.

have not been available in the past because of a lack of public capital stock data. Moreover, they can be used in the estimation of models based on dynamic panel regression techniques aiming to uncover the average effect of public capital in OECD countries. Finally, they can be employed in the estimation of VAR models—the most commonly used framework in the recent empirical literature.

REFERENCES

Aschauer, David Alan, 1989a, "Does Public Capital Crowd Out Private Capital?" *Journal of Monetary Economics,* Vol. 24 (September), pp. 171–88.

———, 1989b, "Is Public Expenditure Productive?" *Journal of Monetary Economics,* Vol. 23 (March), pp. 177–200.

———, 1995, "Infrastructure and Macroeconomic Performance: Direct and Indirect Effects," in *The OECD Jobs Study: Investment, Productivity and Employment* (Paris: Organization for Economic Cooperation and Development).

Berndt, Ernst R., and Bengt Hansson, 1991, "Measuring the Contribution of Public Infrastructure Capital in Sweden," NBER Working Paper No. 3842 (Cambridge, Massachusetts: National Bureau of Economic Research).

Boskin, Michael J., Marc S. Robinson, and Alan M. Huber, 1987, "New Estimates of State and Local Government Tangible Capital and Net Investment," NBER Working Paper No. 2131 (Cambridge, Massachusetts: National Bureau of Economic Research).

Broadberry, Stephen, 1988, "The Impact of the World Wars on the Long Run Performance of the British Economy," *Oxford Review of Economic Policy,* Vol. 4 (Spring), pp. 25–37.

Demetriades, Panicos O., and Theofanis P. Mamuneas, 2000, "Intertemporal Output and Employment Effects of Public Infrastructure Capital: Evidence from 12 OECD Economies," *Economic Journal,* Vol. 110 (July), pp. 687–712.

Dickey, David A., and Wayne A. Fuller, 1979, "Distribution of Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association,* Vol. 74 (June), pp. 427–31.

———, 1981, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica,* Vol. 49 (July), pp. 1057–72.

Eichengreen, Barry, and Albrecht Ritschl, 1998, "Winning the War, Losing the Peace? Britain's Post-War Recovery in a West German Mirror," CEPR Discussion Paper No. 1809 (London: Center for Economic Policy Research).

Engle, Robert F., and Clive W. J. Granger, 1987, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica,* Vol. 55 (March), pp. 251–76.

European Commission, International Monetary Fund, Organization for Economic Cooperation and Development, United Nations, and World Bank, 1993, *System of National Accounts 1993* (Brussels: Commission of the European Communities).

Evans, Paul, and Georgios Karras, 1994, "Is Government Capital Productive? Evidence from a Panel of Seven Countries," *Journal of Macroeconomics,* Vol. 16 (Spring), pp. 271–79.

Feinstein, Charles H., 1972, *National Income, Expenditure and Output of the United Kingdom, 1855–1965,* Studies in the National Income and Expenditure of the United Kingdom, Vol. 6 (Cambridge, United Kingdom: Cambridge University Press).

Fitz Gerald, John, Ide Kearney, and Edgar Morgenroth, eds., 1999, "National Investment Priorities for the Period 2000–2006," ESRI Policy Research Series, No. 33 (Dublin: Economic and Social Research Institute).

Florio, Massimo, 2001, "On Cross-Country Comparability of Government Statistics: Public Expenditure Trends in OECD National Accounts," *International Review of Applied Economics,* Vol. 15 (April), pp. 181–98.

Ford, Robert, and Pierre Poret, 1991, "Infrastructure and Private-Sector Productivity," *OECD Economic Studies,* Vol. 17 (Autumn), pp. 63–89.

Fuente, Angel de la, and Rafael Doménech, 2002, "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make? An Update and Further Results," CEPR Discussion Paper No. 3587 (London: Center for Economic Policy Research).

Giersch, Herbert, Karl-Heinz Paqué, and Holger Schmieding, 1992, *The Fading Miracle: Four Decades of Market Economy in Germany* (Cambridge, United Kingdom: Cambridge University Press).

Gramlich, Edward M., 1994, "Infrastructure Investment: A Review Essay," *Journal of Economic Literature,* Vol. 32 (September), pp. 1176–96.

Holtz-Eakin, Douglas, 1993, "State-Specific Estimates of State and Local Government Capital," *Regional Science and Urban Economics,* Vol. 23 (April), pp. 185–209.

Im, Kyung So, M. Hashem Pesaran, and Yongcheol Shin, 2003, "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics,* Vol. 115 (July), pp. 53–74.

International Monetary Fund, 2001, *Government Finance Statistics Manual 2001* (Washington).

Jacob, Verughese, Subhash C. Sharma, and Richard Grabowski, 1997, "Capital Stock Estimates for Major Sectors and Disaggregated Manufacturing in Selected OECD Countries," *Applied Economics,* Vol. 29 (May), pp. 563–79.

Kamps, Christophe, 2004, "New Estimates of Government Net Capital Stocks for 22 OECD Countries 1960–2001," IMF Working Paper 04/67 (Washington: International Monetary Fund).

———, 2005, "The Dynamic Effects of Public Capital: VAR Evidence for 22 OECD Countries," *International Tax and Public Finance,* Vol. 12 (August), pp. 533–58.

Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu, 2002, "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," *Journal of Econometrics,* Vol. 108 (May), pp. 1–24.

Ligthart, Jenny, 2002, "Public Capital and Output Growth in Portugal: An Empirical Analysis," *European Review of Economics and Finance,* Vol. 1 (September), pp. 3–30.

Luetzel, Heinrich, 1977, "Estimates of Capital Stock by Industries in the Federal Republic of Germany," *Review of Income and Wealth,* Vol. 23 (March), pp. 63–78.

MacKinnon, James G., 1991, "Critical Values for Cointegration Tests," in *Long-Run Economic Relationships: Readings in Cointegration,* ed. by Robert F. Engle and Clive William John Granger (Oxford: Oxford University Press).

Maddison, Angus, 1982, *Phases of Capitalist Development* (Oxford: Oxford University Press).

———, 1987, "Growth and Slowdown in Advanced Capitalist Economies: Techniques of Quantitative Assessment," *Journal of Economic Literature,* Vol. 25 (June), pp. 649–98.

———, 1991, *Dynamic Forces in Capitalist Development: A Long-Run Comparative View* (Oxford: Oxford University Press).

———, 1995, "Standardised Estimates of Fixed Capital Stock: A Six Country Comparison," in *Explaining the Economic Performance of Nations: Essays in Time and Space,* ed. by Angus Maddison, Economists of the Twentieth Century series (Brookfield, Vermont: Edward Elgar).

Mitchell, Brian R., 1975, *European Historical Statistics 1750–1970* (London: Macmillan).

Mittnik, Stefan, and Thorsten Neumann, 2001, "Dynamic Effects of Public Investment: Vector Autoregressive Evidence from Six Industrialized Countries," *Empirical Economics,* Vol. 26 (May), pp. 429–46.

Munnell, Alicia H., 1990, "How Does Public Infrastructure Affect Regional Economic Performance," in *Is There a Shortfall in Public Capital Investment?* ed. by Alicia Munnell, Federal Reserve Bank of Boston Conference Series No. 34 (Boston: Federal Reserve Bank of Boston).

Newey, Whitney K., and Kenneth D. West, 1987, "Notes and Comments: A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* Vol. 55 (May), pp. 703–08.

O'Mahony, Mary, 1996, "Measures of Fixed Capital Stocks in the Post-War Period: A Five-Country Study," in *Quantitative Aspects of Post-War European Economic Growth,* ed. by Bart van Ark and Nicholas Crafts (Cambridge, United Kingdom: Cambridge University Press).

Organization for Economic Cooperation and Development, 1997, *Flows and Stocks of Fixed Capital 1971–1996* (Paris).

———, 1998, *National Accounts: Main Aggregates, 1960–1996, 1998 Edition, Vol. 1* (Paris).

———, 2001, *Measuring Capital: A Manual on the Measurement of Capital Stocks, Consumption of Fixed Capital and Capital Services* (Paris). Available via the Internet: www.oecd.org/dataoecd/61/57/1876369.pdf.

———, 2002a, *Measuring the Information Economy* (Paris).

———, 2002b, *National Accounts of OECD Countries: Main Aggregates, 1989–2000, 2002 Edition, Vol. 1* (Paris).

———, 2002c, *Purchasing Power Parities and Real Expenditures: 1999 Benchmark Year, 2002 Edition* (Paris).

Ritschl, Albrecht, 2003, "Germany: Modern Germany," in *The Oxford Encyclopedia of Economic History,* Vol. 2, ed. by Joel Mokyr (Oxford: Oxford University Press).

———, and Mark Spoerer, 1997, "Das Bruttosozialprodukt in Deutschland nach den amtlichen Volkseinkommens—und Sozialproduktsstatistiken 1901–1995," *Jahrbuch für Wirtschaftsgeschichte,* No. 2, pp. 27–54.

Seitz, Helmut, 2001, "Infrastructure Investment and Economic Activity: Theoretical Issues and International Evidence," in *Investing Today for the World of Tomorrow,* ed. by Deutsche Bundesbank (Berlin: Springer-Verlag).

Sturm, Jan-Egbert, and Jakob de Haan, 1995, "Is Public Expenditure Really Productive? New Evidence for the USA and the Netherlands," *Economic Modelling,* Vol. 12 (January), pp. 60–72.

Sturm, Jan-Egbert, Gerald H. Kuper, and Jakob de Haan, 1998, "Modelling Government Investment and Economic Growth on a Macro Level: A Review," in *Market Behaviour and Macroeconomic Modelling,* ed. by Steven Brakman, Hans van Ees, and Simon Kuipers (London: Macmillan Press).

U.S. Bureau of Economic Analysis (BEA), 1999, "Fixed Reproducible Tangible Wealth in the United States 1925–94" (Washington: U.S. Department of Commerce).

Voss, Graham M., 2002, "Public and Private Investment in the United States and Canada," *Economic Modelling,* Vol. 19 (August), pp. 641–64.

IMF Staff Papers
Vol. 53, No. 1

The IMF and Russia in the 1990s

JOHN ODLING-SMEE*

This paper explains the IMF's impact on economic policies in Russia, focusing on where the IMF made a difference. The Russian economic and political leadership essentially determined economic policies. The IMF's influence was modest: it had a limited impact on overall fiscal policy and major structural reforms, but it had a positive impact on monetary policy. A tougher position on fiscal policy in 1996–98 might have produced a better outcome. The G-7's concerns weakened the IMF. However, the IMF played a major role in transferring knowledge about macroeconomic policymaking and implementation. [JEL E60, E62, P20, P34, P35]

A remarkably wide range of views has been expressed about the role of the IMF in Russia in the 1990s. Some people believe that most of the economic problems of Russia were attributable to the economic reform policies of successive Russian governments and that those policies were in turn pressed on them by the IMF.[1] Others believe that the IMF's influence in Russia was minor. Both Russian and foreign commentators can be found in both camps. The purpose of this paper is to explain the role of the IMF in Russia from the perspective of a senior member of the IMF staff team working on Russia.

The IMF's role can be assessed in various ways. First, one could assess the quality and relevance of the advice and technical assistance given by the IMF.

*John Odling-Smee was Director of the European II Department of the IMF from 1992 to 2003. He thanks Gérard Bélanger, José Fajgenbaum, Yegor Gaidar, Oleh Havrylyshyn, Ernesto Hernández-Catá, David Owen, Emil Sunley, Poul Thomsen, Siddharth Tiwari, Thomas Wolf, and Mikhail Zadornov for many helpful comments. Aidar Abdychev and Franziska Ohnsorge provided valuable assistance.

[1]For example, Stiglitz (2002) wrote, "The IMF strategy did not work: GDP in post-1989 Russia fell, year after year" (p. 146).

Second, one could identify where and how the IMF affected economic policy in practice. This would include questions such as whether Russia would have pursued different policies had there been no IMF activity and whether the policies actually implemented were those recommended by the IMF. Third, one could assess how beneficial or harmful the impact of IMF advice and technical assistance was on economic policy and outcomes in Russia. This would build on the conclusions of the second approach by assessing the quality of the impact the IMF actually had.

This paper uses the second approach. It does not attempt the first approach, namely, an analysis of the quality and relevance of IMF policy advice. Much of the debate about economic policies in Russia, especially among non-Russian commentators, has been about this issue.[2] There has been much less discussion about the impact the IMF actually had, which is the focus of the second approach. The third approach—assessing the quality of IMF efforts—is also not used here in a systematic way.

The paper is intended to contribute to a better understanding of one particular aspect of economic policymaking in Russia in the 1990s. It also places on the public record information about the IMF's activities at a very important time.

I. Context

Economic Policy Views of the Russian Authorities

Throughout the 1990s, the Russian government's public position was to pursue liberal economic reforms that would lead to the rapid establishment of a market economy. The intellectual leadership for these policies was provided by a relatively small group of economists and others who were committed to market reforms. The starting point was the government formed by President Boris Yeltsin in November 1991, in which Yegor Gaidar was the deputy prime minister responsible for economic reforms. Gaidar, who remained in government until December 1992 (for the last six months as acting prime minister) and returned for four months as first deputy prime minister in September 1993, and Anatoly Chubais, deputy prime minister from November 1991 until early 1996 and again in 1997, were the main leaders of the reform group.

However, there was opposition within the government to economic reforms. The strongest strand of alternative thinking came initially from those who saw the primary economic function of government as supporting production. Their views reflected the Soviet system rather than a market economy, in which enterprises adjust to changing conditions—especially changing relative prices and profitability and the collapse in demand for military products—within a framework set by the government. They did not wish to return to central planning, but they sought to use the government's fiscal and regulatory powers, and the allocation of Central Bank of Russia (CBR) credits, to encourage the continuation and growth of production, even in sectors that objectively had little future. Various deputy prime ministers

[2]See Owen and Robinson (2003) for a discussion of economic policy issues since 1998 by IMF staff.

(for example, Oleg Soskovets, Oleg Lobov, and Vladimir Kadannikov) and ministers dealing with sectors of the economy formed the heart of this group. Viktor Chernomyrdin, prime minister from December 1992 until March 1998, was in the same camp by background, but as prime minister, he had to listen to the president and take account of reformers' views. Yeltsin's own position, as revealed by his speeches and appointments, suggested that he was close to the reformers, although he sometimes turned against them when politics required. Parliament also had an influence on economic policy through the legislation it passed. The Supreme Soviet, until it was dissolved in September 1993 (and its buildings stormed in October), and the Duma, from its inauguration in December 1993, were more often on the side of protecting production than supporting liberal economic reforms.

The opposition to the reformers not only had different ideas about economic reforms; in many cases, they also had more operational experience and knew how to make the bureaucracy work in pursuit of their interests or to impede the reformers' policy initiatives. This was especially true in 1992, when the reformers underestimated the bureaucratic and other obstacles in their way and placed too much faith in Yeltsin's ability to make things happen. Gradually, the reformers learned how to get things done, with Chubais emerging as a particularly skilled operator.

Economic policymaking suffered not only from differences of view within the government but also from the weakness of the state itself. A vacuum had been created by the collapse of the coercive communist system and the Communist Party network, which together had ensured that government policy was implemented. This vacuum was filled by various business and interest groups as well as by criminals. Many businessmen and some criminals had already learned in the Soviet period how to exploit loopholes in the system; for example, by buying goods at controlled prices and selling them at black market prices, and by bribing officials whenever necessary. Regional governors and legislators openly defied the federal authorities. Yeltsin was not able to retain the enormous authority he had acquired from standing up to the leaders of the coup in August 1991 and precipitating the breakup of the Soviet Union, and he was handicapped by poor health, especially after 1995. He tried to retain what influence he could; for example, by agreeing to transfers of power to regional levels on an ad hoc basis.[3] But neither he nor the government ever had enough broad-based popular support to pursue economic reforms without at least tacit support from powerful groups in Parliament, the regions of Russia, and the economic sector.

Business leaders had a very important influence on economic policy in the early 1990s. Enterprises were finding it difficult to adapt to the new environment, in which they had to react to financial signals (e.g., changes in profitability and prices) rather than instructions from economic planners. They sought protection or financial assistance from the government and the CBR to help them through the adjustment phase. While in public they usually asked for temporary assistance, it is likely that in practice many hoped for permanent assistance or even an end to the market reforms that were causing them so much difficulty. Viktor Gerashchenko,

[3]Klyamkin and Shevtsova (1999) have written about Yeltsin's great weakness in practice, despite his considerable constitutional powers.

chairman of the CBR from July 1992 to November 1994, was initially sympathetic to the view that the CBR and the government should deemphasize the goal of cutting inflation and instead support production, but he gradually moved away from this position. Beginning in the mid-1990s, a small group of businessmen, known as the oligarchs, became increasingly powerful. They consolidated their power dramatically in the loans-for-shares operation in late 1995 and as a result of their support for Yeltsin's reelection campaign in 1996. They used this power to influence both the economic policy agenda, including the preparation of legislation, and the implementation of policy. They sponsored members of the Duma and even government ministers.[4] They made special arrangements with the government about the amount of taxes they would pay. They were not opposed to market-oriented reforms in principle and, indeed, provided a bulwark against a return to a planned, state-owned economy. But they sought to ensure that their own businesses were protected from competition and had favorable financial dealings with the government (through sales to government and, for enterprises with government stakes, dividend payments, as well as through privatization and tax deals). When other businessmen saw what the oligarchs were able to get away with, they all tried to emulate them, further weakening the state.[5]

The lack of consensus in favor of liberal economic reforms among the Russian leadership and the weakness of the Russian state were the two most important reasons why economic reforms did not progress more smoothly, despite the stated policies of the government.

IMF Policy Advice

The IMF's basic view of the necessary economic reforms was that it was important to move as quickly as possible with all the key changes, especially macroeconomic stabilization, liberalization, and privatization.[6] The IMF recognized that many reforms would take years to complete, especially the construction of the legal infrastructure for a market economy, privatization and restructuring of large enterprises, and the creation of a market system for banking and finance.[7] But this was not seen as a reason for postponing the main stabilization and liberalization measures, with some specific exceptions, such as gradually rather than suddenly raising the prices of public utilities and housing rents, and reducing export taxes

[4]For a short time, from August 1996 to March 1997, Vladimir Potanin, a leading oligarch, was first deputy prime minister responsible for economic policy.

[5]See Nagy (2000) and Graham (2002) for more discussion of the weak state. Nagy focused primarily on the implications for economic reforms and the economy.

[6]Because this paper is concerned with the IMF's impact on Russia's economic policies, references to the IMF's views, advice, position, and so on refer to the official position adopted by the IMF in its relations with Russia. Different views existed within the IMF on various issues, sometimes reflecting the different responsibilities and perspectives of the executive board, management, and individual departments. Some of these differences are mentioned in the paper.

[7]Camdessus (1992) noted that the process of basic economic transformation would take many years. Odling-Smee and Wolf (1994) classified reforms according to whether they could be undertaken over short-, medium-, or long-term time horizons.

on energy and other basic imports.[8] Thus, IMF views were similar to those of Russian reformers and many foreign observers. As a consequence, the IMF was often in the position of endorsing policies proposed by reformers, with little need to push for major changes in them.[9]

Although this paper does not attempt to analyze the quality and relevance of IMF policy advice, a few brief comments regarding the case for rapid reforms may help set the scene.[10] First, the disintegration of the old central planning system had reached such an advanced stage by the end of 1991 that there was little scope for pursuing a gradual strategy in which the new system would be phased in as the old one was dismantled. Second, in transition countries (such as the Baltic countries) that were able to implement a strategy of rapid reform, living standards rose from their low point earlier and more rapidly than in countries that reformed more slowly. The countries that sought to maintain many features of the old system, such as Belarus and Uzbekistan, appeared to avoid as sharp an initial fall in GDP as other transition countries, but they are now growing more slowly and still face some potentially very disruptive reforms.[11] Third, the many difficulties of the reform process in Russia, especially those stemming from the opposition from old vested interests and new powerful oligarchs, would not have been more manageable had the reforms been introduced more slowly. The underlying problems of the collapse of the old system and the weak state would still have presented major challenges. For these and other reasons that are discussed in more detail elsewhere, the IMF's view of the appropriate reform strategy has remained broadly unchanged despite the difficulties Russia has endured.[12]

In the early 1990s, IMF advice was directed primarily toward avoiding hyperinflation and bringing inflation down quickly to low levels, to create conditions for the resumption of growth. In those years, the IMF emphasized the necessity of tight monetary and fiscal policies to contain inflation. This contrasted with the alternative view, popular in some Russian circles, that more credit expansion was needed to stimulate output. In this view, inflation was the result of the absence of competition, not of excessive monetary growth. While the IMF was concerned in the early years about both monetary and fiscal policies being too loose, once inflation had been brought under control by the end of 1995, the IMF focused more sharply on

[8]The earliest official statement of the IMF's views on economic reforms can be found in IMF and others, (1990). See, especially, Chapter V (pp. 16–19) for a discussion of the speed and sequencing of reforms.

[9]Vladimir Mau, a leading reformer, explained the situation as follows: "A good part of the 'IMF conditions' were developed in Moscow, not Washington. Russian politicians are the ones who initiated many of these conditions" (Mau, 2000, p. 108).

[10]The case for a gradualist strategy as presented by a number of authors is discussed in Åslund (2002). See, especially, Chapter 3 on Strategic Policy Choices.

[11]The experience of China and other Asian transition countries is not very relevant to Russia and other former Soviet area countries, because political systems in the Asian countries did not collapse. More important, the Asian economies were largely agricultural, and the share of large industrial enterprises in total production and employment was much smaller than in Russia and other Soviet area economies. See Sachs and Woo (1994), Mau (2000), and Kalra and Sløk (2001).

[12]See Anderson, Citrin, and Lahiri (1995); Citrin (1995); Hernández-Catá (1995); Odling-Smee (1996); Fischer and Sahay (2000); and Owen and Robinson (2003) for detailed discussions of these issues by IMF staff.

fiscal policy. This reflected the difficulties the Russian government encountered in attempting to tighten fiscal policy, difficulties that culminated in the financial crisis of August 1998. After 1999, with macroeconomic stability reestablished and the economy growing at a good rate, the IMF's policy advice included such issues as how to manage monetary policy in the face of large increases in international reserves and medium-term fiscal policy in a world of fluctuating oil prices and revenues.

A summary of the broad policy issues on which the IMF advised cannot capture the full range and depth of the many detailed aspects of macroeconomic policy formulation and implementation on which the IMF gave advice and technical assistance. Missions from IMF headquarters and staff in the Moscow office discussed with their Russian counterparts in the CBR, the ministry of finance, and other government departments all the relevant details, such as how to set monetary policy to achieve a desired reduction in inflation, the design of instruments for controlling the money supply, the creation of a treasury in the finance ministry, and the meaning and importance of quasi-fiscal deficits. (See the Appendix for a description of the channels of IMF advice.) Whereas the reformers were often familiar with the broad policy issues they discussed with the IMF, they and the technicians in the CBR and government did not necessarily know how to give operational meaning to their policy choices or how to handle many of the issues that arose. The IMF played a major role in transferring knowledge about such issues, all in the context of market-oriented macroeconomic policies aimed at producing low inflation and a favorable environment for growth.[13]

While the official policy was to request technical and financial assistance from the IMF, and many individual Russian officials benefited from discussions with their IMF counterparts, it was not easy for Russians, individually or collectively, to seek and accept advice from IMF staff members. Russia saw itself as a strong, independent nation with centuries of history as a major European power and recent experience as a technically sophisticated world superpower. Many Russians, especially older and more nationalistic ones, felt that it was humiliating to have to negotiate with the IMF for loans and to adjust economic policies to IMF requirements. IMF staff tried to be sensitive to this need for country "ownership" of economic policies, which is an issue in the IMF's relations with all sovereign member states, albeit magnified in the case of Russia. But there was inevitably some tension in the relationship between the IMF and Russia arising from Russia's self-image and history.[14]

[13]In addition to the direct channels through which it gave advice and transferred knowledge, the IMF also made a contribution through papers and participation in conferences. These had an indirect impact on Russian policymakers through their influence on western governments, other international organizations, and independent analysts and, hence, their contribution to the creation of a broad international consensus behind the policies advocated by the IMF.

[14]Yevgeny Primakov, prime minister from September 1998 to May 1999 and previously a senior member of the Soviet foreign policy establishment, was representative of those who found it most difficult to accept the IMF's role in Russia. In the Russian version of his memoirs published in 2001, he said that he was boiling inwardly when he thought that Gérard Bélanger, IMF mission chief in 1999–2002, was "mor-

The IMF's policy and technical advice focused on macroeconomic issues and the structural and institutional reforms needed in the monetary and fiscal areas to design and execute macroeconomic policy. But the IMF was fully aware that the major—and most difficult—parts of the transition to a market economy were the structural and institutional reforms of the whole economic system. This was noted as early as April 1992 by Michel Camdessus, IMF Managing Director from 1987 to 2000, who placed the "speedy adoption of a legal institutional framework" first out of five broad reform areas.[15] Macroeconomic stabilization was the third reform area on his list. The IMF took every opportunity to emphasize, in discussions with Russian authorities and in public, the importance of the broad structural and institutional reform agenda. A failure to make sufficient progress with structural and institutional reforms would not only threaten macroeconomic stability, it would prevent the resumption of economic growth. An example was the initial response of many unreformed enterprises to the sharp changes in demand and relative prices at the beginning of the transition. Rather than make internal adjustments and seek new products or markets, as enterprises that operate under strict financial discipline in market economies would do, they sought credits, tax breaks, and subsidies from the authorities, thereby complicating the task of macroeconomic policy.

Although the IMF recognized the paramount importance of structural and institutional reforms, it was not able to give much practical policy advice in these areas. Its mandate was to focus on macroeconomic issues, while other international financial institutions (especially the World Bank) focused on the structural and institutional areas.[16] Nevertheless, the IMF worked closely with the World Bank to ensure that the government included important structural reforms in its programs. The World Bank was responsible for the discussions with the government about the details of the reforms. The IMF's role was to assist in identifying broad reform priorities and to insist that key reforms were included in the programs before the IMF would agree to support them financially. This helped the government to obtain support from the Duma and other groups and to explain its policies at home and abroad.

The high point of the IMF's involvement in advising on structural and institutional reforms was the preparation, with the authorities and the World Bank, of the government's medium-term economic program in 1995–96. Many discussions took place and papers were prepared over the six or more months leading up to the approval of the program and of a loan, under the Extended Fund Facility (EFF),

alizing"; this passage did not appear in the English version (Primakov, 2004). Earlier, in a November 1998 meeting with Jorge Márquez-Ruarte, IMF mission chief in 1997–99, Primakov had accused the IMF of dictating to Russia and acting like Bolsheviks.

[15]Camdessus (1992).

[16]Thus, the World Bank's Country Assistance Evaluation for Russia noted that the Bank, IMF, and European Bank for Reconstruction and Development (EBRD) were asked by their shareholders to work together to facilitate Russia's transition: "The Bank was entrusted with the responsibility of encouraging and overseeing structural reforms. . . . A series of strategy documents established that, beyond its complementary assistance in support for IMF-funded stabilization efforts, the Bank's focus would be on helping build the institutions of a market economy, developing the private sector, and mitigating the social costs of transition" (World Bank, 2003, p. xi).

by the IMF's Executive Board in March 1996. The outcome of this work was the structural reform agenda summarized in Box 1. The IMF's contribution was primarily in the banking, finance, and fiscal areas, although it also tried to keep up the momentum of efforts in other areas. In practice, with the exception of some critical measures in the fiscal area, the IMF did not subsequently insist on the implementation of the original structural agenda. The authorities made rather slow progress in most areas, which raises the question of whether the whole effort was misconceived. The counterargument, which was the prevalent (although not the only) view in the IMF at the time, was that a reform agenda such as this one would strengthen the hand of reformers in government and increase the likelihood that reforms would be implemented. With the benefit of hindsight, one can add that, although structural reforms may have been slow to get off the ground in the 1990s, the acceleration after 2000 probably owes something to the earlier preparations.

Box 1. Structural Reform Agenda, 1996–98

Banking system: Conduct examinations of the financial condition of the largest banks; introduce legislation on bank bankruptcy; begin bank rehabilitation; improve transparency of banks' disclosed financial information, including introduction of International Accounting Standards (IAS); strengthen contract enforcement; facilitate entry of foreign banks.
Payments system: Develop a strategy for improving the payments system; introduce same-day settlement for large transfers among the financial centers; improve the efficiency of clearing centers; introduce electronic settlements.
Bank supervision: Introduce additional, tightened, and more transparent reporting requirements based on IAS; widen bank audits; introduce penalties for noncompliance with prudential requirements.
Capital markets: Introduce securities legislation; adopt minimum standards for market participants; amend the law on bankruptcy and the civil code to strengthen creditor rights.
Foreign trade: Reduce the average and maximum import tariff; remove export duties; eliminate duty exemptions; remove mandatory registration of export contracts; remove import quotas on alcohol; harmonize excise taxation of domestic and imported goods.
Budget system and process: Consolidate extrabudgetary expenditures and revenues; improve expenditure management, including through monitoring successive stages of the spending process with strengthened budgetary accounting; strengthen internal auditing.
Energy taxation and revenue measures: Improve efficiency and effectiveness of the tax system applicable to oil production; revise tax system of gas and power sectors; increase revenues from remaining state ownership in the energy sector.
Intergovernmental fiscal relations: Redesign the tax-sharing system; limit State Tax Service responsibilities to collecting taxes; conduct an overall evaluation of intergovernmental relations.

Box 1 *(concluded)*

Agriculture: Introduce a land register and facilitate land registration; increase transferability of land; untie the provision of inputs from state procurement; introduce procedures for bankruptcy of agricultural enterprises; introduce a land tax to finance social services at the local government level; develop an agricultural tax strategy for inclusion in the federal budget; target transfers to the regions or specific activities to link them to reforms.
Urban land and real estate: Introduce a land cadastre; facilitate purchase of leased land; improve transparency of land use rights, strengthen loan collateral rights, and introduce mortgage legislation.
Privatization and corporate governance: Set targets for privatization revenues; improve information on enterprises to be privatized; use privatization receipts to clear tax arrears; apply uniform and transparent rules to cash privatizations; broaden monitoring and disclosure of financial information of enterprises.
Natural monopolies: Eliminate tax exemptions for natural monopolies; allow natural monopolies to discontinue services to nonpaying customers; improve governance and efficiency of each of the four natural monopolies (gas, electric power, telecommunications, and rail); prepare a restructuring/privatization plan for each natural monopoly; strengthen regulatory institutions to oversee tariff regulation; introduce and foster competition among competitive segments of the system.
Social safety net: Delink pension and unemployment benefits from the minimum wage; amend legislation to introduce a two-tier pension system; strengthen payroll tax revenue collections through broadening the tax base; strengthen eligibility criteria for the pension system; conduct an assessment of the effectiveness of the social safety net in three poor regions; trim benefits of the employment fund, increase its centralized share and transfer its revenue collection activities to the State Tax Service; strengthen auditing of social funds; initiate a new pension law; introduce a basic poverty benefit; adopt a transparent cash management system for the pension fund; introduce a program of support for long-term unemployed based on the budgetary constraints of the employment fund.
Health care: Audit the medical insurance fund; define a package of minimum benefits in relation to resource availability at the regional and federal levels; offer a choice of health insurance companies and health service providers; introduce oversight and regulation of insurance companies.
Education: Introduce a system of user fees and capitation formulas for the 1997 budget; facilitate entry of new educational institutions; develop an action plan for ensuring minimum standards for school achievement and testing; make the allocation of research funds competitive.

It must be admitted that the IMF did not have a significant influence on structural and institutional reforms other than those related to the design and implementation of monetary and fiscal policies (e.g., the introduction of money market instruments, the design of the treasury, and tax administration). This partly reflected the staff's limited expertise and the IMF's macroeconomic mandate. More important, the IMF had to focus its efforts on macroeconomic, especially fiscal, issues

during the difficult years of the 1990s and had little leverage left to persuade the government to implement its structural reform program. If macroeconomic policies were being successfully implemented, the IMF was usually unwilling to delay disbursements because of problems on the structural side, other than problems with measures (such as those related to revenue collection) that directly affected monetary or fiscal policy.

There were two prominent exceptions to this practice. In January 1997, the IMF postponed a disbursement because of delays in repealing a resolution introducing quotas on alcohol and vodka imports and in differentiating natural gas prices to reflect transportation costs. And in November 1999, a disbursement was delayed (in fact, was never made at all), ostensibly because the authorities did not fully implement various structural measures relating to the CBR's foreign subsidiaries and its internal accounting and control arrangements. As discussed later, the real reason for this delay was more complicated. These two instances notwithstanding, the Russian authorities understood that the IMF was unlikely to delay disbursements for reasons not directly connected to macroeconomic policy problems. Partly because of this situation, they gave lower priority to implementing the structural and institutional measures in their programs than to macroeconomic measures.

There was quite a contrast between the IMF's limited influence on structural and institutional reforms and the public perception in Russia and elsewhere that the IMF was responsible for policy advice in these areas and, indeed, for what actually did or did not happen. This perception was reasonable, given that the programs the IMF supported with loans (notably the 1996 program) had a large structural component and speeches by IMF leaders stressed the importance of structural reforms. The media focus on the drama of IMF-Russia negotiations added to the perception that the IMF had influence across a much wider range of policies than was the case. The role of the World Bank, the European Bank for Reconstruction and Development (EBRD), and others who were actively involved in advising on structural and institutional reforms received less public attention. While a general perception that the IMF was responsible for advice and outcomes in the structural and institutional area was understandable, it is surprising that some people who knew what the IMF's role actually was have continued to hold the IMF accountable for problematic outcomes.[17]

Influence of the G-7

The transition of Russia from a planned to a market economy and the associated transition to a democracy was, of course, of great concern to the rest of the world, especially the major western countries. There was some discussion at the begin-

[17]For example, Stiglitz (2002) has strongly criticized the IMF for insisting on rapid privatization, whatever the price, and for not being concerned about whether market institutions, such as a competitive system, were in place first. This led, he argued, to IMF acquiescence in the loans-for-shares scheme and to corruption, state capture, and growing income inequalities. (See below for more on the loans-for-shares scheme.) As Chief Economist at the World Bank in 1997–2000, Stiglitz knew that the World Bank rather than the IMF was in the lead in advising on these issues and that the IMF had little influence. Similarly, Lopez-Claros (2002), who represented the IMF in Moscow in 1992–95, deplored the fact that the IMF did not stop the loans-for-shares scheme.

ning of the 1990s about a so-called "Grand Bargain." This was the name given to the idea that the major western countries would give Russia large sums of money in exchange for deep and sustained economic reform efforts.[18] The main problem with the Grand Bargain was that the western countries were reluctant to spend the large sums of money that were being suggested without evidence that major reforms were being undertaken. They were unwilling even to commit to the idea because of a deep skepticism about whether the Soviet Union—and, subsequently, Russia—really was ready to implement such reforms. Instead, they turned to the international financial institutions (IFIs), which had significant resources as well as the capacity to apply conditionality to any lending.

By not providing large-scale financial support to Russia themselves, the G-7 were effectively assigning two tasks to the IFIs that were sometimes contradictory.[19] First, the IFIs were expected to perform their usual function of lending on the basis of economic policies or projects that met the normal standards of the institutions. In insisting that Russia align its economic policies with the IMF, the G-7 caused Russia to take the IMF's views more seriously than it might otherwise have done. Second, the G-7 sometimes wanted the IFIs to lend in order to show support for the government (or for President Yeltsin), even when economic policies or projects were not up to the standard normally required by the IFIs. The advocates in G-7 countries of these two tasks were sometimes split between institutions, with the finance ministry being most concerned that the normal standards of the IFIs should be upheld and the foreign ministry wishing to show support for Russia (or not, according to the circumstances) based on broader political and bilateral considerations. The president or prime minister would synthesize these views. The distinction was clear in the United States, where the Treasury Department, to which the U.S. executive director of the IMF reports, supported strict IMF conditionality and the U.S. State Department focused on the political and bilateral issues. The White House generally supported the Treasury position, although sometimes it allowed its impatience to see more IMF lending to overcome Treasury concerns about weak Russian policies. One incident in December 1993 illustrates this: Vice President Al Gore complained in public in Moscow about IMF policies being too tough, to the annoyance of the Treasury Department, which supported the IMF's attempts to tighten macroeconomic policies (see Box 2).

This situation was unsatisfactory for both the Russians and the IMF. Russian governments thought that the west should be prepared to make available large sums of money without difficult policy conditions attached. This was a small price to pay for ensuring that Russia never returned to communism or confrontation with the west. The Russians resented the need to negotiate with the IMF and the

[18]See Allison and Yavlinsky (1991).

[19]In fact, quite large sums of money were provided by the G-7, although most of it was tied to particular imports (suppliers' credits) or expenditures (e.g., housing for troops returning from Germany), or was debt relief. In 1992 and 1993, all bilateral creditors provided a total of US$36 billion of financial assistance, with the G-7 accounting for the bulk of it (Citrin and Lahiri, 1995, Table 7.5).

Box 2. U.S. Attitudes Toward the IMF After the 1993 Duma Elections

The U.S. political leadership was concerned about the outcome of the Duma elections in December 1993, in which the nationalists and communists emerged with a majority and liberal, pro-western groups were in the minority. On a visit to Moscow a few days after the elections, Vice President Al Gore stood alongside Viktor Chernomyrdin at a press conference and said that the IMF had been "slow to recognize some of the hardships that are caused by some of the conditions that have been overly insisted upon in the past." According to Strobe Talbot (2003), the U.S. Deputy Secretary of State who was accompanying Gore, the vice president already had some sympathy for the view that the IMF was too tough on Russia, dating back at least to a discussion in Washington in September 1993 between Chernomyrdin, who asked for IMF leniency, and Treasury Undersecretary Lawrence Summers, who defended IMF conditionality. Talbot himself appeared to call for a softer line when he said in a press briefing later in December that the administration wanted to promote "less shock and more therapy." Both the Russian reformers and the U.S. Treasury disliked what Gore and Talbot said, because it undercut the efforts of the reformers, notably First Deputy Prime Minister Yegor Gaidar and Deputy Prime Minister and Finance Minister Boris Fedorov, to persuade the government, especially Chernomyrdin, to pursue tight monetary and fiscal policies in 1994.[1] In February 1994, the U.S. Senate Committee on Banking held hearings on the impact of IMF and World Bank policies on the Russian economy. Witnesses included Jeffrey Sachs—who had been advising the Russian government and who argued that the IMF should permit a bigger fiscal deficit that would be financed by issuing bonds and more borrowing abroad—as well as Russia experts Marshall Goldman, Peter Reddaway, and Jude Wanniski.

It is difficult to say how great an impact these public statements by, and discussions among, U.S. political leaders had on IMF-Russia relations. On the one hand, the Treasury continued to support the IMF's attempt to persuade the Russians to follow strict monetary and fiscal policies. On the other hand, the IMF leadership and Treasury were aware that the political mood in Washington and other G-7 capitals could swing so much that the good intentions of finance ministries to support the IMF could be swept aside. Camdessus was, therefore, willing to settle, during his visit to Moscow in March 1994, for an economic program that was less strict than he would have liked, to preempt any undue pressure from the G-7. Perhaps more important, the U.S. statements produced a hardening of the Russian position. One aspect of this was the resignation in January 1994 of both Gaidar and Fedorov from the government, which chose to pursue the less strict of the macroeconomic policy options under consideration. Gaidar (1997) believed that the U.S. criticism of the IMF for excessive tightness weakened the IMF's ability to put its weight behind those, including himself, who were arguing for the radical disinflation option.

[1]In his memoir, Talbot (2003) noted that he infuriated "both Russian liberals and my colleagues at Treasury" and regretted the language he used, while also arguing that the hardships accompanying reforms were backfiring against the reformers.

delay this involved. The IMF was unhappy because it feared that the political concerns the G-7 had about making IMF money available to Russia, even when the normal economic policy conditions were not met, might cause the G-7 to push the IMF to lend with insufficiently strong conditionality.[20] Although this never happened in an overt way, an atmosphere was sometimes created (for example, at the end of 1993 as described in Box 2 and in 1996–98 as discussed in Section III) in which the IMF felt that it should err on the side of supporting weak policies rather than interrupt disbursements altogether. The Russians understood the situation and sometimes saw negotiations with the IMF as a charade since they believed that the G-7 would ultimately insist that the IMF go ahead with the loans.

Among the G-7 countries, the United States effectively determined the collective position. It was unusual for other countries to initiate approaches different from that of the United States or to strongly oppose U.S. initiatives, except on matters relating to Russia's debt to G-7 countries. Here Germany, by far the largest creditor, was able to modify U.S. policy positions, which tended to favor more generous debt relief than Germany could accept. The Russian authorities understood the preeminent position of the United States and frequently sought U.S. support to put pressure on the IMF.[21]

To sum up, during the years before the 1998 crisis, while the G-7 helped to create an important role for the IMF in Russia, its political and bilateral interests caused a weakening of the IMF's influence. Many IMF staff members believed that it would have been better for the Russian economy and for the consistency of the IMF's position across its member countries if it had insisted on more ambitious policies, especially fiscal policy, before agreeing to lend, and on fuller implementation of agreed-upon policies before resuming lending after a break because of poor implementation. A stronger position would also have embraced some conditions on the effective implementation of structural measures. Managing Director Michel Camdessus and First Deputy Managing Director Stanley Fischer, who were keen to avoid any formal proposals from the G-7, were more prepared to make disbursements on the basis of the best promises and prior actions that the Russians could present. Apart from the G-7 angle, they, together with some staff members, also believed that the IMF's ability to influence Russian economic policies was greater when there was an active dialogue without lengthy interruptions in lending. The Executive Board was split, with some members doubting the wisdom of continuing to support weak Russian policies, although this rarely led to a rupture in the Board's traditional consensus. After 1998, and especially after the newspaper stories suggesting money laundering scandals and the renewed fighting in Chechnya in 1999, the G-7 took a much harder line on IMF lending. This was reflected in the IMF's behavior.

[20]IMF here means not only the management and staff but also many members of the Executive Board, albeit a minority, who did not like to see the majority pushing the Board in this direction.

[21]Talbot (2003) described several occasions when Russian leaders pressed U.S. leaders to ensure IMF financial assistance. Primakov (2004) commented that progress toward an agreement with the IMF was determined largely by the United States. He also noted that President Yeltsin spoke to German Chancellor Gerhard Schröder about speeding up an agreement with the IMF, and French President Jacques Chirac promised Primakov that he would "convince Camdessus."

II. Where the IMF Made a Difference

This section begins with an overview of the IMF's involvement in macroeconomic policy in the 1990s, followed by a discussion of the IMF's impact in the fiscal, monetary, and exchange rate policy areas. The treatment is selective, focusing on only a few key issues.

Overview

By the time the IMF had its first discussions with the new Russian government in November 1991, Gaidar and his team already had a clear idea of the economic reforms they wanted to introduce.[22] IMF staff teams met with the new government almost from the beginning and sought to provide technical advice, especially on the importance and modalities of designing a consistent macroeconomic policy framework with a quantified program of monetary and fiscal measures to achieve specific macroeconomic objectives. In February 1992, IMF staff helped the Russian authorities prepare a policy program that the government and the CBR then published. The policy program did not contain a quantified monetary program because of a lack of consensus between the government and the CBR, problems arising from other countries' use of the ruble, and inadequate CBR databases. Subsequent discussions in May, June, and July 1992 about a more detailed program that the IMF could support with financial assistance, after Russia became a member on June 1, built further on this first program. However, the policies themselves were essentially those chosen by the authorities. At the time of the February program, the IMF had little substantive disagreement with the Russians, but in the summer negotiations, the government was already moving toward less tight monetary and fiscal policies in response to pressures from the political opposition to support production. The IMF staff opposed this move, but in the end decided to support the program sought by the Russians in the interest of providing some financial assistance, as desired (and as had been promised) by the G-7.[23]

From about April 1992 on, political opposition to the government's strict monetary and fiscal policies grew sharply. The IMF was criticized by reformers in the government and their western supporters for not providing financial assistance on a larger scale and much earlier than the first US$1 billion approved by the Executive Board on August 5, 1992 (Table 1). The main reason for the schedule was that time was needed to make Russia a member of the IMF and also to agree on an economic program with the authorities.[24] The IMF limited the amount of the loan

[22]Gaidar (1999) describes the debates among the reform team.

[23]One example of a difference on which the IMF gave way to the Russians was the oil price. The IMF argued for a larger increase in the domestic price of oil than the Russians were willing to implement, on the grounds that this was a move toward market pricing and would allow the government to collect more revenues from the oil sector and achieve a more ambitious target for the fiscal deficit.

[24]Under an accelerated mechanism for admitting new members, Russia joined the IMF on June 1, 1992. Difficulties in agreeing on monetary policy and some other issues delayed agreement on an economic program until early July. The Executive Board meeting to approve the first loan was held on August 5.

Table 1. IMF Lending to Russia, by Type of Arrangement
(In millions of SDRs)

Type of Arrangement	IMF Board Approval Date	Amount Approved	Amount Disbursed
Stand-By Arrangement (First Credit Tranche)	August 5, 1992	719	719
Systemic Transformation Facility (First Tranche)	June 30, 1993	1,078	1,078
Systemic Transformation Facility (Second Tranche)	March 22, 1994	1,078	1,078
Stand-By Arrangement	April 11, 1995	4,313	4,313
Extended Fund Facility	March 26, 1996	6,901	5,105
Extended Fund Facility Augmentation	July 20, 1998	1,248	0
Supplemental Reserve Facility	July 20, 1998	3,992	675
Compensatory and Contingency Financing Facility	July 20, 1998	2,157	2,157
Stand-By Arrangement	July 28, 1999	3,300	471
Total		24,786	15,596

Source: *IMF Survey,* various issues.

because Russia did not have complete control over its monetary policy as a result of the uncoordinated monetary policies among all the countries that used the ruble. The IMF had, in May, sought to make the ruble area work by brokering an agreement among ruble area countries, but Russia was not willing to share power and the agreement never materialized. Russia would have preferred the IMF to act decisively to end the ruble area. As explained elsewhere (Odling-Smee and Pastor, 2002), the IMF, as a cooperative institution, could not do this as long as some members wanted to retain the area, and it was too early to conclude that a successful ruble area was politically impossible.[25] Gaidar (1997, 1999), recognizing the IMF's rules-based lending policies and the constraints on the speed at which it could move, argued that the IMF was not the right instrument to handle the large-scale political problem of Russia's transition in 1992, when the situation was so unpredictable. But in the absence of major untied G-7 financial assistance, the IMF was the main instrument available.

The period from mid-1992 until the end of 1994 was one of serious conflict between reformers and their opponents, both inside and outside the Russian government. The outcome was that macroeconomic policies were less tight than the

[25]Some IMF staff members believed that a cooperative ruble area would never work, and events proved them correct. Others, including executive directors representing some major countries, believed that the importance of the ruble area for regional trade and payments flows was such that every effort should be made to make it work. In hindsight, one can say that it would have been better for Russia—and probably for the other ruble area countries, too—if the ruble area had broken up earlier than it did, perhaps even by the middle of 1992. See Odling-Smee and Pastor (2002) for further discussion of these issues.

reformers would have liked, as they were forced to bow to pressures from interest groups that sought more CBR credits and budget subsidies. The IMF encouraged the authorities to implement reasonably tight monetary and fiscal policies, and the annual programs agreed to in the early months of 1993 and 1994 (which were supported by loans under the new Systemic Transformation Facility [STF]) were in line with this.[26] The Russian authorities were in agreement with these programs, mainly because they had persuaded the IMF that anything more ambitious in terms of reducing inflation and the budget deficit—however desirable—would not be feasible. As noted above, at the time of the negotiations for the 1994 program, U.S. authorities were openly expressing concern that the IMF was being too tough (see Box 2). This stance weakened the IMF's negotiating position, which was already more fluid than in normal situations because precedents had not yet been set for the strength of programs to be backed by STF loans. While this situation produced a less ambitious disinflation program for 1994 than the reformers and the IMF would have liked, any other outcome for the year would have been unlikely, because the pressures for looser monetary and fiscal policies that built up during the year (including unbudgeted expenditures on agriculture and the Northern Territories) would still have existed.

It is possible to see the 1992–94 period as one during which the reformers, supported by the IMF, were gradually able to move toward macroeconomic stabilization. Sergei Dubinin (1995), who was acting finance minister from early 1994 until October 1994, talked about a hardening of financial policies in a phased manner, to allow the continuation of some inflationary financing until enterprises were able to adjust to the new market conditions. A more accurate description might be that the reformers were doing the best they could to stabilize the economy, winning battles whenever possible and otherwise retreating. The 21 percent collapse in the ruble exchange rate on Black Tuesday (October 11, 1994) was a direct consequence of the loose monetary policies. Dubinin (1995) noted that this collapse signified the failure of the gradualist strategy. It certainly led to a significant change in attitude among Russian leaders outside the reformist group. Chernomyrdin, in particular, was determined to pursue policies that would avoid a repetition of such an event.

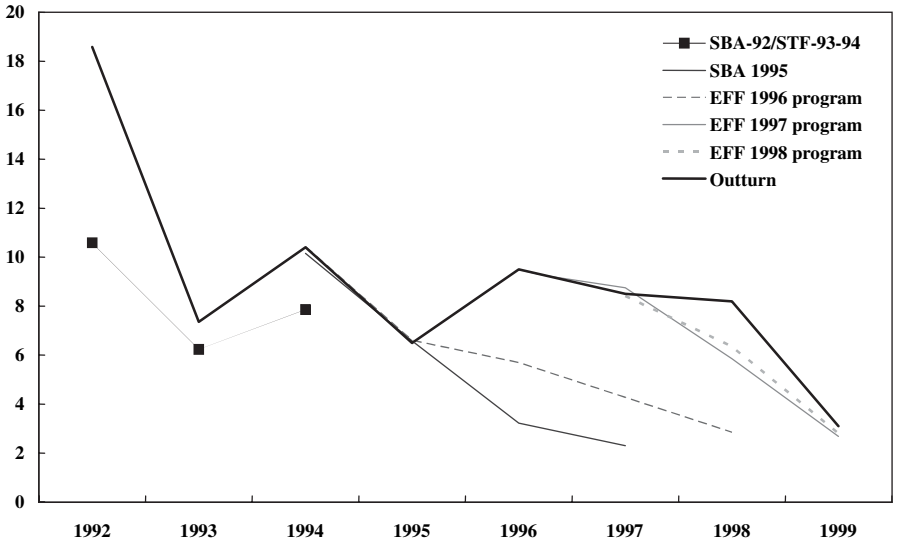
The failure of the gradualist or (depending on one's point of view) ad hoc approach to macroeconomic stabilization was brought home not only by Black Tuesday but also by the continued decline in output. The argument that loose monetary policy was needed to protect output was shown to be incorrect. These facts, as much as the IMF's repeated recommendations for tightening monetary and fiscal policies, were behind the tighter policies implemented in 1995.

The new leaders set about preparing an economic program for 1995.[27] The IMF worked closely with them and supported the program with a one-year Stand-By

[26]The IMF introduced the STF in April 1993 to enable it to lend to countries that were making the transition from centrally planned to market economies before they were able to implement strong macroeconomic stabilization and reform programs.

[27]Tatiana Paramonova was appointed acting chair of the CBR in October 1994, after Viktor Gerashchenko resigned. Anatoly Chubais, who was already a deputy prime minister, took over responsibility for macroeconomic policy issues from Alexander Shokhin, who also resigned.

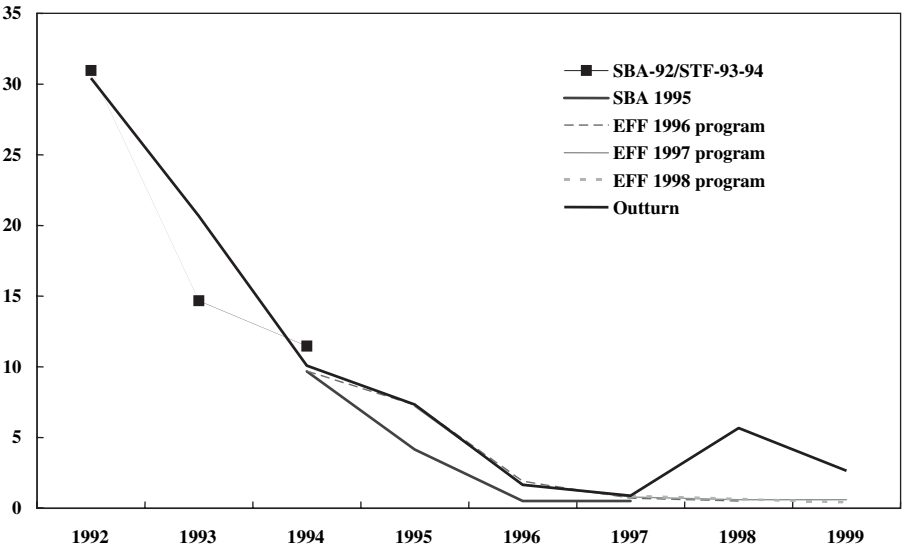
Figure 1. General Government Deficit
(*In percent of actual GDP*)



Sources: IMF staff estimates based on data from Russian authorities.

Notes: EFF=Extended Fund Facility; SBA=Stand-by Arrangement; and STF=Systemic Transformation Facility.

Figure 2. CPI Inflation
(*In percent, average monthly rate*)



Sources: IMF staff estimates based on data from Russian authorities.

Notes: EFF=Extended Fund Facility; SBA=Stand-by Arrangement; and STF=Systemic Transformation Facility.

Arrangement (SBA) approved by the Executive Board on April 11, 1995. The macroeconomic stabilization objectives were more demanding than in earlier years, with monthly inflation targeted to fall to 1 percent by July and then stay there, and the general government deficit set at 6.5 percent of GDP (Figure 1). No direct CBR lending to the budget was permitted after the first quarter. More important, the program was successfully implemented, although inflation did not fall quite as much as planned (Figure 2). All the quantitative targets were met in 9 out of 12 months of the program, and the full IMF loan—about US$6.5 billion—was disbursed (see Table 2).[28]

One of the consequences of the successful monetary tightening was the strong appreciation of the nominal exchange rate beginning in April 1995.[29] The authorities were concerned about this, partly because of its impact on competitiveness and the real value of the population's foreign currency savings but also because of fears of a sharp reversal, with economic and political consequences even more severe than those following Black Tuesday. Therefore, they introduced an exchange rate corridor in July to contain appreciation and any subsequent depreciation pressures.

Although the actual fiscal deficit in 1995 was in line with the target, revenues and expenditures were not as high as originally intended. As in most years in the 1990s, revenues were below the projected level, mainly because well-connected enterprises had been able to reduce their tax liabilities, either by obtaining formal tax exemptions or by informal arrangements. The government responded by holding expenditures below budgeted levels, which subsequently created serious problems. First, by allowing wage and pension arrears to build up, the government only postponed the problem until 1996, when special loans had to be raised to pay off the arrears (in part because it was an election year). Second, the fact that the government was not paying its bills—for procurement of goods and services, as well as wages and pensions—was a major reason for the financial indiscipline throughout the economy. It contributed to the withholding of tax revenues by taxpayers who were not themselves being paid by the government and to the growth of the pernicious practice of offsets.[30]

The loans-for-shares scheme introduced in 1995 also had lasting adverse economic consequences.[31] Starting in May 1995, the IMF expressed concern about the lack of transparency and the unsatisfactory arrangements for competitive bidding. The scheme, as it was endorsed by the president in October, looked more reasonable on paper, but there remained scope for collusion, nontransparency, and asset-stripping. IMF staff warned about this in October and, in December, criticized the

[28]The IMF monitored progress on a monthly basis, as described later.

[29]With inflation still high, the real appreciation was especially rapid.

[30]Offsets were arrangements between the government and individual enterprises whereby the government's unpaid obligations to an enterprise and an equivalent amount of the enterprise's unpaid tax liabilities were cancelled. Enterprises were thus encouraged not to pay their taxes in the expectation of future offsets. The IMF pushed for many years for the practice to be banned.

[31]Under this scheme, banks controlled by oligarchs made loans to the government on the security of attractive but undervalued state assets (e.g., oil enterprises), then took possession of the assets when the government failed to repay the loans.

Table 2. IMF Lending to Russia, by Year

Year	Disbursements	Repayments	Outstanding (end of year)
		(In millions of SDRs)	
1992	719	0	719
1993	1,078	0	1,797
1994	1,078	0	2,876
1995	3,594	0	6,470
1996	2,588	359	8,698
1997	1,467	360	9,806
1998	4,600	674	13,732
1999	471	3,101	11,102
2000	0	2,189	8,913
2001	0	2,998	5,915
2002	0	1,148	4,767
2003	0	1,356	3,411
		(In billions of U.S. dollars)[1]	
1992	1.0	0	1.0
1993	1.5	0	2.5
1994	1.5	0	4.2
1995	5.5	0	9.6
1996	3.8	0.5	12.5
1997	2.0	0.5	13.2
1998	6.2	0.9	19.3
1999	0.6	4.2	15.2
2000	0	2.9	11.6
2001	0	3.8	7.4
2002	0	1.5	6.5
2003	0	1.9	5.1

Source: IMF.
[1]Converted from SDRs at the exchange rate at the time (annual average and end-year).

implementation of the scheme.[32] However, the IMF did not consider postponing disbursements on this account for the general reasons relating to structural reforms noted earlier, namely, the IMF's macroeconomic mandate and modest expertise in structural issues, and the need to focus its limited influence on macroeconomic policy issues.[33]

Despite these problems with arrears and loans-for-shares (and others not discussed here), the IMF judged the 1995 program to have been broadly successful as a major step toward macroeconomic stability. From the middle of 1995, the IMF staff discussed with the authorities a medium-term economic program that the IMF could support with a three-year loan under the Extended Fund Facility (EFF). The

[32]These concerns and criticisms were voiced directly to the authorities and to all IMF members in papers for the Executive Board.

[33]Given the very strong political backing of the scheme, it is doubtful whether the IMF, or the World Bank for that matter, could have had much influence on its actual implementation. See Freeland (2000) for a full account that emphasizes the political context.

program that was approved by the Executive Board in March 1996 envisaged further disinflation and reductions in the general government deficit to 4 percent, 3 percent, and 2 percent in 1996, 1997, and 1998, respectively (see Figure 1, where the deficits are shown as a percentage of actual GDP as now estimated, rather than GDP as estimated at the time). However, performance under the EFF was not as good as under the SBA.

The fiscal situation deteriorated sharply during 1996, partly as a direct consequence of the buildup to the presidential elections, which brought promises of more expenditures from the president and an increased capital outflow. To keep the exchange rate within the band, interest rates had to be raised sharply, which added significantly to the government's interest expenditures. Moreover, revenue shortfalls worsened, for two main reasons. The first was the reduced tax discipline as a result of the government's own expenditure arrears, noted above. Second, the powerful oligarchs, who helped secure Yeltsin's reelection, reduced their tax payments. It was no coincidence that nearly all the main enterprises with tax arrears at the end of 1996 were oil or gas companies.

New hope was kindled by Yeltsin's hard-hitting State of the Nation speech in March 1997 and his appointment of Anatoly Chubais and Boris Nemtsov as first deputy prime ministers. These moves appeared to signal a willingness to tackle the oligarchs and restore some of the authority of the state. In a speech in Moscow in April 1997, Camdessus (1997) strongly urged the authorities to establish the proper role for the state in a market economy, including collecting taxes, meeting the government's own financial obligations, and avoiding excessively close relationships with large enterprises. But tax collection did not improve as the intense struggle between the oligarchs and the government and among the oligarchs themselves progressed. The Emergency Tax Commission that had been set up in October 1996 was not effective at increasing revenues.

The effects of the Asian crisis began to be felt in Russia toward the end of 1997. Contagion led to a large outflow of capital in November, to which the authorities responded by tightening monetary policy. During the following months, the price of oil collapsed, with adverse consequences for Russia's balance of payments and government revenues.

The IMF's advice in the first half of 1998 was to tighten monetary policy and introduce new fiscal measures to deal with the growing fiscal gap. The former worked in January, but the very high interest rates in May did not contain the loss of international reserves. While the authorities recognized the need for further fiscal measures, they were paralyzed by the unexpected dismissal on March 23 of Chernomyrdin as prime minister and the delay until Sergei Kiriyenko was approved on April 24. With confidence not restored in financial markets after the May crisis, calls were heard for more drastic measures, including devaluation, refinancing of government debt to extend maturities, and a massive external loan. The IMF's room for maneuver was narrowed by the announcement by President Bill Clinton's office on May 31 that the United States endorsed additional financial support from the IFIs to promote stability and structural reforms. During the months that followed, until the crisis in August, the IMF focused on attempting to strengthen Russia's fiscal policy, knowing that the government was weak and could

not deliver very much. The key event here was the approval on July 20 by the Executive Board of a major loan (US$4.8 billion) in support of a new package of fiscal and other measures, many of which would not take effect until January 1999. The aim of this loan was to provide Russia with additional reserves so it could manage any short-term outflows until confidence in financial markets was restored. Unfortunately, after a temporary improvement in July, confidence continued to weaken in August; drastic measures, such as those announced on August 17, became inevitable.

After a failed effort to avert a financial crisis, it always appears that the attempt was a mistake. The increase in Russia's international reserves after the new loan was disbursed was reversed in the subsequent weeks as financial players converted rubles into dollars.[34] But it is important to assess the situation as it appeared at the time. Judged from the point of view of the balance of payments, which was not in "fundamental disequilibrium," the exchange rate was not obviously overvalued.[35] The government under Kiriyenko was firmly committed to trying to bring the fiscal situation under control, although the absence of strong support from the Duma was a matter of some concern.[36] There was a reasonable chance that confidence would return to financial markets and the crisis would be averted. As the government clearly wanted to take the chance and was prepared to take the measures necessary to strengthen policies, especially in the fiscal area, the IMF chose to support it, although there were naturally some misgivings among the executive directors and the IMF staff.[37]

Following the crisis, the IMF's main concern was that the new leadership of the government and the CBR should not move toward more inflationary policies, thereby losing the degree of macroeconomic stability that had been achieved with difficulty over the previous few years.[38] In fact, the authorities pursued reasonably tight monetary and fiscal policies, after a short injection of liquidity to the banking system that was motivated mainly by the need to protect the payments system. Although the IMF made no further disbursements under the EFF, the authorities hoped to reach an agreement with the IMF that would lead to disbursements, and this gave the IMF some influence. But more important was the message of the crisis: that the fiscal deficit must be kept under control if problems were to be

[34]Some months later, a story emerged that the money had been stolen and ended up in the bank accounts of people close to the president and government leaders. There was no truth to this story; the documents purporting to prove it were forged, and a detailed audit of the CBR's reserves found that they had been used to support the ruble, as they were intended to be, if needed.

[35]The boost to growth from the devaluation might suggest that an earlier devaluation would have led to an earlier recovery. But, as discussed below, this conclusion is not clear-cut, because higher inflation would have worked in the opposite direction. Even if it is partly correct, it does not in itself mean that the exchange rate was overvalued.

[36]The Duma did, in fact, pass a large number of measures in a few days in mid-July but rejected two key measures, one to increase personal income tax revenues and the other to restore balance to the pension fund.

[37]The comments of Fischer (2001) and Summers (2001) on Kharas, Pinto, and Ulatov (2001) constitute an excellent exposition of the case for supporting the government's policies at that time.

[38]In September, Yevgeny Primakov became prime minister and appointed Yuri Maslyukov as first deputy prime minister for economic policy. Mikhail Zadornov remained as finance minister. Gerashchenko returned as CBR chairman.

avoided. In addition, of course, the government's inability to borrow abroad and limited scope at home imposed a much tighter limit on the fiscal deficit than before the crisis.

With macroeconomic policies reasonably sound in the winter of 1998–99, the IMF was able to reach agreement with the authorities in the spring of 1999 on a program to be supported by a new SBA. Following delays caused by, among other things, a change of government and the need to investigate accounting irregularities in the CBR dating back to 1996, the Executive Board approved the SBA in July 1999. Macroeconomic policy performance was generally in line with the program, helped by higher oil prices and the recovery of growth in the economy. But stories in the press about corruption and misuse of public funds led the IMF, encouraged by the G-7, to ask for various structural improvements in accounting and control practices in the CBR, including its relationship with its subsidiaries. These were not fully implemented, which gave the IMF a reason for not making the disbursement under the SBA that was due in November, as noted earlier. However, attitudes in major western countries toward helping Russia had hardened because of the press stories and the start of the second Chechen war. As a result, the G-7 made clear to IMF management that they did not favor a disbursement at that time, and the IMF adopted an unusually strict position toward the nonimplementation of structural measures. Thus, the IMF's relations with Russia were tested (and strained, because the government partly blamed the messenger) by the G-7's sharply cooler attitude, after many years of being under stress because of G-7 pressures in the opposite direction.

As the macroeconomic situation, especially the balance of payments, continued to improve during 2000–03, it became apparent that there was no need for the IMF to lend to Russia.[39] Moreover, most government leaders were eager to see an end to negotiations over economic policies. The relationship between the IMF and the authorities, therefore, changed from one involving negotiations and detailed monitoring by the IMF to one in which the IMF was a more detached observer and advisor, as is the case with the IMF's relations with other major countries. The absence of negotiations made for an easier dialogue between the IMF and the authorities but also released the latter from the obligation to agree their economic policies with the IMF. Despite this, IMF policy and technical advice was generally welcomed and had some impact on policymaking and implementation.

A final point in this section relates to the development during the course of the 1990s of the authorities' understanding about macroeconomic policy and how to implement it. Whereas in the early and mid-1990s the IMF's messages about the importance of macroeconomic stability and how to achieve it were understood and accepted by relatively few reformers in the government and the CBR, by the early 2000s these messages were fully accepted by most important policymakers, including the leaders of the Duma. The IMF itself contributed to this development through the continuous dialogue its staff and management maintained with their counterparts at every level. This included the highest level, where the personal relationship of Camdessus and Chernomyrdin was a factor in the latter's coming to trust

[39]Developments since 1998 are discussed in detail in Owen and Robinson (2003).

IMF advice about macroeconomic policy.[40] The spread of knowledge was assisted by the gradual changing of the guard, as some of the younger reformers of the early and mid-1990s, who were most open to modern macroeconomic analysis, moved into senior positions.

But, given the deep differences in perspective within Russia about economic policy in the transition to the market, the competition for political and economic power, and the suspicious attitude toward foreign advisors inherited from Soviet times, IMF advice alone was not enough to change things. The experience of the 1990s was a necessary part of the learning process, especially the failure of the attempts in 1992–94 to stimulate output through expansionary monetary and fiscal policies, the crises of Black Tuesday (October 11, 1994) and August 1998, and the coincidence of the post-1998 recovery with the tightening of fiscal policy. But the importance of real-world experience does not diminish the major contribution that the IMF made to the transformation of the authorities' understanding of macroeconomic policy issues and implementation.

Fiscal Policy

The IMF and the reformers in the government generally agreed that macroeconomic stabilization required a reduction in the fiscal deficit from the high levels of the early 1990s. However, good intentions to reduce the deficit were thwarted in most years before 1999 by unanticipated shortfalls in revenue, and in some years by excessive spending (see Figures 1 and 3). The failure to reduce the deficit decisively was, of course, the single most important cause of the 1998 financial crisis. This section first discusses the IMF's role in the reduction of the deficit and related issues, especially the revenue shortfalls. It then briefly considers the IMF's role with respect to social expenditures, oil export duty, and the liberalization of the treasury bill market.

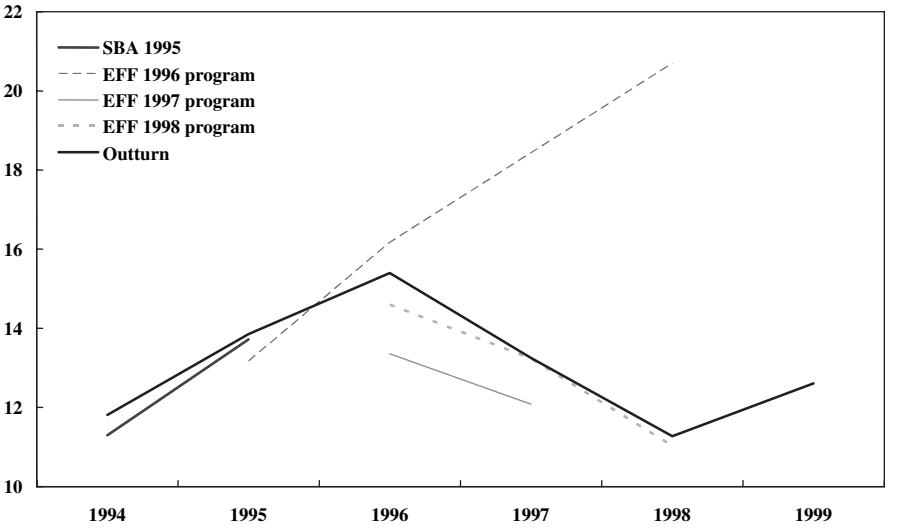
Impact of the IMF

A distinction should be made between the planned deficit and the actual outturn. Both the reformers and the IMF sought a similar path for the reduction of the deficit. Without the IMF, it is unlikely that the reformers would have been able to resist the pressure from others in government and the Duma to plan for a bigger deficit, especially in 1995 and 1996. In some years—for example, 1995—the IMF was able to agree easily with the government's plans for the deficit, but it had to argue for additional measures to raise revenues or cut expenditures, because the revenue projections were too optimistic and the planned deficit was unattainable without additional measures.[41] In 1997, the IMF had little influence on the planned deficit. It accepted the authorities' argument that only a small reduction from the

[40]There was, however, a downside to this personal relationship in Camdessus' increased willingness to accept Chernomyrdin's pleas for continued disbursements when conditions had not been met.

[41]It was common for the Duma to increase expenditures in the draft budget and to "finance" the increase by arbitrarily raising projected revenues.

Figure 3. Federal Revenues
(*In percent of actual GDP*)



Sources: IMF staff estimates based on data from Russian authorities.

Notes: EFF=Extended Fund Facility; SBA=Stand-by Arrangement; and STF=Systemic Transformation Facility.

high outturn of 1996 was possible during the following year or two and that major improvements had to be postponed.

The planned deficit was one thing; what actually happened was quite another. The IMF's impact on the implementation of fiscal policy and, hence, on the short-term fiscal deficit was rather weak. As Figure 1 shows, the failure to keep the actual deficit in line with the planned deficit, except in 1995 and 1997, was the main reason deficits remained so high before 1999. Most of the overshooting was the result of revenue shortfalls; some came from spending overruns. Both of these were in most cases under the control of the authorities, with some important exceptions, such as the higher interest expenditures before the presidential elections in June 1996 and after the Asian crisis in 1997, and the lower oil revenues in 1998.

The IMF attempted to influence the implementation of fiscal policy by monitoring it during the course of the year and urging the authorities to take corrective measures if necessary. Such encouragement had no teeth in 1992–94, because the IMF made only one disbursement in each year and had little leverage over developments in the later months of the year. For two years, beginning with the SBA approved in April 1995, a formal system of monthly disbursements was linked to monthly reviews of progress toward the targets of the program. While this enabled the IMF to check almost continuously on policy implementation, it also made it more difficult to resist proposals from the authorities to change the next month's targets because of insufficient time to take measures to meet them. For this reason, and to encourage the authorities to take more responsibility, in 1997 and 1998 the IMF switched to the more typical quarterly disbursements linked to quarterly reviews.

The IMF's ultimate lever for influencing fiscal policy implementation was its disbursements. A number of disbursements between the middle of 1996 and the middle of 1998 were delayed because of insufficient progress toward the targets or inadequate corrective measures, mostly but not always in the fiscal area. The slippages in disbursements were only minor in the second half of 1996 but became larger in the first half of 1997 (see Table 3). There was an improvement in the summer months of 1997, but this was followed by a deterioration later in the year and into 1998, as large slippages reemerged. Associated with the delays was the easing of targets, both those that had already been missed and future targets that could not be achieved (see Table 4). Although some of the easing of targets reflected factors outside the government's control, as noted above, most (especially in the fiscal area) resulted from the authorities' failure to implement the agreed-upon policies in full. The summary in Table 4 does not distinguish between targets within and outside the government's control, but a careful analysis indicates that weak implementation was often the main factor.[42] The IMF's hope at the time that close monitoring would lead to an improvement in implementation, especially of fiscal policy, was, therefore, only partly realized during the 1996–98 EFF period.

What was the IMF's impact on the level of revenues? The IMF certainly tried through various means to prevent the revenue shortfalls. It provided extensive technical assistance on tax administration and posted a senior tax administration expert in the State Tax Service to assist the authorities.[43] Beginning in August 1996, it set targets for revenue collection as conditions for disbursements of IMF loans. It frequently insisted on the introduction of new tax measures to increase the effective tax rates or to improve tax administration before agreeing to disbursements. As shown in Table 4, most reviews under the EFF involved new tax measures, ranging widely from short-term fixes such as collecting tax arrears from specific debtors to fundamental reforms of the tax system.[44] Despite all this activity, in a broad sense the IMF failed to have much impact on revenues, which continued to fall short of targets until 1999.

Assessment

A superficial reading of the impact of the IMF on the fiscal deficit might suggest that the IMF did not have a major impact. The IMF was pushing for a substantial reduction in the deficit from the very high levels at the beginning of the 1990s, and

[42]A complete list of the reasons why some reviews were delayed and the measures that were taken to enable reviews to be completed is available on request from the author: jodlingsmee@juno.com.

[43]Richard Highfield, previously a second commissioner of the Australian Tax Office, was an advisor to the Ministry of Taxation from September 1997 to August 1999.

[44]A full list of the revenue measures is available on request from the author: jodlingsmee@juno.com. The wide scope of the measures reflected the fact that many things needed to be put right, as well as uncertainty about the underlying causes of poor tax collections. Sometimes the problem seemed to be that the weak government was not able to force strong enterprises to pay their taxes; at other times, the inadequacies of the tax administration system seemed to be the problem. The government urged the IMF not to insist on higher tax rates when the taxes due at existing rates were not being collected, but the government did not itself solve the collection problem. The sharp improvement in tax collections after 1998, as the economy grew, suggests that tax evasion rather than the tax administration system had been the main problem before 1998.

Table 3. Delays in Disbursements Under the Extended Fund Facility (EFF)
(In millions of SDRs)

Expected Disbursements		Actual Disbursements		
Earliest date	Amount	Date	Amount	Review on which disbursement based
		A. Schedule approved on March 26, 1996		
March 26, 1996	233.6	March 29	233.6	Approval of EFF
April 15, 1996	233.6	May 3	233.6	First monthly
May 15, 1996	233.6	June 10	233.6	First quarterly
June 15, 1996	233.6	June 28	233.6	Third monthly
July 15, 1996	233.6	August 26	233.6	Fourth monthly
August 15, 1996	233.6	September 18	233.6	Second quarterly
September 15, 1996	233.6	October 11	233.6	Sixth monthly
October 15, 1996	233.6			
November 15, 1996	233.6	December 18	233.6	Third quarterly
December 15, 1996	233.6			
January 15, 1997	233.6	February 12	467.3	Eighth monthly
February 15, 1997	233.6			
May 15, 1997, and later dates[1]	4,097.6			
		B. Revised schedule approved on May 16, 1997		
May 16, 1997	500.0	May 21	500.0	Second annual
August 1, 1997	500.0	September 8	500.0	Fifth quarterly
November 1, 1997	500.0	January 13	500.0	Sixth quarterly
February 1, 1998	500.0	June 30	500.0	Seventh quarterly
May 16, 1998, and later dates[1]	2,564.7			
		C. Revised schedule approved on July 20, 1998[2]		
July 20, 1998	768.4	July 22	768.4	Emergency review
September 15, 1998 and later dates[1]	1,796.3			

Source: IMF.

[1]The detailed schedules for these disbursements are not shown because they were replaced by revised schedules in the first two cases and no further disbursements were made in the third case.

[2]The amounts in this panel refer only to those committed under the original EFF. The additional amounts approved on July 20 under the EFF augmentation, the Supplemental Reserve Facility, and the Compensatory and Contingency Financing Facility are shown in Table 1. The actual disbursements under the three facilities were all made on July 22.

some reduction was achieved. But the reduction was much less than the IMF recommended or the government announced that it would bring about. Strong vested interests prevented weak governments from delivering on their commitments. Taxes were not paid and unbudgeted expenditures were extracted from the government. The result of the failure to bring the fiscal deficit down was the August 1998 financial crisis, which could have been avoided if the fiscal situation had been under better control.[45]

Despite what the large deficits of the 1990s and the 1998 crisis suggest, the IMF was not without some influence. First, the government wanted to continue to qualify for IMF disbursements and therefore tried hard, albeit with rather modest results (except in 1995), to bring the deficit below what it would have been in the absence of the IMF. The IMF helped tilt the internal argument toward reformers who were pushing for more fiscal consolidation. Also, despite the disappointing overall outcome, many of the specific measures taken to enable Russia to continue to qualify for disbursements were good reform measures with long-term benefits. Second, the extensive discussions between the government and the IMF about fiscal (and monetary) policy issues led over time to the existence of a strong cadre within the government that was able to articulate the case for a restrained fiscal policy in public and in private, and to design an appropriate policy. Third, IMF advice and technical assistance contributed to many improvements in budget procedures, tax policy, and tax administration that made it easier to design and execute the planned budget. As described in more detail in the Appendix, technical assistance in the fiscal area was concentrated on tax administration and the creation of the treasury, although assistance also was provided on tax policy, public expenditure management, the social sector, and other issues. The impact of such assistance built up gradually over time, as administrative systems were reformed and increasing numbers of Russian officials learned the new procedures. In the case of tax policy reform, major changes did not come until 2001 under President Vladimir Putin. The seeds of some components of this reform had been sown earlier, partly by the IMF and other technical assistance advisors. The IMF staff had opposed other components, however, notably the move to a flat-rate personal income tax, which, although desirable on structural grounds, had seemed to the staff to carry the risk of a temporary loss of revenue.

Would the IMF's influence on fiscal performance have been greater if it had postponed disbursements more often in response to missed targets? Consider first the accumulation of arrears as a by-product of the successful reduction of the deficit in 1995. The IMF was aware at the time of the adverse consequences, although it may not have envisaged the full scale of the nonpayment problems that were partly caused by government arrears.[46] The IMF was so pleased that progress had been made toward macroeconomic stability that it did not aggressively challenge the manner in which the progress had been achieved. Had it done so—for example, by postponing some disbursements in 1995 until revenues were raised or expenditures

[45]Hernández-Catá warned as early as the beginning of 1995 that the continuation of a strategy of large budget deficits was dangerous (Hernández-Catá, 1995, p. 121).

[46]See Cheasty and Davis (1996).

Table 4. Reviews Under the Extended Fund Facility (EFF)

Date of Board Meeting	Review	Delay	Easing of Targets	Fiscal Measures
March 26, 1996	Approval of EFF			
April 29, 1996	First monthly	On time	Fiscal deficits increased from April to enable payment of wage arrears using new foreign loans; deficits reduced toward end-1996 so that annual deficit target unchanged.	
June 5, 1996	First quarterly	On time	Fiscal deficits increased from May because of higher interest payments than expected.	
June 24, 1996	Third monthly	On time	May target for international reserves missed.	
August 21, 1996	Fourth monthly	Delayed by one month	June targets for international reserves, net domestic assets, net credit to government, and federal government deficit missed. Fiscal deficits from July onward raised because of weak revenue collections and higher interest payments. Targets for international reserves, net domestic assets, and net credit to government eased to reflect higher growth in nominal GDP and money demand.	Measures taken to strengthen revenues.
September 13, 1996	Second quarterly	On time		
October 8, 1996	Sixth monthly	On time		
December 13, 1996	Third quarterly	On time	September and October targets for international reserves, tax revenues, and fiscal deficits missed. November and subsequent targets for tax revenues, fiscal deficits, and reserves eased. Targets for net domestic assets and net credit to government tightened.	Measures taken to strengthen tax collection. Major program of revenue measures developed.

February 7, 1997	Eighth monthly	Delayed by two months, but two monthly disbursements made together	No further easing; delay caused by the delays in October–December 1996.	
May 16, 1997	Second annual and ninth monthly	On time, but two monthly disbursements missed	Most January targets missed, but March targets hit. New targets set for second year.	Measures taken to strengthen tax revenues.
September 3, 1997	Fifth quarterly	On time	June target for tax revenues missed by a small amount. Credit targets tightened and reserves targets raised for next two quarters.	
January 8, 1998	Sixth quarterly	Delayed by 1–2 months	September targets for international reserves and tax revenues missed. Targets for international reserves, tax revenues, and credit growth eased for the fourth quarter of 1997 and subsequent quarters.	Measures taken to strengthen tax collection.
June 25, 1998	Third annual and seventh quarterly	Seventh quarterly delayed by 4 months	March targets for tax revenues and fiscal deficits missed. Delays in implementing revenue reform measures.	Major program of tax and expenditure measures initiated.
July 20, 1998	Emergency review, with enhanced commitments and disbursements	Not applicable		Major program of measures, mostly in the fiscal area, launched.

Source: IMF.

cut in a more economically sensible and socially acceptable way—future problems might have been diminished, although at the price of less progress in reducing the deficit and inflation.

Turning next to the missed targets for the deficit and for revenues—especially from mid-1996 through mid-1998, when major taxpayers were able to make arrangements with the government to pay less than their full liabilities—some disbursements were postponed, but the postponements were fairly short (see Table 3). At the time, the IMF generally believed that the chances of implementing the fiscal policy set forth in the program were greater if postponements were not too frequent or prolonged. If delays were prolonged, the relationship between the IMF and the authorities would become strained; the authorities would be less, rather than more, inclined to follow the IMF's advice; and the position of the reformers in government who shared the IMF's views would be weakened. These were years when the government appeared to be especially weak. Yeltsin was frequently sick and not exercising strong leadership. The oligarchs were confident of their power vis-à-vis the state and were manipulating the law, or worse, to add to their wealth.[47] Under these circumstances, the G-7 favored IMF engagement with Russia as a way of adding some weight to those in government who were struggling to bring the fiscal situation under control. As noted above, the IMF was not impervious to the views of the G-7; however, the IMF management itself believed that close engagement with the Russians, with continuous dialogue about economic programs and the next IMF disbursement, could effect some improvements in policies and avoid major reversals, which could have been very disruptive. Given the importance of Russia in the region and the world, it was thought to be worth the risk to the IMF's reputation and finances to persist with the engagement.

The alternative view was that the government would be strengthened in its relations with recalcitrant taxpayers and others if it were deprived of IMF backing, especially financial support.[48] This view was held by some reformers in Russia, notably Boris Fedorov and Andrei Illarionov,[49] and it was increasingly held by some in the IMF, both executive directors and staff.[50] Changes in Russia, such as the appointment of Anatoly Chubais and Boris Nemtsov as first deputy prime min-

[47]A very senior representative of one of the biggest oil and banking groups told the author in November 1997 that the period of asset grabbing would last only a few more years and that after that his company would have to start earning money in the usual way of western companies.

[48]An incident in December 1997 is an example of IMF flexibility weakening the efforts of some people in government to persuade taxpayers that they were serious about strengthening tax collection. Chubais was trying to use the Emergency Tax Commission to seize the assets of two delinquent taxpayers. Chernomyrdin wanted to take a more lenient line with the two companies and urged Camdessus to agree with him, while promising to improve tax collections. Camdessus took him at his word and decided that the IMF should not insist on specific actions by the Emergency Tax Commission as a condition of disbursing the next installment of the EFF. This incident had wider ramifications for other taxpayers' assessment of the government's willingness to tackle evasion.

[49]Federov had been deputy prime minister and finance minister in 1993, and Illarionov, who was an economic advisor to Prime Minister Chernomyrdin in the early 1990s, was President Putin's advisor in 2000–05.

[50]Notably, Michael Mussa, economic counselor, was an early advocate of terminating IMF financial support.

isters in March 1997, gave rise to new hope that the fiscal problems would be solved and held the growing disillusionment in the IMF at bay. The very large disbursement of US$4.8 billion in July 1998 was seen by the IMF leadership as the last chance for Russia. When this disbursement did not succeed in restoring confidence, there could be no question of more money without a major strengthening of the country's fiscal position.

The decision to withhold financial support in August 1998 had important ramifications in Russia and international financial markets. It contributed to the determination of subsequent Russian governments to keep the fiscal situation under control so that they would never again be dependent on the IMF. (However, more important influences on government behavior were the absence of new borrowing possibilities and their wish to avoid both another crisis and renewed dependence on financial markets.) International financial markets learned that the G-7 and the IMF would not always provide assistance to important countries just because they were "too big to fail."

We have discussed whether a firmer policy by the IMF, with less willingness to continue disbursing funds in the face of continued fiscal slippages, would have been better. A tougher line, with more interruptions to disbursements and less IMF money to finance deficits, might have had some impact on the behavior of recalcitrant taxpayers and the Duma majority that was opposed to the government's policy, although these groups were very powerful and were focused on improving their own financial and political positions at the expense of the weak state. A tougher line might also have led to a crisis, perhaps a less damaging one, well before August 1998. Taking into account what we now know, including the cathartic effect of the crisis, it seems that it would have been better for economic reforms in Russia, and for the IMF, if the IMF had taken a tougher line in 1996–98. It is, of course, much easier to say this now than at the time, when the possibility was very real that without active lending from the IMF, Russia would face economic collapse, or at least economic policy going seriously wrong.

Social expenditures

One of the major casualties of the failure to bring fiscal policy under control before 1999 was the social sector. The health and education sectors were starved in the budget. Moreover, even the meager budgeted amounts for the social sector were often not made available, as resources were sequestered during the year and bills—including for wages and pensions—were not paid. The overall standard of living and general well-being of those who could not afford to pay out of their own pockets for health and education services were severely reduced. The social safety net still consisted primarily of heavily subsidized prices, which were being raised over time, for housing and related services and for privileged groups, rather than income transfers based on need. Should the IMF have tried harder to protect social expenditures?

Some reduction in the real value of social sector expenditures was inevitable because of the decline in GDP. A further reduction would have been justified if reforms had introduced more private financing of and charges for health, education,

and other social services in a structured way. But, although there was a willingness in principle to introduce private financing and charging, the necessary structural reforms for this and for the overall reduction in the size of the sector were never undertaken in a comprehensive way. As a result, the adverse effect of the deterioration of the social sector on the standard of living was greater than it had to be.

The IMF's view from the beginning of the transition was that a strong social safety net was critically important, and Camdessus frequently referred to that in his speeches.[51] The IMF provided technical assistance regarding ways of targeting the social safety net to the most vulnerable groups instead of continuing with expensive subsidies that held down the cost of housing, utilities, and services for everyone.[52] The IMF also encouraged the government to include various social protection measures in its programs.[53] Reforms to social safety nets featured prominently in the government's program in 1992 and again in the 1996 EFF-supported program.

In practice, however, little effort was made to reform the social sector or to protect social expenditures from general fiscal tightness. This reflected the government's priorities against the background of its weakness in the face of strong taxpayers who were able to avoid paying their full obligations. Despite its concerns about this situation, the IMF did not push hard for higher social expenditures or the reform of the social safety net. It never made shortfalls in these areas a reason for delaying disbursements or a major sticking point in program negotiations. The Russians knew this. In the exception that proved the rule, they were surprised by the intensity of Camdessus's concern in 1999 to limit the large cuts in real wages and pensions that were being planned by the government as part of the adjustment to the 1998 crisis.[54] But, in general, the IMF felt that it could not, against a background of weak revenues, insist on achieving both a satisfactory overall fiscal balance and the protection of social expenditures.[55]

Specific fiscal issues

In two specific fiscal issues, the IMF made a difference and, in hindsight, a mistake. The first was the abolition of the oil export duty, which the IMF insisted, against the advice of the authorities, should be a condition of the approval of the EFF in March 1996. This was seen as a major step in a program that had been run-

[51]In summarizing the desirable reform strategy, Camdessus (1992) gave equal billing to social security and to structural reforms and macroeconomic stabilization. He said that "the need for adequate social safety nets is greatest when society is undergoing far-reaching upheavals, as it is now." (Camdessus, 1992. See also Camdessus 1994 and 1997.)

[52]The main lines of IMF advice are summarized in Chu and Gupta (1996) and IMF (1995).

[53]The IMF was not qualified to advise on the health and education sectors. The World Bank was active in these areas.

[54]Primakov complained to Camdessus about being outflanked on the left.

[55]Some people—for example, Lopez-Claros (2002)—have argued that the IMF's failure to protect the social safety net while allowing the oligarchs to minimize tax payments was a major mistake, because it undermined support for reforms.

ning for a few years to remove all taxes and restrictions on exports. It was to be accompanied by an increase in excise duties on oil and oil products. As it turned out, the enforcement of the higher excise duties proved difficult, as they were much easier to evade than the export duty, and export duties were reintroduced after the 1998 crisis. Given the importance of collecting as much revenue as possible in the 1996–98 period, it would have been better to have postponed the abolition of the export duty on oil until the excise tax collection system and the government's authority were stronger.

Second, the IMF supported the government's wish to liberalize access to the treasury bill market by foreign investors in 1996. The government's intention was to reduce interest rates by opening up the market. The IMF was sympathetic to this view and, more generally, to the liberalization of capital markets. There was also strong support from the G-7, especially the United States, for the same reasons. However, the CBR was opposed to this measure because it feared the possible volatility of external capital flows. The IMF pressed for liberalization, and in early 1997 the CBR agreed to a phased program of liberalization to be completed in January 1998. This added a significant element to the financial market volatility that culminated in the August 1998 crisis, although, because of the behavior of domestic investors, the crisis would not have been avoided had the liberalization not occurred.[56]

III. Monetary Policy

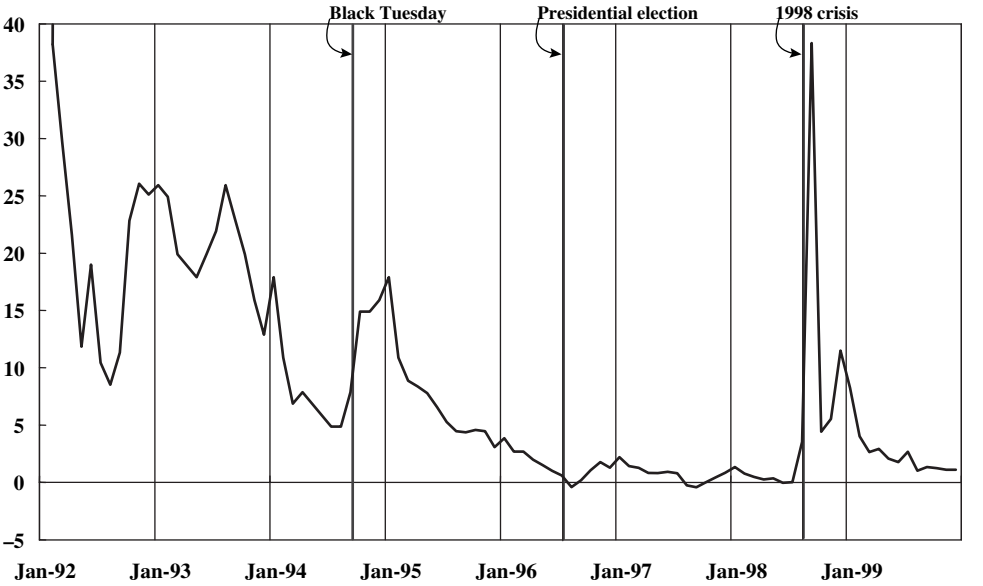
Monetary policy in the 1990s falls into two periods: 1992–94, when the reformers were struggling against opposition to bring about tighter monetary policy, and 1995–99, when the reformers had largely won the battle and monetary policy was mostly under control. Inflation was high in the first period and low (in single digits in 1996–97) in the second period, apart from the blip caused by the 1998 crisis (see Figures 4 and 5).

The efforts of the IMF were devoted in the first period to persuading the authorities that inflation was a serious problem, and that it could and should be tackled by tightening monetary policy. Technical assistance was provided to the CBR to strengthen its ability to plan and execute monetary policy. There was only limited understanding, except among reformers, that inflation was caused by loose monetary policy and excessive credit expansion. Various alternative views could be found, including that inflation was caused by price liberalization, the high degree of monopolization of the economy, or other market imperfections. Some people were concerned that tightening monetary policy would reduce output growth rather than inflation. The IMF was actively engaged in countering these arguments, drawing on experience in other countries.[57] It was helped by events, especially the

[56]While the IMF appeared to play an important role in bringing about liberalization, some government officials and observers have argued that if the IMF had not pressed, the government would itself have insisted on liberalization.

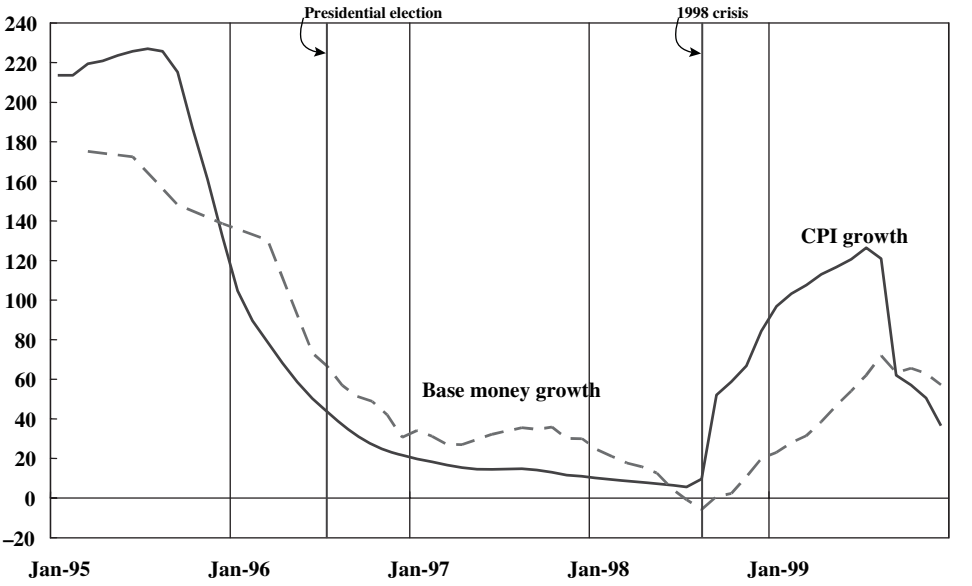
[57]This paragraph is based on Hernández-Catá (1995), who provides additional details.

Figure 4. Inflation
(*In percent, month-on-month change in CPI*)



Source: IMF staff estimates based on data from Russian authorities.
Note: CPI=Consumer Price Index

Figure 5. Inflation and Money
(*In percent, 12-month growth rate*)



Source: IMF staff estimates based on data from Russian authorities.

continuation of high inflation well past the initial price liberalization, the failure of output to respond positively to the rapid credit expansion in the summers of 1992 and 1993 (and 1994), and the rise in inflation after the monetary expansion in the third quarter of 1994 and Black Tuesday in October. In turn, the IMF's arguments helped the reformers in their internal discussions.

Although the effect built up slowly, by 1995 the IMF had had a significant influence on the authorities' views of the role of monetary policy and inflation. The IMF also had an influence on the measures used to tighten monetary policy. For example, the program agreed upon with the IMF in mid-1993 included, at the IMF's suggestion, a rule for tying the CBR's bank lending rate to a market rate. When implemented along with a reduction in monetary credit growth, this rule produced a sharp rise in the lending rate to a positive real level. The 1995 program targeted a major reduction in monthly inflation to 1 percent in July, to be achieved by tightening monetary policy, and the government and the CBR endorsed it. While they might not have done so had they not been shocked by Black Tuesday (which led, among other things, to reformers taking leading positions in the government and the CBR), the IMF also had an impact. The implementation of the program was made difficult by unexpected capital inflows that, in the absence of adequate sterilization instruments, caused excessive monetary growth in the middle of the year, and inflation remained above the target path. The IMF worked closely with the authorities during the year (there were monthly missions) on corrective measures to deal with the situation.

From 1995 onward, the IMF's influence was less on major themes, such as tightening monetary policy to produce a big reduction in inflation, and more on the details of setting and implementing monetary policy. The ongoing dialogue between IMF staff and the authorities—about issues such as how to respond to inflows (1995 and 1997) and outflows (1996 and 1998) of international reserves, how and how much to sterilize changes in reserves, and how to keep credit growth in the economy within the targeted range—was the channel for a major transfer of knowledge and experience from the IMF to the authorities.

IV. Exchange Rate Policy

This section discusses the impact of the IMF on exchange rate policy at three critical stages: the liberalization of the exchange system in 1992, the introduction of the corridor in 1995, and the decision to maintain the pegged rate regime right up to August 1998.

The liberalization of the exchange system was one of the major reforms of 1992, after the initial reforms in January. In July 1992, a unified exchange system replaced most of the system of multiple exchange rates that had existed during the first half of the year. Some minor exceptions to the unified system were brought in later in the year. The authorities attached great importance to making the ruble free and convertible, and strong pressure from the IMF was not needed. The IMF did, however, play a significant role in helping the authorities identify and phase out the various elements of the multiple exchange system and resist proposals that would have segmented the market again. The IMF's contribution included a major

educational element—some Russian officials, while sympathetic to liberalization and unification, did not initially understand all the ramifications of the policy.

The idea of introducing the exchange rate band in July 1995 came from the government; the CBR was significantly less enthusiastic. The government was concerned about the strong pressures for the exchange rate to appreciate. If appreciation was permitted (and some did take place), this would reduce competitiveness quickly, as Russia's inflation was already considerably above that of western trading partners. This could damage export prospects or produce a sharp reversal in the appreciation (the political and economic consequences of Black Tuesday were still fresh in people's minds). If, on the other hand, the appreciation was averted by intervention, in the absence of adequate sterilization instruments noted above, the consequent rapid monetary growth would threaten the achievement of the inflation target. This was already of concern to the government and the IMF. The government hoped that the introduction of the band would persuade financial markets that there was a ceiling to the exchange rate and that the markets themselves would keep the rate below the ceiling with little need for intervention. The government also hoped that the band would provide some protection against a sharp depreciation should sentiment move in the other direction.

Since the 1992 liberalization, most IMF staff working on Russia had opposed pegging the exchange rate, and they continued to have doubts about whether fiscal policy would be strong enough to support the peg. Some staff, noting the success of exchange rate–based stabilizations in Poland, Estonia, and some other transition countries, favored pegging the rate. Stanley Fischer had for some years advocated exchange rate–based stabilizations in countries with high inflation and had encouraged the Russian government to expect a favorable response from the IMF to any proposal it might make to switch to a pegged regime. The IMF was in favor of pegs in many other emerging market countries and had a general policy of accepting a country's choice of exchange rate regime, provided that the supporting policies were in place. Therefore, despite some doubts about the fiscal situation and the ultimate exit strategy, the IMF officially supported the Russian authorities' proposal, and the next mission worked closely with them to design a monetary policy framework that was compatible with the new exchange rate policy.

In the following three years, the IMF continued to support the exchange rate band policy. Before each announcement of the band for the next period, there was an exchange of views between the IMF and the CBR. On all occasions, agreement was reached quickly with little disagreement. The outcome was usually very close to the initial plans of the CBR.

With fiscal policy continuing to be very weak in 1996–98, the burden of supporting the exchange rate peg fell on monetary policy. The resulting policy mix produced a more appreciated real exchange rate than would have been the case if fiscal policy had been tighter. However, an earlier devaluation without a tightening of fiscal policy would not necessarily have produced a more depreciated real exchange rate for longer than a temporary period. It would, however, have produced higher inflation, which would have tended to postpone the recovery in output, thus working in the opposite direction from any real exchange rate effect. The IMF continued to urge the government to go for the first best solution—namely,

tightening fiscal policy—and did not advocate the alternative approach of a devaluation in, say, 1997.[58]

Given the fiscal policy actually pursued in 1996–98, it is appropriate to ask whether the IMF should have opposed the introduction of the peg in 1995, or at least insisted on an earlier exit. The absence of the peg might have led to a better outcome for growth in the short term, greater flexibility to deal with the shocks from the Asian crisis and the oil price drop, and no major crisis associated with the exit from the peg. But it would have been at the price of higher inflation, and growth over the medium and long term might have been no better. It is, anyway, difficult to see the grounds on which the IMF could have opposed the peg, given its openness to such arrangements and its commitment to helping the government achieve the reduction in the fiscal deficit that it appeared to want. In supporting the government's fiscal plans—for example, the proposed deficits of 4 percent, 3 percent, and 2 percent in 1996, 1997, and 1998, respectively, in the EFF-supported program—the IMF was implicitly saying that the exchange rate peg was sustainable.

Should the IMF have pressed for an earlier exit as it became clear that the fiscal problems were not being solved? Here one faces the usual dilemma of exit strategies. As long as exchange market conditions are benign, as they were in Russia until the onset of the Asian crisis in late 1997, there is no reason to change the strategy. When market conditions worsen, any attempt to exit the peg arrangement risks creating a crisis. Nevertheless, there was probably a window from November 1997 to March 1998 when people might have been persuaded of the need for an exit, and an orderly exit might have been possible. In hindsight, it is regrettable that the IMF did not spot this opportunity at the time.[59]

V. Conclusions

This paper is about the IMF's impact on economic policies as they were implemented in Russia in the 1990s. By far the most important determinants of economic policies were the Russian leadership and the many economic agents who were in a position to support or thwart that leadership. The leadership was itself divided, most visibly between the government, which usually had broadly reformist tendencies, and Parliament, which often opposed reforms, in many cases under the influence of economic agents. The government itself was often a coalition of opposing views, and the state was weak. In this situation, economic agents were able to manipulate the law and its implementation to their advantage through

[58]It should be noted that many commentators—for example, Berglöf and Vaitilingam (1999); Kharas, Pinto, and Ulatov (2001); and Stiglitz (2002)—have argued that the exchange rate was overvalued before the crisis and that an earlier devaluation would have been beneficial. For a defense of the IMF's position, see the comments of Fischer (2001) and Summers (2001) on the paper by Kharas and others (2001).

[59]The government was opposed to a devaluation because its credibility and one of its few macroeconomic achievements—the low inflation rate—were linked to the exchange rate peg. It would, therefore, have resisted any pressure from the IMF for an early devaluation, although it is possible that it would have eventually relented.

their influence, with or without bribes, over government officials and members of Parliament.

Given the turbulence of the struggle for economic and political power in the 1990s, it is not surprising that the IMF was not a dominant force in determining economic policies. Two other factors weakened the IMF's influence. First, in the early 1990s, some suspicion of foreign advisors remained, along with a reluctance (natural in a former superpower) to admit that there were things to be learned. Although these attitudes dissipated gradually during the decade, the perception persists in some quarters that the IMF's advice is not in Russia's best interest. Second, the G-7's wish to see close IMF engagement in Russia reduced the IMF's ability to insist on ambitious policies and full implementation of agreed-upon programs. Thus, a weakened IMF urged the implementation of strong macroeconomic policies that were supported by the reform wing in the government but were often opposed by stronger groups in the government, Parliament, and the economy.

The biggest consequence of the IMF's limited influence was the failure to raise revenues and reduce fiscal deficits, which led to the 1998 crisis. Had the fiscal consolidation that was agreed upon between the government and the IMF in every year up to 1996 (when the medium-term program to 1998 was set) been achieved, it is unlikely that the crisis would have occurred. It is even possible that a tougher stance by the IMF might have caused the authorities to bring about enough of a consolidation to avert the crisis. The crisis did, however, teach the important lesson that fiscal prudence is crucial, just as Black Tuesday (October 11, 1994) had taught the lesson that monetary prudence mattered.

Another consequence of the IMF's relative weakness was that it had to focus its efforts on key monetary and fiscal policy issues rather than attempting to affect the whole range of reform policies. Thus, in practice, it had very little influence over structural and institutional reforms outside areas closely linked to the design and implementation of monetary and fiscal policies. In addition, it did not have the mandate or expertise to have an impact on the detailed implementation of the wider agenda of structural and institutional reforms. Its main contribution was to help keep such reforms on the agenda and to emphasize their importance. Its most ambitious effort in this direction was its work, together with the World Bank, on the government's structural reform agenda for 1996–98 in preparation for the EFF. The protection of social expenditures also received a relatively low weight in decisions about how to exert the IMF's influence. A special case was the increase in arrears in 1995, which met with undue tolerance from the IMF because the overall fiscal situation had been brought under control for the first time.

With regard to areas in which the IMF exerted more influence, the most important was monetary policy. The IMF repeatedly presented the case to the authorities for directing monetary policy toward producing low inflation and, just as important, provided technical assistance on how the CBR should do this operationally. Although the political conditions in 1992–94 were not conducive to the sustained implementation of a serious disinflation strategy, the IMF laid the foundation for later policies through its operational advice and debates with the CBR and the government. When conditions changed and the CBR and government economic policy came under new leadership in late 1994, this foundation supported disinflationary

monetary policies. The IMF continued to provide close support during 1995 and subsequent years, and the general success of the policies is seen in the decline in inflation in the years before the 1998 crisis.

Even in fiscal policy the IMF had some impact. Despite the failure at the macroeconomic level, much of the specific advice of the IMF was taken up and implemented by the government. For example, the long struggle to get rid of offsets would probably not have been successful so soon if the IMF had not been pushing. And some components of the tax reforms of 2001 were derived in part from technical advice from the IMF in the mid-1990s. However, the IMF's influence was not always positive. It was, for example, a mistake to push for early abolition of the oil export duty and liberalization of the treasury bill market.

Given what we now know about fiscal policy, it might have been better not to have introduced the pegged exchange rate regime in 1995. But the IMF could not have easily prevented it in the face of the government's insistence, as it was in principle open to such arrangements, and it did not expect fiscal policy to be so weak. Similarly, a devaluation before March 1998 might have been better than what actually happened. But this was not obvious at the time, and the government might have resisted any IMF attempts to persuade it to change policy in this direction. Less ambiguous was the positive assistance that the IMF gave to the unification and liberalization of the exchange rate regime in 1992.

This mixed picture of the IMF's impact on specific policies and reforms is only part of the story of the IMF's role in Russia in the 1990s. Another part was the transfer of knowledge to individual officials and experts in the CBR and the government, which took place almost continuously through ongoing discussions about policy issues between the authorities and IMF missions, Moscow office representatives, and resident experts. It occurred at every level, from Camdessus' interactions with Chernomyrdin to discussions between economists about data spreadsheets. This ongoing dialogue produced, in addition to specific policy improvements, a large number of officials and experts with a fine understanding of macroeconomic policy issues. As time has passed, more of these people have reached senior positions, and the knowledge has spread widely through key agencies and public bodies, ensuring that the CBR and the government as a whole now design and manage macroeconomic policies in ways that are recognizably similar to those used in western industrial countries. There has also been a related improvement in the quality of the public debate about macroeconomic issues. While the IMF was only one of the elements in this critical transformation of the government's role from central planning to managing a market economy, the transfer of knowledge it brought about may stand as its major legacy from the turbulent 1990s.

APPENDIX

Channels for IMF Advice

The main channels for IMF advice between 1992 and 2003 were the discussions between visiting staff missions from the European II Department (EU2) of the IMF and their Russian counterparts. In 2003, the IMF's area departments were reorganized; EU2 was disbanded, and responsibility for Russia was transferred to the European Department (EUR). In 1992–94 and

1997–2001, there were usually five or six missions a year, each lasting about two weeks. In 1995–96, there were monthly missions, although they were shorter (7–10 days). Since 2001, the frequency of missions has declined to two or three a year.

During 1992–2001, when Russia was borrowing or seeking to borrow from the IMF, much of the policy discussion between the missions and the authorities took the form of a debate about the details of the economic program for the coming period. The annual programs were most important, but quarterly and monthly programs were also discussed. Typically, drafts of the program documents were prepared by the IMF team, taking into account both the IMF's views about what was needed and the authorities' objectives. The drafts were amended to take into account the authorities' views, and differences were negotiated. Since 2001, the Russian authorities have not sought to borrow from the IMF and the policy dialogue has been less intense. The IMF staff from EU2/EUR has focused on giving advice on a few key issues and backing it up with convincing technical work.

A related channel to the EU2 missions was the work of the staff at the IMF Moscow office, whose role has varied according to the frequency of EU2 missions. When these were very frequent, Moscow office staff spent most of their time supporting the missions. Since the missions have become less frequent in recent years, the Moscow office staff have played an important role in advising the Russian authorities directly. One advantage they have in this role over their colleagues at IMF headquarters is that they have time to develop a deeper knowledge of specific issues and closer personal relationships with their counterparts. The latter quality led in a few cases to Moscow office staff members acting as part-time advisors to senior officials in the CBR or the finance ministry.

A third channel for advice was technical assistance, which took the form of missions of technical experts and resident technical advisors who advised on particular issues, especially the development of the capacity of the CBR and the ministry of finance to design and implement monetary and fiscal policies. In the 10-year period from May 1991 to April 2001, the IMF provided 63 person-years of technical assistance (TA), which was nearly 4 percent of its worldwide TA provision (Table A.1). In the four peak years, May 1992 to April 1996, nearly 6 percent of the IMF's worldwide TA effort was devoted to Russia. In the early 1990s, well over half of the technical assistance to Russia concerned the development of the CBR as a

Table A.1. IMF Technical Assistance Work in Russia

Fiscal Year*	Number of Person-Years	In Percent of Worldwide Total
1992	3.2	2.4
1993	7.2	4.8
1994	9.6	6.7
1995	12.1	7.2
1996	8.8	4.2
1997	6.1	3.3
1998	4.0	2.7
1999	4.6	2.3
2000	3.9	2.0
2001	3.6	1.9
Total	63.1	3.7

Source: IMF.
*The fiscal year runs from May of the previous year to April of the year shown in this column.

modern central bank. The rest was mostly assistance in the fiscal area, including advice on tax policy, tax administration, and the creation of the treasury. As Table A.2 shows, this assistance was concentrated on tax administration and the treasury. All the long-term resident advisors (who accounted for over half of the total fiscal TA provision) worked in these two

Table A.2. IMF Technical Assistance in the Fiscal Area

A. Short visits (number of visits)[1]

Fiscal year[2]	Tax policy	Tax administration	Treasury	Social sector	Public expenditure management	Other[3]	Total
1992	2	2	1	1		2	8
1993	1	3	6			2	12
1994		2	14	1			17
1995	1	6	6			1	14
1996	2	1	5				8
1997	2	6	1			2	11
1998		3		1		2	6
1999		2	1			1	4
2000		2	2		2		6
2001	1	3	2		6	2	14

B. Long-term resident advisors

		Total (person-years)
February 1993–June 1997	Tax administration	4.4
January 1994–December 1995	Treasury software	2.0
March 1994–September 1995	Tax administration	1.6
July 1994–December 1995	Treasury	1.5
September 1997–August 1999	Tax administration	2.0
February 1999–August 2001	Treasury	2.6

C. Short- and long-term technical assistance (person-years)

Fiscal year[2]	Short visits[1]	Long-term advisors	Total
1992	1.7	0.0	1.7
1993	2.5	0.3	2.8
1994	2.3	1.5	3.8
1995	1.3	3.8	5.1
1996	0.6	2.8	3.4
1997	1.0	1.0	2.0
1998	0.4	0.9	1.3
1999	0.4	1.2	1.6
2000	0.7	1.3	2.0
2001	1.0	1.0	2.0
Total	11.9	13.8	25.7

Source: IMF.

[1]Includes visits lasting from a few days to a few months, comprising both single experts and teams.

[2]The fiscal year runs from May of the previous year to April of the year shown in this column.

[3]Includes missions on general budget making and fiscal monitoring, participation in World Bank missions, and other visits.

areas. There was also assistance on statistics and drafting legislation. Details of the TA in the central banking area in Russia and other transition countries in the region are available in Zulu and others (1994), Knight (1997) and Sundararajan and others (1997). Details of the work on setting up treasuries in Russia and the other countries of the region can be found in Potter and Diamond (2000), and details on reforming budget systems in Russia in Diamond (2002a, 2002b).

The fourth channel for IMF advice was less direct. The IMF organized many courses for government and CBR officials at its training institutes in Washington and Vienna. It also invited officials, parliamentarians, and journalists to seminars in Washington, Vienna, and centers in Russia. These courses and seminars covered economic policies in general, economic policies in transition, and specific Russian policy issues. Typically, they did not address current policy issues, unlike the policy advice provided by EU2 missions and Moscow office staff.

REFERENCES

Allison, Graham, and Grigory Yavlinsky, 1991, *Window of Opportunity* (New York: Pantheon Books).

Anderson, Jonathan, Daniel A. Citrin, and Ashok K. Lahiri, 1995, "The Decline in Output," in *Policy Experiences and Issues in the Baltics, Russia, and Other Countries of the Former Soviet Union,* ed. by Daniel Citrin and Ashok Lahiri, IMF Occasional Paper No. 133 (Washington: International Monetary Fund).

Åslund, Anders, 2002, *Building Capitalism: The Transformation of the Former Soviet Bloc* (Cambridge, United Kingdom: Cambridge University Press).

Berglöf, Erik, and Ramesh Vaitilingam, 1999, *Stuck in Transit: Rethinking Russian Economic Reform* (London: CEPR; Moscow: RECEP; and Stockholm: SITE).

Camdessus, Michel, 1992, "Economic Transformation in the Fifteen Republics of the Former USSR: A Challenge or an Opportunity for the World?" address to Georgetown University School of Foreign Service, Washington, DC, April 15.

———, 1994, "The Transformation of the Russian Economy: Progress Made, Challenges Remaining, and the Role of the IMF," address at the Moscow Finance Academy, Moscow, March 21.

———, 1997, address at the Moscow Institute of International Affairs, Moscow, April 2.

Cheasty, Adrienne, and Jeffrey M. Davis, 1996, "Fiscal Transition in Countries of the Former Soviet Union: An Interim Assessment," *MOCT-MOST,* Vol. 6, No. 3, pp. 7–34.

Chu, Ke-young, and Sanjeev Gupta, 1996, "Social Protection in Transition Economies: Emerging Issues," *MOCT-MOST,* Vol. 6, No. 3, pp. 107–23.

Citrin, Daniel A., 1995, "Overview," in *Policy Experiences and Issues in the Baltics, Russia, and Other Countries of the Former Soviet Union,* ed. by Daniel Citrin and Ashok Lahiri, IMF Occasional Paper No. 133 (Washington: International Monetary Fund).

Citrin, Daniel, and Ashok K. Lahiri, eds., 1995, *Policy Experiences and Issues in the Baltics, Russia, and Other Countries of the Former Soviet Union,* IMF Occasional Paper No. 133 (Washington: International Monetary Fund).

Diamond, Jack, 2002a, "Budget System Reform in Transitional Economies: The Experience of Russia," IMF Working Paper No. 02/22 (Washington: International Monetary Fund).

———, 2002b, "The New Russian Budget System: A Critical Assessment and Future Reform Agenda," IMF Working Paper No. 02/21 (Washington: International Monetary Fund).

Dubinin, Sergei K., 1995, "Macroeconomic Stabilization in Russia: The Lessons of 1992–95 and the Outlook for 1996–97," in *Economic Transformation: The Tasks Still Ahead,* ed. by Jan Svejnar, Oleh Havrylyshyn, and Sergei Dubinin (Washington: Per Jacobsson Foundation), pp. 46–66. Available via the Internet: http://www.perjacobsson.org/lectures/1995.pdf.

Fischer, Stanley, 2001, "Comments and Discussion," in *Brookings Papers on Economic Activity: 1,* Brookings Institution, pp. 62–64.

Fischer, Stanley, and Ratna Sahay, 2000, "The Transition Economies after Ten Years," IMF Working Paper No. 00/30 (Washington: International Monetary Fund).

Freeland, Chrystia, 2000, *Sale of the Century: Russia's Wild Ride from Communism to Capitalism* (New York: Crown Business).

Gaidar, Yegor, 1997, "The IMF and Russia," *American Economic Association Papers and Proceedings,* Vol. 87, No. 2, pp. 13–16.

———, 1999, *Days of Defeat and Victory* (Seattle: University of Washington Press).

Graham, Jr., Thomas E., 2002, *Russia's Decline and Uncertain Recovery* (Washington: Carnegie Endowment for International Peace).

Hernández-Catá, Ernesto, 1995, "Russia and the IMF: The Political Economy of Macrostabilization," in *Policy Experiences and Issues in the Baltics, Russia, and Other Countries of the Former Soviet Union,* ed. by Daniel Citrin and Ashok Lahiri, IMF Occasional Paper No. 133 (Washington: International Monetary Fund).

International Monetary Fund, Fiscal Affairs Department, 1995, "Social Safety Nets for Economic Transition: Options and Recent Experiences," IMF Paper on Policy Analysis and Assessment 95/3 (Washington: International Monetary Fund).

International Monetary Fund, International Bank for Reconstruction and Development, Organization for Economic Cooperation and Development, and European Bank for Reconstruction and Development, 1990, *The Economy of the USSR: Summary and Recommendations,* a study undertaken in response to a request by the Houston Summit.

Kalra, Sanjay, and Torsten Sløk, 2001, "Inflation and Growth in Transition: Are the Asian Economies Different?" in *A Decade of Transition: Achievements and Challenges,* ed. by Oleh Havrylyshyn and Saleh M. Nsouli (Washington: International Monetary Fund).

Kharas, Homi, Brian Pinto, and Sergei Ulatov, 2001, "An Analysis of Russia's 1998 Meltdown: Fundamentals and Market Signals," *Brookings Papers on Economic Activity: 1,* Brookings Institution, pp. 1–50.

Klyamkin, Igor, and Lilia Shevtsova, 1999, *This Omnipotent and Impotent Government: The Evolution of the Political System in Post-Soviet Russia* (Washington: Carnegie Endowment for International Peace).

Knight, Malcolm, and others, 1997, *Central Bank Reforms in the Baltics, Russia, and the Other Countries of the Former Soviet Union,* IMF Occasional Paper No. 157 (Washington: International Monetary Fund).

Lopez-Claros, Augusto, 2002, "The Role of the International Financial Institutions During the Transition in Russia," Institute for the Economy in Transition, Moscow, December 18. Available via the Internet: http://iet.ru/guest/lopez/lopezbook.pdf.

Mau, Vladimir, 2000, "Russian Economic Reforms as Perceived by Western Critics," in *Russian Crisis and Its Effects,* ed. by Tuomas Komulainen and Iikka Korhunen (Helsinki: Kikimora Publications).

Nagy, Piroska Mohácsi, 2000, *The Meltdown of the Russian State: The Deformation and Collapse of the State in Russia* (Cheltenham, United Kingdom, and Northampton, Massachusetts: Edward Elgar Publishing).

Odling-Smee, John, 1996, "The IMF's Approach to Economies in Transition," in *Fifty Years After Bretton Woods: The New Challenge of East-West Partnership for Economic Progress,* ed. by Miklós Szabó-Pelsöczi (Aldershot, United Kingdom: Avebury).

———, and Gonzalo Pastor, 2002, "The IMF and the Ruble Area, 1991–93," *Comparative Economic Studies,* Vol. XLIV, No. 4 (Winter), pp. 3–29 and 81–84.

Odling-Smee, John, and Thomas Wolf, 1994, "Economic Reforms in Transition Economies: A Macroeconomic Perspective on the Baltic States, Russia, and Other Countries of the Former Soviet Union," paper presented at a seminar on "Economic Reform in Russia and Other Countries in Transition: Issues and Prospects," Moscow, October 18–19.

Owen, David, and David Robinson, eds., 2003, *Russia Rebounds* (Washington: International Monetary Fund).

Potter, Barry H., and Jack Diamond, 2000, *Setting Up Treasuries in the Baltics, Russia, and Other Countries of the Former Soviet Union: An Assessment of IMF Technical Assistance,* IMF Occasional Paper No. 198 (Washington: International Monetary Fund).

Primakov, Yevgeny, 2004, *Russian Crossroads: Toward the New Millennium* (New Haven: Yale University Press).

Sachs, Jeffrey D., and Wing Thye Woo, 1994, "Reform in China and Russia," *Economic Policy,* No. 18, pp. 101–45.

Stiglitz, Joseph, 2002, *Globalization and Its Discontents* (New York: Norton).

Summers, Lawrence, 2001, "Comments and Discussion," *Brookings Papers on Economic Activity: 1,* Brookings Institution, pp. 51–57.

Sundararajan, V., Arne B. Petersen, and Gabriel Sensenbrenner, eds., 1997, *Central Bank Reform in the Transition Economies* (Washington: International Monetary Fund).

Talbot, Strobe, 2003, *The Russia Hand: A Memoir of Presidential Diplomacy* (New York: Random House).

World Bank, 2003, *Assisting Russia's Transition: An Unprecedented Challenge* (Washington).

Zulu, J.B., ed., 1994, *Central Banking Technical Assistance to Countries in Transition* (Washington: International Monetary Fund).

In statistical matter throughout this issue,

dots (. . .) indicate that the data are not available;

a dash (—) indicates that the figure is zero or less than half the final digit shown, or that the item does not exist;

a single dot (.) indicates decimals;

a comma (,) separates thousands and millions;

"billion" means a thousand million; and "trillion" means a thousand billion;

a short dash (–) is used between years or months (for example, 1998–99 or January–June) to indicate a total of the years or months inclusive of the beginning and ending years or months;

a slash (/) is used between years (for example, 1998/99) to indicate a fiscal year or a crop year; and

components of tables may not add to totals shown because of rounding.

The term "country," as used in this publication, may not refer to a territorial entity that is a state as understood by international law and practice; the term may also cover some territorial entities that are not states but for which statistical data are maintained and provided internationally on a separate and independent basis.

Design: Luisa Menjivar-Macdonald and Sanaa Elaroussi