INTERNATIONAL MONETARY FUND

STAFF PAPERS

Openness, Human Development, and Fiscal Policies
DELANO VILLANUEVA • 1

Collection Lags and the Optimal Inflation Tax
ALEX MOURMOURAS and JOSÉ A. TIJERINA • 30

The Parallel Market Premium
PETER J. MONTIEL and JONATHAN D. OSTRY • 55

Exchange Rate Strategies and Fiscal Performance in Sub-Saharan Africa
KARIM NASHASHIBI and STEFANIA BAZZONI • 76

The Stabilizing Effect of the ERM on Exchange Rates and Interest Rates
MICHAEL J. ARTIS and MARK P. TAYLOR • 123

Shorter Paper and Comments: CATHARINA J. HOOYMAN; CHING-CHONG LAI and WEN-YA CHANG; PIERRE-RICHARD AGÉNOR and ROBERT P. FLOOD • 149

Vol. 41 No. 1 **MARCH 1994**

STAFF PAPERS

Among the responsibilities of the International Monetary Fund, as set forth in its Articles of Agreement, is the obligation to "act as a centre for the collection and exchange of information on monetary and financial problems." *Staff Papers* makes available to a wider audience papers prepared by the members of the Fund staff. The views presented in the papers are those of the authors and are not to be interpreted as necessarily indicating the position of the Executive Board or of the Fund.

To facilitate electronic storage and retrieval of bibliographic data, *Staff Papers* has adopted the subject classification scheme developed by the *Journal of Economic Literature.*

Subscription: US$50.00 a volume or the approximate equivalent in the currencies of most countries. Four numbers constitute a volume. Single copies may be purchased at $16.00. Individual academic rate to full-time professors and students of universities and colleges: $25.00 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
Telefax: (202) 623-7201

INTERNATIONAL MONETARY FUND

STAFF PAPERS

Vol. 41 No. 1 MARCH 1994

EDITOR'S NOTE

The Editor invites from contributors outside the Fund brief comments (not more than 1,000 words) on published articles in *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 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.

International Standard Serial Number: ISSN 0020-8027

The U.S. Library of Congress has cataloged this serial publication as follows:

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

v. tables, diagrs. 23 cm.

Three no. a year, 1950–1977; four no. a year, 1978–
Indexes:
Vols. 1–27, 1950–80, 1 v.
ISSN 0020-8027 = Staff papers — International Monetary Fund.

1. Foreign exchange—Periodicals. 2. Commerce—Periodicals.
3. Currency question—Periodicals.

HG3810.15 332.082 53-35483

C O N T E N T S

Vol. 41 No. 1 March 1994

Openness, Human Development, and Fiscal Policies:
Effects of Economic Growth and Speed of Adjustment
DELANO VILLANUEVA • 1

Collection Lags and the Optimal Inflation Tax:
A Reconsideration
ALEX MOURMOURAS and JOSÉ A. TIJERINA • 30

The Parallel Market Premium:
Is It a Reliable Indicator of Real Exchange Rate
Misalignment in Developing Countries?
PETER J. MONTIEL and JONATHAN D. OSTRY • 55

Exchange Rate Strategies and Fiscal Performance
in Sub-Saharan Africa
KARIM NASHASHIBI and STEFANIA BAZZONI • 76

The Stabilizing Effect of the ERM
on Exchange Rates and Interest Rates:
Some Nonparametric Tests
MICHAEL J. ARTIS and MARK P. TAYLOR • 123

Shorter Paper and Comments

The Use of Foreign Exchange Swaps by Central Banks
CATHARINA J. HOOYMAN • 149

Unification of Foreign Exchange Markets:
A Comment on Agénor and Flood
CHING-CHONG LAI and WEN-YA CHANG • 163

Unification of Foreign Exchange Markets:
A Rejoinder to Lai and Chang
PIERRE-RICHARD AGÉNOR and ROBERT P. FLOOD • 171

IMF Working Papers • 173

IMF Papers on Policy Analysis and Assessment • 174

This page intentionally left blank

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

Openness, Human Development, and Fiscal Policies

Effects on Economic Growth and Speed of Adjustment

DELANO VILLANUEVA*

The model developed here postulates that learning through experience raises labor productivity with three major consequences. First, the steady-state growth rate of output becomes endogenous and is influenced by government policies. Second, the speed of adjustment to steady-state growth increases, and enhanced learning further reduces adjustment time. Third, both steady-state growth and the optimal net rate of return to capital are higher than the sum of the exogenous rates of technical change and population growth. Simulation results confirm the model's faster speed of adjustment, and regression analysis finds that a large part of the divergent growth patterns across countries is related to the extent of economic openness, the depth of human development, and the quality of fiscal policies. [JEL F43, H3, H5]

THE BASIC NEOCLASSICAL growth model developed by Solow (1956) and Swan (1956) has been the workhorse of growth theory for the past three and a half decades. Its simple structure—consisting of a well-behaved neoclassical production function, investment-saving relation, and a labor growth function—is an elegant solution to the "knife-edge" problem posed by Harrod (1939) and Domar (1946). By allowing smooth factor substitution and wage-price flexibility, the capital-output ratio is

* Delano Villanueva is Assistant to the Director of the Middle Eastern Department. He received his Ph.D. from the University of Wisconsin.
The author wishes to thank Donald Mathieson and Julio Santaella for useful comments and Brooks Dana Calvo and Ravina Malkani for efficient research assistance.

made a monotonic function of the capital-labor ratio. The growth rate of the capital stock (the warranted rate) adjusts to the exogenously given growth rate of the labor force (the natural rate) to maintain full-employment real output.

The Solow-Swan model, however, has certain equilibrium properties that bother many growth theorists: an increase in the saving rate, while raising the level of per capita real income, has no effect on the growth rate of output. This surprising result on growth neutrality has a simple explanation: although a higher saving rate raises the growth rate of output by increasing the investment rate, the increase in economic growth occurs only during the transition toward the next equilibrium; sooner or later, the labor input becomes a bottleneck, restricting further output expansion. The growth rate of output would eventually fall back to the constant natural rate of growth.

The time it takes the economy to reach this balanced growth path is of considerable interest—particularly to policymakers. In the context of the Solow-Swan model, if the objective of economic policy is to raise the equilibrium level of per capita real income (for example, by raising the government saving rate), a fast adjustment would be desirable.

Using a Cobb-Douglas production function with constant returns to scale and Harrod-neutral technical progress, Sato (1963) has shown that the time required for the Solow-Swan model to reach equilibrium is about a hundred years![1] Moreover, the lower the rate of depreciation or the higher the share of capital, the slower the adjustment. An intuitive explanation for these results is that a slower rate of depreciation or a larger share of capital would enable firms to substitute capital for labor and thus postpone for a longer period the bottleneck posed by a fixed rate of labor growth.

The Solow-Swan model's prediction that the rates of saving, depreciation, and population growth as well as government policies cannot affect the equilibrium growth rate of per capita real income, which is fixed by an exogenously determined rate of labor-augmenting technological progress, appears counterfactual. It seems reasonable to conjecture that, over the long haul, countries that promote saving and investment, reduce the depreciation of the capital stock, and create more open trading systems tend to grow faster and that those with rapid population growth, sluggish expansion in expenditures on human development and basic needs, and high ratios of government deficits to GDP tend to grow slower.

[1] Such a slow adjustment would render somewhat irrelevant the equilibrium behavior of the model because of the likelihood that the other parameters of the system would have changed in the interim.

The relatively slow adjustment of the Solow-Swan model toward its steady state is partly due to the (assumed) inability of the natural rate to adjust to changes in capital intensity as the economy moves from one equilibrium to another in response to an exogenous shock. It seems plausible to consider that a partly endogenous natural rate, via learning through experience, would contribute to a faster speed of adjustment. If so, the limiting behavior of the Solow-Swan model would assume much more relevance to policymakers.

This study is both theoretical and empirical and belongs to the class of new "endogenous growth" models.[2] It is a variant of Conlisk's (1967) endogenous–technical change model and Arrow's (1962) "learning by doing" model, wherein experience (measured in terms of either cumulative past investment or output) plays a critical role in raising labor productivity over time. The presence of learning through experience has three major theoretical consequences. First, equilibrium growth becomes endogenous and is influenced by government policies.[3] Second, the speed of adjustment to growth equilibrium is faster, and enhanced learning further reduces adjustment time. Third, both equilibrium economic growth and the optimal net rate of return to capital are higher than the sum of the exogenous rates of technical change and population growth.

The endogenous growth model's equilibrium behavior is found to be consistent with the substantial diversity in per capita growth patterns observed among countries. Such diverse growth experiences, which are predicted by the model, can be explained by differences in saving rates, ratios of government deficits to GDP, population growth rates, and certain parameters that influence the learning coefficient, such as changes in openness to world trade and growth in government outlays on education and health.

I. Endogenous Growth

The model is summarized by the following relationships:

$$Y = F(K, N) = Nf(k), \tag{1}$$

$$dK/dt = s(\theta, .)Y - \delta(\mu)K, \tag{2}$$

[2] See, among others, Romer (1986), Lucas (1988), Becker, Murphy, and Tamura (1990), Grossman and Helpman (1990), and Rivera-Batiz and Romer (1991).

[3] Equilibrium growth in Arrow's learning-by-doing model, although a function of the "learning coefficient," nevertheless remains independent of the saving rate and the depreciation rate. See footnote 5 for details.

$$dL/dt = nL, \tag{3}$$

$$dT/dt = \alpha(\chi, \xi, \omega, .)K/L + \lambda T, \tag{4}$$

$$N = TL, \tag{5}$$

$$k = K/N; \tag{6}$$

where the variables are defined as

Y = real GDP,
K = capital stock,
N = labor (manhours in efficiency units),
L = population (manhours),
T = labor productivity or technical change multiplier (index number),
k = ratio of K to N,
s = ratio of real saving to Y,
δ = depreciation of capital,
α = learning coefficient;

and the parameters are defined as

n = population growth rate,
χ = change in ratio of foreign trade (sum of exports and imports) to GDP,
ξ = growth rate of real government expenditures on education and health,
ω = growth rate of real government expenditures on social security and housing,
μ = growth rate of real government expenditures on operations and maintenance,
θ = ratio of government deficits to GDP,
λ = rate of exogenous labor-augmenting technical change,
$d(.)/dt$ = time derivative.

Equation (1) is a standard neoclassical production function satisfying the Inada (1963) conditions.[4] Equation (2) is the expression for capital accumulation: the increment in the capital stock is equal to gross domestic saving less depreciation. The proportion s of GDP saved and invested is assumed to be sensitive to government policies, in particular to θ, the ratio of the fiscal deficit to GDP. High values of θ directly lower s, as the

[4] $\lim \partial F/\partial K = \infty$ as $K \rightarrow 0$; $\lim \partial F/\partial K = 0$ as $K \rightarrow \infty$; $f(0) \geq 0$; $f'(k) > 0$; and $f''(k) < 0$.

public sector dissaves. There are indirect effects as well. High levels of θ indicate large government borrowings from financial markets. Either through higher interest rates or lower credit availability, private sector capital accumulation is adversely affected. Thus, it is assumed that $s'(\theta) < 0$.

There are other (unspecified) factors affecting s. For example, interest rate liberalization may increase the private saving rate, which would tend to pull up aggregate s, but may also entail increases in the rate of government dissaving in the presence of a large stock of public debt, which would drag down total s both directly and indirectly (via negative effects on the private saving-investment rate). It is also assumed that $\delta'(\mu) < 0$—the rate of depreciation, δ, is a negative function of the real growth of expenditures on operations and maintenance, μ. In other words, a higher μ lowers the rate of depreciation of existing capital stock K. The population grows at an exogenously constant rate n in equation (3).

The key relationship in the model is equation (4). It postulates that technical change dT/dt improves with the capital stock per capita K/L. Output per capita Y/L can be used instead. For example, manhours in the production of an airframe (the structural frame of an airplane) during the 1930s tended to decline with the number of airframes produced. A more current example is the introduction of both high-speed and personal computers, which has improved the productivity of engineers and scientists (including economists). Since $(dT/dt)/T$ is a function of $Y/TL = Y/N = f(k)$, using K/L is equivalent to using Y/L as the forcing variable behind improvements in labor productivity. The parameter α is the learning coefficient. If $\alpha = 0$, T grows exogenously at a constant rate λ and the endogenous growth model collapses into the Solow-Swan model. (The restrictions $\alpha \geq 0$ are assumed and empirically tested in a later section. Since the assumption that $\alpha > 0$ is crucial to the arguments in this paper, the extended discussion of its rationale is useful.)

The Solow-Swan model's characterizing assumption $\alpha = 0$ may be true in a world devoid of technical change, since labor supply may be measured by the size of population. In this case, it may be plausible to assume that labor has no endogenous growth component, since population in many countries appears to grow independently of the economic system. But the real world is one of continuous technical change. While a portion of this change may be exogenous, some technical change is clearly endogenous and partly labor augmenting. Workers learn through experience, and their productivity is likely to be enhanced by the arrival of new and advanced capital goods. That is, the endogenous growth model's

assumption that $\alpha > 0$ seems more plausible than the Solow-Swan model's assumption that $\alpha = 0$.[5]

In the restriction $\alpha > 0$, the learning coefficient α is allowed to vary positively with changes in the ratio of foreign trade to GDP, χ; real growth of outlays on education and health, ξ; social security, housing, and recreation, ω; and other unspecified factors. The role of rapid growth of foreign trade in stimulating a higher learning coefficient is twofold.[6] First, the import-export sector serves as a vehicle for technology transfer through the importation of advanced capital goods, as elucidated by Bardhan and Lewis (1970), Chen (1979), and Khang (1987), and as a channel for positive intersectoral externalities through the development of efficient and internationally competitive management, the training of skilled workers, and the spillover consequences of scale expansion (Keesing (1967), and Feder (1983)). Second, rising exports relieve the foreign exchange constraint. The importation of technologically superior capital goods is enhanced by growing export receipts and higher flows of foreign credits and direct investment, which take into account the country's ability to repay out of export earnings.[7]

It is also reasonable to posit that accelerated growth of real outlays on education and health would be associated with a higher learning potential of labor, as would growth in real expenditures on social security, housing, and recreation. Finally, equations (5) and (6) are standard definitional relations involving N and k.

Reduced Model

The growth rate of the capital stock is derived by dividing equation (2) by K, using equations (1) and (6):

[5] Arrow's (1962) learning-by-doing model has a steady-state solution for the growth rate of output equal to $(\lambda + n)/(1 - \alpha)$, where the technical change function is $(dT/dt)/T = \alpha(dK/dt)/K + \lambda, 0 < \alpha < 1$. Although steady-state growth is thus a multiple of $\lambda + n$, growth remains independent of s and δ; besides, this model has the property that $\partial(g^* - n)/\partial n = \alpha/(1 - \alpha) > 0$. That is, an increase in population growth raises equilibrium rate of per capita growth. This proposition is rejected by the empirical finding reported later that an increase in the rate of population growth depresses the average growth rate of per capita output during 1975–86 in a sample of 36 developing countries from five regions.

[6] See the discussion on the production linkage summarized in Khan and Villanueva (1991). Edwards (1992) and Knight, Loayza, and Villanueva (1993) present evidence on the relationship between trade openness and economic growth.

[7] The transfer of efficient technologies and the availability of foreign exchange have featured prominently in recent experiences of rapid economic growth (Thirlwall (1979)).

$$(dK/dt)/K = s(\theta, .)f(k)/k - \delta(\mu). \tag{7}$$

The growth rate of efficient labor is derived by differentiating equation (5) with respect to time, using equations (1) and (3)–(6):

$$(dN/dt)/N = \alpha(\chi, \xi, \omega, .)k + n + \lambda. \tag{8}$$

Differentiating equation (6) with respect to time and substituting equations (7) and (8), the growth rate of the capital-labor ratio k is thus equal to

$$\begin{aligned}(dk/dt)/k &= (dK/dt)/K - (dN/dt)/N \\ &= s(\theta, .)f(k)/k - \alpha(\chi, \xi, \omega, .)k - [n + \lambda + \delta(\mu)]. \end{aligned} \tag{9}$$

The reduced model, equation (9), is a single differential equation involving the variables $(dk/dt)/k$ and k alone.

Per capita income, Y/L, grows according to

$$(dY/dt)/Y - n = \alpha(\chi, \xi, \omega, .)k + \pi(k)(dk/dt)/k + \lambda, \tag{10}$$

which is also a single-valued function of k. Here, π is the share of income going to capital; in general, it is a function of k.[8] The equilibrium capital intensity k^* is the root of equation (9) equated to zero:

$$s(\theta, .)f(k^*)/k^* - \alpha(\chi, \xi, \omega, .)k^* - [n + \lambda + \delta(\mu)] = 0. \tag{11}$$

And the equilibrium growth rate of per capita income is given by

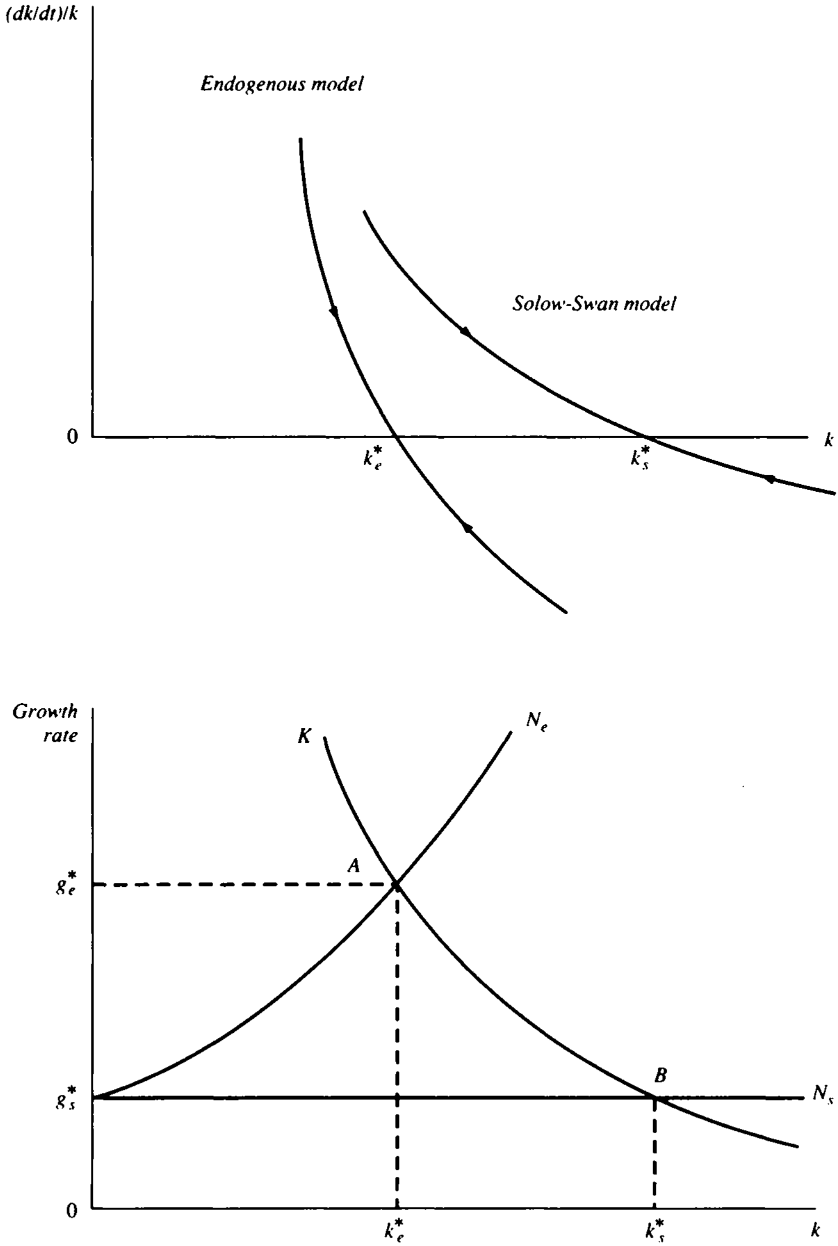
$$\begin{aligned}[(dY/dt)/Y]^* - n = [(dK/dt)/K]^* - n &= s(\theta, .)f(k^*)/k^* \\ &\quad - [n + \delta(\mu)] \qquad (12a) \\ &= [(dN/dt)/N]^* - n = \alpha(\chi, \xi, \omega, .)k^* \\ &\quad + \lambda. \qquad (12b)\end{aligned}$$

Stability

Given the Inada conditions on the production function, equations (7)–(9) are depicted according to Figure 1. The upper panel graphs

[8] For a degree-β homogeneous production function $Y = F(K, N)$, $\pi(k) = kf'(k)/\beta f(k)$. The sign of $\pi'(k)$ follows the sign of $\epsilon(k) - 1$, where $\epsilon(k) = f'(k)[\beta f(k) - kf'(k)]/k[(\beta - 1)f'(k)^2 - \beta f(k)f''(k)]$ is the elasticity of substitution. If F is Cobb-Douglas, $\pi(k) = \rho$, where α is the constant exponent of K, and $\epsilon(k) = 1$. If F has constant elasticity of substitution (CES), $\pi(k) = 1/[1 + (1 - \rho)(1/\rho)k^{-\sigma}]$ and $\epsilon(k) = 1/(1 - \sigma)$. Notice that if $\sigma = 0$, CES reduces to Cobb-Douglas. Also see footnote 8.

Figure 1. *Endogenous and Solow-Swan Growth Models*



equation (9), and the lower panel graphs equations (7) and (8). The downward slopes of the curves representing equations (7) and (9) and the upward slope of the curve representing equation (8) follow from the assumption of a positive but diminishing marginal product of capital. The reasons why the $(dk/dt)/k$ curve lies partly in the first quadrant and partly in the fourth quadrant in Figure 1 are given by the other Inada conditions—that is, for some initial values of the capital-labor ratio, it is possible for capital to grow either faster or slower than labor. It is obvious by inspection that, at any point on the $(dk/dt)/k_e$ curve, the economy would move in the direction indicated by the arrows. Thus, k tends to settle at an equilibrium value k_e^*, which is globally stable. Other points along the curve imply nonzero rates of change in k, and k will change toward k_e^*. In Figure 1, for example, points to the left of k_e^* imply positive values of $(dk/dt)/k$. This means that K is growing faster than N, and the ratio K/N will rise. The increase in k lowers the income-capital ratio, and hence the saving- and investment-capital ratios. The growth of K slows. Meanwhile, a higher k induces an increase in labor-augmenting technical change through enhanced learning and experience. The growth of N is stimulated. This process would continue until the growth rates of K and N converge at the stationary value k_e^*.[9] At this equilibrium point, K and N would grow at the same rate g_e^*, and by the constant returns assumption output Y would also grow at this rate, given by equations (12a) and (12b).

Equilibrium Capital Intensity and Growth

The Solow-Swan and endogenous growth models are graphically portrayed in Figure 1. In the lower panel, the natural rate schedule, N_e, is upward sloping in the endogenous growth model, owing to the presence of learning-by-doing and the assumption of a positive marginal product of capital. The natural rate schedule in the Solow-Swan model is shown as the horizontal line N_s with vertical height equal to a constant growth rate $g_s^* (= \lambda + n)$. The warranted-rate schedule K is assumed to be identical in the two models.

In the upper panel, reflecting the different natural rate schedules, the capital accumulation schedules assume the shape and intersection with the k-axis indicated by the two curves, with $(dk/dt)/k_s$ flatter than and to the right of $(dk/dt)/k_e$. The equilibrium positions of the two types of models are indicated by the points A and B, respectively, in the lower

[9] The opposite sequence of events is true for points to the right of k_e^*, implying negative values of $(dk/dt)/k$).

Table 1. *Comparative Effects of Structural Parameters on Equilibrium Values of Capital Intensity* (k^*) *and the Per Capita Growth Rate* ($g^* - n$)

	Endogenous growth		Solow-Swan	
An increase in	k^*	$g^* - n$	k^*	$g^* - n$
Saving rate (s)	+	+	+	0
Ratio of foreign trade to GDP (χ)	+	+	na	na
Growth in real spending on education and health (ξ)	−	+	na	na
Growth in real spending on social security (ω)	−	+	na	na
Growth in real spending on operations and maintenance (μ)	+	+	na	na
Ratio of fiscal deficit to GDP (θ)	−	−	na	na
Population growth (n)	−	−	−	0
Exogenous technical change (λ)	−	+	−	+

Notes: + = increase; − = decrease; 0 = no change; na = not applicable.
Source: For endogenous growth model, equations (11) and (12a)–(12b) in text. For Solow-Swan model, same set of equations with α set equal to zero.

panel. The growth rate of output is higher in the endogenous growth model by the magnitude $\alpha(\,.\,)f(k^*)$: that is, $g_e^* > g_s^*$. The capital-labor ratio, however, is lower in the endogenous growth model ($k_e^* < k_s^*$); the growth rate is higher because of induced learning-by-doing. The model's capital intensity level is lower because of a higher level of the effective labor input.

Comparative Dynamics

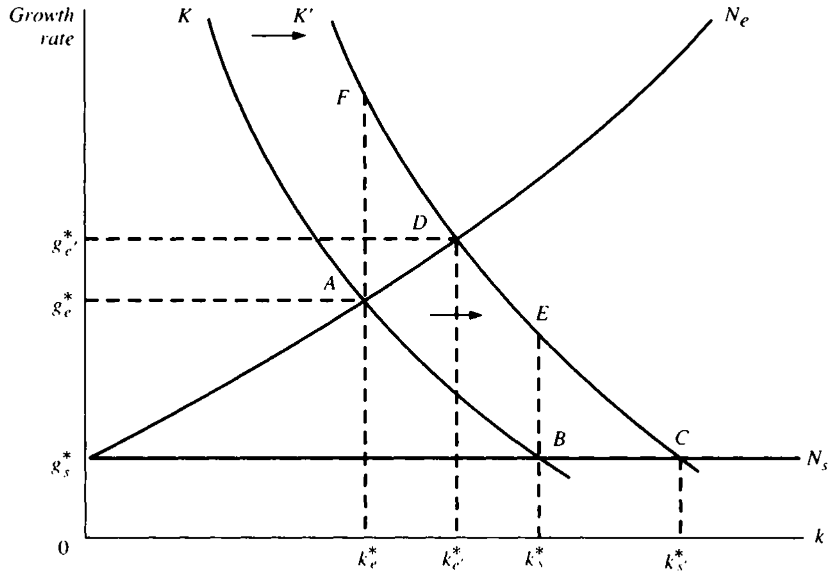
Table 1 summarizes the qualitative effects of changes in the structural parameters on the equilibrium capital intensity, k^*, and on the equilibrium per capita growth rate of income, $g^* - n$. Algebraically, the partial derivatives of k^* and $g^* - n$ with respect to any structural parameter may be obtained by the differentiation of equations (11) and (12a)–(12b).

Effects of a Higher Saving Rate[10]

The effects of an increase in the saving rate s on the transitional and equilibrium growth rates of output can be analyzed using Figure 2, in which the initial equilibrium positions in the endogenous growth and

[10] The effects of a reduction in the rate of depreciation—exogenously in the Solow-Swan model and endogenously in the endogenous growth model via a higher rate of real expenditures on operations and maintenance—are similar.

Figure 2. *Effects of an Increase in the Saving Rate*



Solow-Swan models are indicated by points A and B, respectively. An increase in the saving rate shifts the warranted-rate curve to K' in either model. The new equilibrium positions are indicated by points D in the endogenous growth model and C in the Solow-Swan model. In both models, the capital-labor ratio rises, although the new ratio remains lower in the endogenous growth model than in the Solow-Swan model, owing to positive learning-by-doing. However, the new equilibrium growth rate increases in the endogenous growth model but remains unchanged in the other. The discussion below traces the adjustment dynamics to the new growth equilibrium in the two models in the wake of an increase in the saving rate.

During the transition between equilibrium points B and C, the rate of output growth in the Solow-Swan model is momentarily higher—by EB—than the natural rate g^*_s because the warranted rate rises owing to a higher ratio of saving to income.[11] The capital-labor ratio begins to rise,

[11] The transitional growth rate of output, $(dY/dt)/Y$, is equal to $\lambda + n + \pi(k)\dot{k}/k$, where $\pi(k) = kf'(k)/f(k)$. Now, both $\pi(k)$ and $\dot{k}/k$ are positive anywhere between k^*_s and $k^{*\prime}_s$. It follows that $(dY/dt)/Y > \lambda + n$ during the transition from B to C. At either B or C, $\pi > 0$ and $\dot{k}/k = 0$, so that $(dY/dt)/Y = \lambda +$

which slows the warranted rate. Since the natural rate is completely independent of the capital-labor ratio, only the warranted rate adjusts (downward) along the segment EC. Over time, labor becomes a bottleneck, and the growth rate slows (converges) to the constant natural rate $g_s^*(= n + \lambda)$ at C. At this point, the capital-labor ratio stops rising and settles at a new and higher level $k_s^{*\prime}$. Thus, the effect of an increase in the saving rate is to raise the equilibrium capital-labor ratio (from k_s^* to $k_s^{*\prime}$)[12] because of the permanent upward shift of the warranted-rate curve associated with an increase in the saving rate.

In the endogenous growth model, an increase in the saving rate shifts the equilibrium from A to D. At the starting position A, capital would grow faster than labor (by FA), and the capital-labor ratio would rise (from k_e^* toward $k_e^{*\prime}$). As this happens, the marginal and average product of capital would fall, thus lowering the level of saving per unit of capital and slowing the warranted rate (downward along FD). On the other hand, the natural rate, instead of remaining constant as in the Solow-Swan model, would accelerate (from A to D along the N_e curve) because of a higher rate of labor-augmenting technical change associated with a rising capital-labor ratio. This process would continue until the warranted and natural rates are equalized—through a continuous increase in the capital-labor ratio—at the new equilibrium value $k_e^{*\prime}$ at D, at which point the warranted rate would have fallen to the new and higher value of the natural rate, equal to the new and higher growth rate of output $g_e^{*\prime}(> g_e^*)$.

Effects of Openness, Human Development Spending, and Technical Change

The effects of these factors can be analyzed with the help of Figure 3. Since many of these parameters are absent from the Solow-Swan model,[13] the illustrations refer only to the endogenous growth model. Changes in the ratio of foreign trade (sum of exports and imports) to GDP and growth in real outlays on education, health, social security, housing, and recreation are reflected in changes in the learning coefficient α, while

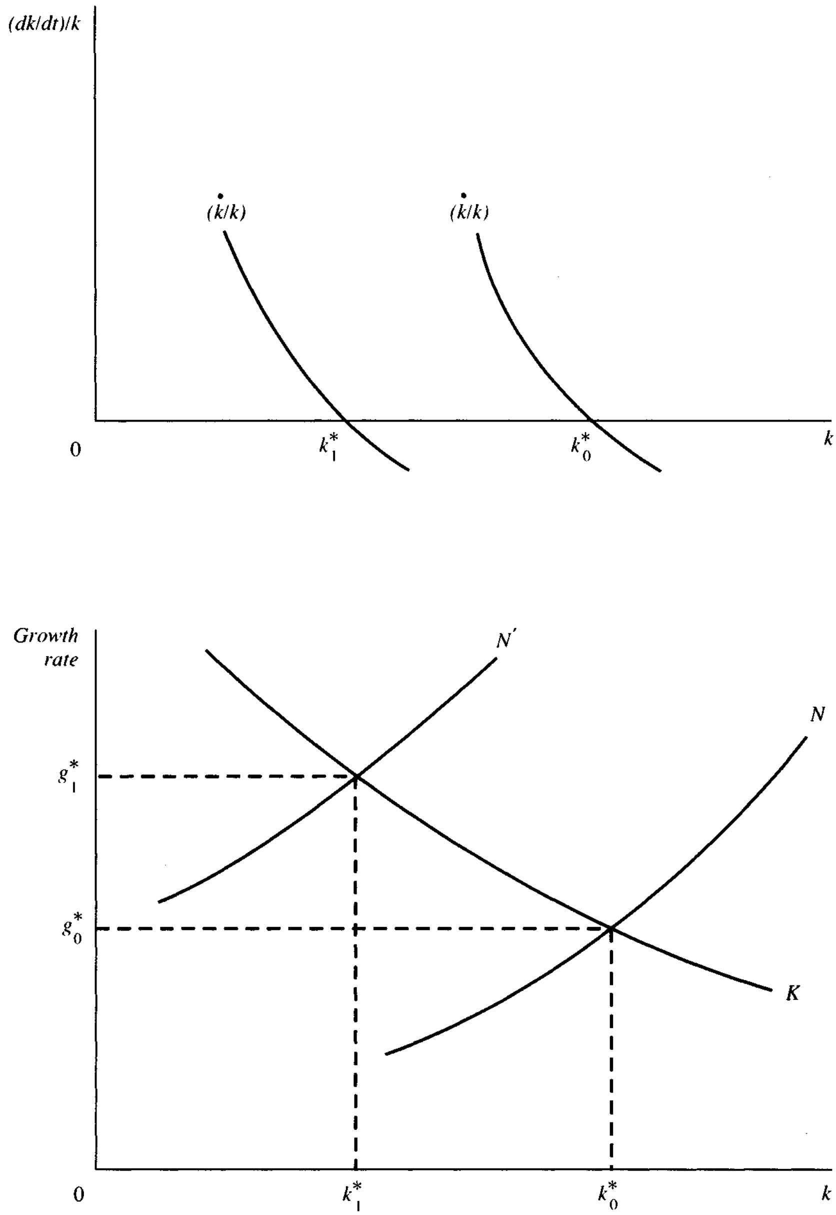
---

n at either equilibrium point. The convergence property of neoclassical growth models can be demonstrated with the aid of Figure 2. As the initial capital intensity (or initial income per worker) moves farther to the left of $k_s^{*\prime}$ (gets smaller), the average growth rate of per capita income rises—that is, the vertical line increases between N_s and any point on the K' curve corresponding to the initial level of capital intensity.

[12] And thus it raises the equilibrium level of real income per efficient worker.

[13] One parameter that is present is the exogenous rate of technical change λ, whose effects on capital intensity and per capita growth are similar in the two models.

Figure 3. *Effects of Increased Openness and Expenditures on Human Development in the Endogenous Growth Model*



changes in the exogenous rate of technical change λ enter the natural rate schedule directly.

An increase in any of these parameters shifts the capital accumulation schedule to the lower left (upper panel) and the natural rate schedule to the upper left (lower panel). With reference to Figure 3, the adjustment dynamics are the following. After the parametric increase, the rate of change in k is negative at the old equilibrium value k_0^*. This means that the natural rate is above the warranted rate, as shown in the lower panel. Thus, the level of capital intensity begins to fall toward k_1^*. As k falls, income per unit of capital rises, stimulating saving and investment, and the warranted rate goes up. At the same time, a lower stock of capital reduces the rate at which technological progress is taking place, depressing the natural rate. This process continues until the two rates meet at k_1^*, where the rate of change in k is again zero. The new equilibrium position is characterized by a lower level of capital intensity and a higher growth rate of per capita output and income.

Effects of Fiscal Deficits and Population Growth

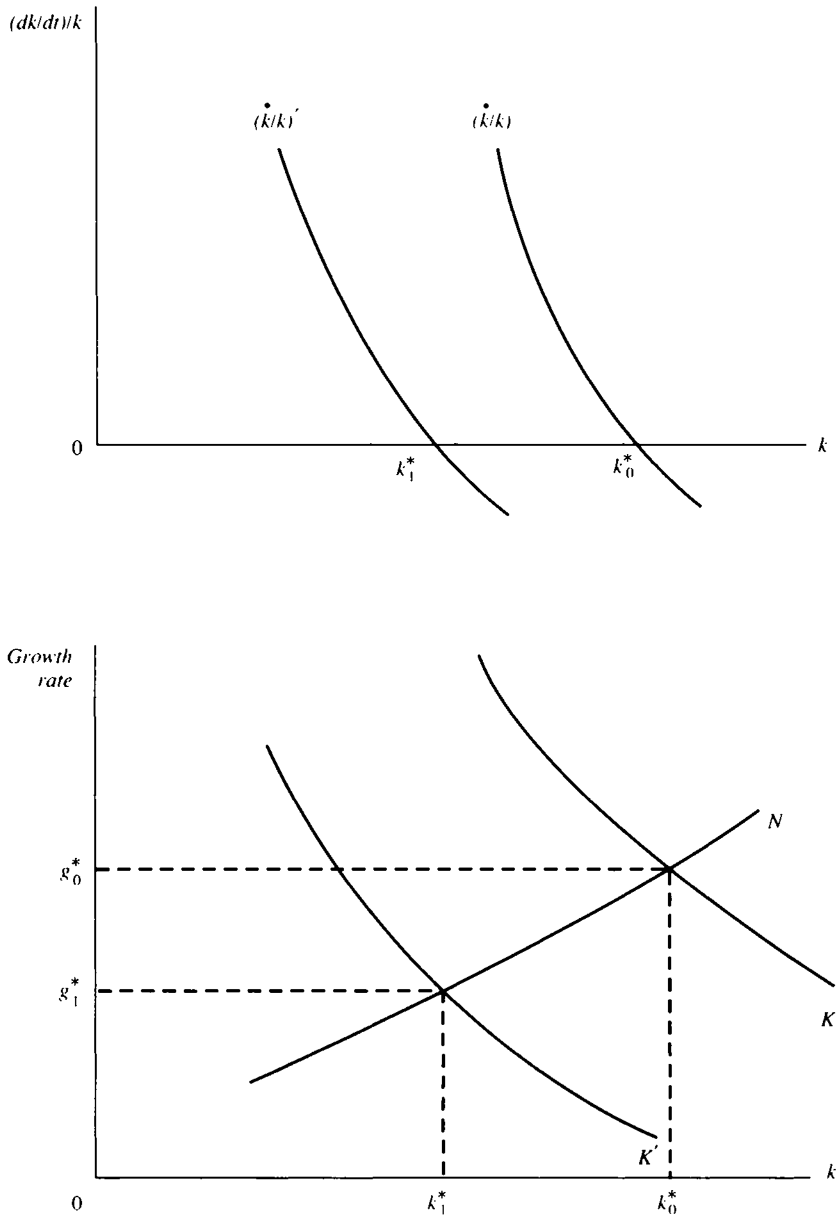
Finally, Figure 4 illustrates the effects of increases in the ratio of the fiscal deficit to GDP and in the rate of population growth on equilibrium capital intensity and on the growth rate of per capita output in the endogenous growth model. An increase in population growth or in the rate of government dissaving[14] (by lowering the saving rate) shifts the capital accumulation schedule to the lower left in both panels. At the old equilibrium capital intensity, the rate of change in k turns negative (upper panel), implying that the warranted rate falls short of the natural rate (lower panel). As k falls, income per unit of capital increases, raising saving and investment, and hence the warranted rate. At the same time, the natural rate decreases, because a lower k induces a lower rate of learning. This process continues until the economy settles at a new equilibrium position characterized by a convergence of the warranted and natural rates, a lower level of capital intensity, and a slower growth rate of per capita income.

II. Optimal Long-Run Growth

In the long run, output per unit of effective labor is $y^* = f(k^*)$. If y^* is considered a measure of the standard of living, and since $f'(k^*) > 0$,

[14] As noted earlier, as the public sector dissaves, fewer resources will be available to accumulate capital. Moreover, the ensuing large government borrowings from financial markets tend to raise interest rates or lower available credit, adversely affecting private capital accumulation.

Figure 4. *Effects of Increases in Ratio of Fiscal Deficits to GDP and in Population Growth*



it is possible to raise living standards by increasing k^*. This can be done by adjusting the saving rate—for example, by lowering the ratio of the fiscal deficit to GDP. If consumption per unit of effective labor (or any monotonically increasing function of it) is taken as a measure of the social welfare of the society, the saving rate that maximizes social welfare by maximizing long-run consumption can be determined. Phelps (1966) refers to this path as the "Golden Rule of Accumulation."

Consumption C per unit of effective labor is $c = C/N = Y/N - S/N$, where S is saving. Y/N is $f(k)$ and $S = I = dK/dt + \delta(\mu)K$, where I is investment. Thus,

$$c = f(k) - [(dK/dt + \delta(\mu)K)]/N = f(k) - k[(dK/dt)/K] - \delta(\mu)k.$$

On the balanced growth path,

$$(dK/dt)/K = \alpha(\,.\,)k^* + \lambda + n,$$

where $\alpha(\,.\,) = \alpha(\chi, \xi, \omega, \,.\,)$. Thus,

$$c^* = f(k^*) - [\alpha(\,.\,)k^* + \lambda + n + \delta(\mu)]k^*. \tag{13}$$

Maximizing c^* with respect to s,

$$\partial c^*/\partial s = \{f'(k^*) - 2\alpha(\,.\,)k^* - [\lambda + n + \delta(\mu)]\}\partial k^*/\partial s = 0. \tag{14}$$

Since $\partial k^*/\partial s > 0$, the Golden Rule condition is

$$f'(k^*) - \delta(\mu) = g^*(k^*) + \alpha(\,.\,)k^*, \tag{15}$$

where $g^*(k^*) = \alpha(\,.\,)k^* + \lambda + n$ is the equilibrium growth rate of output. The second-order condition for a maximum is satisfied, since

$$\partial^2 c^*/\partial s^2 = [f''(k^*) - 2\alpha(\,.\,)]\partial k^*/\partial s < 0. \tag{16}$$

Equation (15) says that for social welfare to be maximized the saving rate should be raised to a point where the net rate of return to capital (which is equal to capital's marginal product less depreciation) equals the long-run growth rate of output plus the product of the learning coefficient and the equilibrium capital intensity. The second term is nothing more than the endogenous component of labor-augmenting technical change—the component of $(dT/dt)/T$ induced by any learning that occurs at a higher level of capital intensity, which, in turn, is caused by a higher saving rate. If there is no learning ($\alpha = 0$), equation (15) reduces to $f'(k^*) - \delta = \lambda + n$, which is the familiar Golden Rule result from standard neoclassical growth theory. It is evident that the optimal net rate of return to capital should be higher than $\lambda + n$ when $\alpha > 0$—when there

is learning-by-doing—because of two factors. First, when the saving rate s is raised, the equilibrium growth g^* will be higher than $\lambda + n$, by the amount $\alpha(.)\partial k^*/\partial s$. Second, capital should be compensated for the effect on equilibrium output growth through the induced learning term $\alpha(.)k^*$.

An alternative interpretation of the above Golden Rule can be given. A standard neoclassical result is that the optimal saving rate s should be set equal to the income share of capital π. With endogenous learning-by-doing, the optimal saving rate should be set at a fraction of π, the fraction being equal to $(g^* + \delta)/[g^* + \delta + \alpha(.)k^*]$.[15] Here, $g^* + \delta + \alpha(.)k^* = f'(k^*)$—given by equation (15)—is the (gross) social marginal product of capital, inclusive of the positive externalities arising from the learning associated with capital accumulation in the endogenous growth model. Equivalently put, income going to capital as a share of total output should be a multiple of the amount saved and invested in order to compensate capital for the additional output generated by endogenous growth and induced learning. A value of π equal to s, implicit in the standard model, would undercompensate capital and thus be suboptimal from a societal point of view.

III. Speed of Adjustment Toward Equilibrium

The equilibrium results derived in the preceding section would not be relevant to the real world if the time period for the model to reach its equilibrium were unduly long. There are three approaches to the analysis of adjustment dynamics in the speed-of-approach literature:

—analytical approach, which garners less explicit results but does not resort to a full-scale numerical simulation;

[15] Equations (11) and (15) and the definition $\pi = k^*f'(k^*)/f(k^*)$ are used to derive this result. When $\alpha = 0$, the proportionality factor assumes a value of unity, and the standard neoclassical result holds. In terms of the parametric values assumed in the simulations reported in Table 2 below, when the learning coefficient α is greater than zero, the optimal saving rate should be set at about three-quarters of the assumed income share of capital π, or at 0.3 when $\pi = 0.4$. The simulations also show that the higher is the learning coefficient, the lower is the optimal saving rate as a proportion of capital's income share. According to the standard model, the optimal saving rate should always be set equal to π, which is at 0.4 in the numerical examples. The higher saving rate implied by the standard model owes to its neglect of endogenous growth and positive externalities through the learning-by-doing associated with saving and capital accumulation. By contrast, in the endogenous growth model the economy benefits from such endogenous growth and positive externalities, so that a smaller saving-investment rate is all that is required (relative to the rate required by the standard model).

—simulation, such as the work of Sato (1963), which uses a specific functional form for the production function and representative values of the structural parameters and which calculates adjustment paths from hypothetical disequilibria to obtain estimates of the time (in years) needed to reach equilibrium; and

—empirical approach, which examines whether the model's equilibrium predictions accord with observed growth patterns of real economies over reasonably long periods.

Analytical Approach

The negative slope of the $(dk/dt)/k_e$ curve (see Figure 1) at the equilibrium capital intensity k_e^* is a measure of the local adjustment speed. The steeper the slope, the faster the steady state k_e^* is reached. The absolute value of the slope at k_e^* may be obtained by differentiating equation (9) with respect to k and evaluating at k_e^*:

$$\begin{aligned} V &= |\{(d/dk)[(dk/dt)/k]\}^*| \\ &= (n + \lambda + \delta + \alpha k_e^*)[(1 - \pi(k_e^*)]/k_e^* + \alpha. \end{aligned} \tag{17}$$

The key feature of the endogenous growth model that distinguishes it from the Solow-Swan model is the assumed presence of learning-by-doing, represented by a positive learning coefficient α. In the absence of learning ($\alpha = 0$), equation (17) reduces to the Solow-Swan expression. It is obvious from equation (17) that with $\alpha > 0$ the slope of the $(dk/dt)/k_e$ curve is steeper than the slope of the $(dk/dt)/k_s$ curve (when $\alpha = 0$). Thus, the endogenous growth model takes relatively less time to reach equilibrium. Moreover, it can be shown that enhanced learning—represented by an increase in α—would further reduce the adjustment time, provided that the elasticity of substitution is not less than one, such as when the production function is CES. This can be seen by differentiating equation (17) with respect to α, which yields

$$\begin{aligned} \frac{\partial V}{\partial \alpha} = \{[(1 - k_e^*)\alpha - (n + \lambda + \delta)][1 - \pi(k_e^*)] \\ - \pi'(k_e^*)(n + \lambda + \delta + \alpha k_e^*)\}(1/k_e^*)(\partial k_e^*/\partial \alpha), \end{aligned}$$

which is positive if the production function is CES (in which case, $\pi'[k_e^*] \geq 0$) and if $k_e^* \geq 1$ (the equilibrium capital per effective worker is not less than a unit of the currency). It has been shown earlier that $\partial k_e^*/\partial \alpha < 0$. The simulations using a Cobb-Douglas production function (a special case of CES) reported in Table 2 below show the same results.

The above results can be given an intuitive interpretation. It has been shown that the equilibrium growth rate of output is $[(dN/dt)/N]^* =$

Table 2. *Estimated Period of Adjustment in Years as y_t Approaches a Limit of y_∞, Using Various α*[a]

Adjustment ratio (p_t)	Adjustment from above ($y_0 - y_\infty > 0$)		Adjustment from below ($y_0 - y_\infty < 0$)	
$\alpha = 0$[b]	$y_0 = 0.045$	$y_0 = 0.035$	$y_0 = 0.015$	$y_0 = 0.025$
0.25	33.8	9.3	3.9	5.4
0.50	55.7	21.1	10.1	13.6
0.75	80.3	39.6	22.6	28.5
0.90	105.9	62.6	41.7	49.3
$\alpha = 0.01$[b]	$y_0 = 0.08$	$y_0 = 0.07$	$y_0 = 0.01$	$y_0 = 0.05$
0.25	6.4	4.3	1.5	2.6
0.50	13.5	9.8	3.9	6.5
0.75	23.3	18.6	9.1	13.6
0.90	34.7	29.4	17.5	23.5
$\alpha = 0.02$[b]				
0.25	3.8	3.0	1.3	2.1
0.50	8.4	6.9	3.3	5.1
0.75	15.3	13.2	7.6	10.5
0.90	23.6	21.2	14.2	18.0

[a] Analysis uses $a = 0.4$, $\delta = 0.04$, $\lambda = 0.005$, $n = 0.025$, and $s = 0.2$.
[b] When $\alpha = 0$, $k^* = 5.75$ and $y_\infty = 0.03$. When $\alpha = 0.01$, $k^* = 3.00$ and $y_\infty = 0.06$. When $\alpha = 0.02$, $k^* = 2.40$ and $y_\infty = 0.078$.

$[(dK/dt)/K]^*$. Both the natural and warranted rates adjust endogenously to changes in capital intensity. With the brunt of adjustment toward equilibrium being shared by changes in the natural rate, the time needed to reach equilibrium is much less in the endogenous growth model. In sharp contrast, the time required to reach equilibrium is much longer in the Solow-Swan model because the adjustment burden is borne entirely by changes in the warranted rate.

Simulation

The reduced model, equation (9), is

$$(dk/dt)/k = sf(k)/k - \alpha k - (n + \lambda + \delta).$$

Assuming a Cobb-Douglas form for $f(k) = k^a$, where $0 < a < 1$ is the exponent of the capital stock (in this particular case, also equal to capital's share in income π, which is constant and independent of k), the reduced model becomes

$$dk/dt = sk^a - \alpha k^2 - (n + \lambda + \delta)k = g(k). \tag{18}$$

The solution to this differential equation is complicated because it is a nonlinear function. However, a linear approximation is possible in the neighborhood of the steady-state constant value k^*:[16]

$$dk/dt = g(k^*) + g'(k^*)(k - k^*)$$
$$= [ask^{*a-1} - 2\alpha k^* - (n + \lambda + \delta)](k - k^*),$$

since $g(k^*) = 0$. Or,

$$dk/dt = A(k - k^*), \tag{19}$$

where $A = ask^{*a-1} - 2\alpha k^* - (n + \lambda + \delta) < 0$.[17]

Equation (19) is of a "variables separable" form, which can be separated as

$$[1/(k - k^*)]dx = Adt. \tag{20}$$

Integrating both sides,

$$\int[1/(k - k^*)]dx = At + constant,$$

$$\log(k - k^*) = At + constant,$$

$$k - k^* = constant\, e^{At},$$

$$k = k^* + Ce^{At}, \tag{21}$$

where C is a constant of integration.[18]

Substituting equation (21) into equation (19)—

$$(dk/dt)/k = A\{1 - [k^*/(k^* + Ce^{At})]\}. \tag{22}$$

Now, from equation (10), the growth rate of output is given by

$$(dY/dt)/Y = y_t = a(dk/dt)/k + y_\infty, \tag{23}$$

where

$$y_\infty = \alpha k^* + \lambda + n. \tag{24}$$

[16] The constant k^* is the unique root of expression (18) equated to zero: $sk^{*a} - \alpha k^{*2} - (n + \lambda + \delta)k^* = 0$. Given $s = 0.2, a = 0.4, \alpha = 0.01, n = 0.025$, $\lambda = 0.005$, and $\delta = 0.04$, k^* assumes the value of 3.00, and the balanced growth path is equal to an annual rate of 0.06. If $\alpha = 0$, as in the Solow-Swan model, and if the other parameters are unchanged, k^* solves to a higher level at 5.75 and balanced growth to a lower rate of 0.03 per annum.

[17] In a follow-up on the preceding footnote, a particular value for A equal to -0.0886 is obtained for $\alpha = 0.01$.

[18] Note that as t goes to infinity, the second term on the right-hand side of equation (21) goes to zero (since $A < 0$), and thus k approaches k^*.

Substituting equations (22) and (24) into equation (23)—

$$y_t = aA\{1 - [k^*/(k^* + Ce^{At})]\} + y_\infty. \quad (25)$$

Setting $y_t = y_0$ and $t = 0$ in equation (25),

$$y_0 = aA\{1 - [k^*/(k^* + C)]\} + y_\infty, \quad (26)$$

which can be solved for the constant C,

$$C = (y_0 - y_\infty)k^*/(y_\infty - y_0 + aA). \quad (27)$$

Substituting equation (27) into equation (25),

$$y_t = aA\big(1 - [\![k^*/\{k^* + [(y_0 - y_\infty)k^*/(y_\infty - y_0 + aA)]e^{At}\}]\!]\big) + y_\infty. \quad (28)$$

Next, define the adjustment ratio p_t as

$$p_t = (y_t - y_0)/(y_\infty - y_0). \quad (29)$$

Substituting equation (28) into equation (29) solves for the time t (in years) required to get a fraction p_t of the way from y_0 to y_∞, from which Table 2 is computed:

$$t = (1/A)\ \ln\{(1 - p_t)(y_\infty - y_0 + aA)/[(1 - p_t)(y_\infty - y_0) + aA]\}, \quad (30)$$

where ln is the natural logarithm operator.

Table 2 reveals that the adjustment times in an endogenous growth model are generally a quarter to a third of those in an exogenous growth model, depending on the value of the learning coefficient α.[19] For example, whereas an exogenous growth model ($\alpha = 0$) takes from 42 to 106 years to approach equilibrium growth, an endogenous growth model ($\alpha > 0$) takes anywhere from 14 to 35 years to achieve 90 percent adjustment to the steady-state growth path, depending on the learning coefficient α. (Table 2 alternately uses values of 0, 0.01, and 0.02 for α.)

Table 2 also illustrates the effects of an increase in the learning coefficient from 0.01 to 0.02: the equilibrium capital intensity falls from 3.0 to 2.4 and equilibrium growth rises from 6.0 to 7.8 percent annually; moreover, adjustment times are reduced by 30–50 percent.[20]

[19] Except for the learning coefficient α, the parameter values for the saving and population growth rates used in the simulation represent historical averages of the data in the sample of countries listed in the Appendix. The value of the income share of capital is within the range of available empirical estimates. The value of the rate of depreciation is a standard approximation used in the literature.

[20] These simulation results are confirmed by the earlier qualitative analysis of the endogenous growth model, summarized in Table 1.

Empirical Approach

The model's predictions about the per capita output growth and capital stock, which are summarized in Table 1, are reproduced below, with the directional impact given by the sign above each argument inside the two functions:

$$g^* - n = \psi(\overset{+}{s}, \overset{-}{\theta}, \overset{+}{\chi}, \overset{+}{\xi}, \overset{+}{\omega}, \overset{+}{\mu}, \overset{-}{n}, \overset{+}{\lambda}); \tag{31}$$

$$k^* = \phi(\overset{+}{s}, \overset{-}{\theta}, \overset{-}{\chi}, \overset{-}{\xi}, \overset{-}{\omega}, \overset{+}{\mu}, \overset{-}{n}, \overset{-}{\lambda}). \tag{32}$$

Equations (31) and (32) are nonlinear functions, in general. Without the fiscal deficit variable θ, a linear approximation to these two equations can produce coefficient estimates of arbitrary magnitude and significance. For example, suppose that growth rates initially rise and then fall as government expenditure continuously grows, with the attendant heavy financing burdens reflected in rising values of θ. In this case, positive coefficients on government expenditure will be obtained for linear regressions using data with low θ, negative coefficients will be obtained for those that rely on high θ, and coefficients biased toward zero will be obtained for linear regressions using both low and high θ. The endogenous growth model developed earlier and the linear regression results reported below thus include θ, the ratio of government deficits to GDP.

No data for k^* exist in developing countries, so that equation (32) cannot be estimated. However, since there are data on $g^* - n$, equation (31) can be tested. In general, the average per capita growth rate, $g^* - n$, is inversely related to the starting value of per capita real income, y_0—the familiar convergence property of neoclassical growth models (including the present one).[21] Thus, for empirical testing, the following linear specification can be considered:

$$g^* - n = a_0 + a_1 s + a_2 \chi + a_3 \xi + a_4 \omega + a_5 \theta + a_6 n$$
$$+ a_7 y_0 + a_8 \lambda + a_9 \mu. \tag{33}$$

Of the nine explanatory variables in equation (33), data on only the last two are unavailable. Recall that μ is the real growth of expenditures on the operation and maintenance of capital assets, while λ is the exogenous rate of labor-augmenting technological progress. The parameter λ can be interpreted as capturing all the unobserved country-specific factors that raise labor productivity—cultural, social, ethnic, political, and religious.

[21] See footnote 11.

Regional dummy variables are included below to reflect such factors. The unobserved series μ is assumed to enter the error term in a well-behaved manner. For present purposes, the following multiple regression can be estimated:

$$g^* - n = a_0 + a_1 s + a_2 \chi + a_3 \xi + a_4 \omega + a_5 \theta + a_6 n + a_7 y_0 + a_8 \, dummy. \tag{34}$$

The endogenous growth model's equilibrium predictions (where the learning coefficient $\alpha > 0$) are that $a_1, a_2, a_3, a_4 > 0$ and that $a_5, a_6, a_7 < 0$. The Solow-Swan model (where $\alpha = 0$) predicts that $a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = 0$ and $a_7 < 0$. The data set consists of annual averages of observations for the period 1975–86 for 36 developing countries from five regions (see the Appendix for details).

The regression results are reported below, where the insignificant coefficients on the regional dummies are suppressed (t-values are in parentheses):

$$\begin{aligned} g^* - n = \underset{(0.50)}{0.01} + \underset{(3.06)}{0.183s} + \underset{(2.43)}{0.038\chi} + \underset{(1.91)}{0.093\xi} + \underset{(1.54)}{0.063\omega} - \underset{(2.39)}{0.189\theta} \\ - \underset{(1.90)}{0.665n} - \underset{(2.59)}{0.000015y_0}; \end{aligned} \tag{35}$$

$R^2 = 0.7952$; standard error of estimate (s.e.e.) = 0.0144.

An R^2 of almost 0.8 is relatively high for a cross-country regression,[22] and all the regression coefficients have the expected signs. The coefficients on the saving rate, the ratio of foreign trade to GDP, the ratio of fiscal deficits to GDP, and the initial level of per capita income are statistically significant at the 5 percent level or better. The coefficients on the growth of real expenditures on education and health and on the rate of population growth are statistically significant at the 10 percent level or better. The coefficient on the growth of real expenditures on social security, housing, and recreation is marginally significant.

Since θ (government dissaving) is a part of total s, further discussion of the coefficients on s and θ would be useful. The endogenous growth model divides the total long-run impact of changes in s on $g^* - n$ into two components: (i) the element arising from changes in the private saving rate induced by changes in its determinants other than changes in

[22] Ramanathan (1982) notes that typical values of R^2 for equations estimating the growth performance of developing countries, using cross-country data, fall in the 0.3–0.4 range.

θ; and (ii) the composite factor stemming from changes in s directly as a result of changes in θ and indirectly via induced changes in the private saving rate. Component (i) is measured by the coefficient on s, while component (ii) is captured by the coefficient on θ. Since the estimates of these two coefficients are nearly identical (with opposite signs), the results suggest a symmetric response of $g^* - n$, though in opposite direction, either to a change in the private saving rate or to a change in the rate of government dissaving.

The empirical results clearly show that the following factors promote per capita economic growth: steady increases in the saving-investment rate, the ratio of foreign trade to GDP, and the growth of real expenditures on education and health. On the other hand, rapid population growth and a high ratio of fiscal deficit to GDP are followed by slow average growth of per capita output. There is also empirical support for the convergence property of the endogenous growth and Solow-Swan models—the significant negative relationship between the initial level of per capita real income and subsequent average growth.

IV. Conclusions

This paper has presented a simple neoclassical growth model with endogenous technical change and contrasted its equilibrium properties with those of the more standard growth model. It has found that, contrary to the predictions of the Solow-Swan model, the equilibrium growth rate of per capita output is influenced in a systematic way by changes in the private rates of saving, depreciation, and population growth, as well as by changes in public policies regarding trade liberalization, fiscal deficits, spending on human resource development, and net investment.

In the absence of learning-by-doing, the model's optimal net rate of return to capital is equal to the sum of the population growth n and the exogenous rate of labor-augmenting technical change λ; alternatively, the optimal saving rate should be set equal to the share of capital in aggregate output—the familiar Golden Rule theorems from standard optimal growth theory. With learning-by-doing, these standard Golden Rule results are revised: the optimal net rate of return to capital is higher than $n + \lambda$, or the optimal saving rate should be set at only a fraction of capital's income share because of endogenous growth and the induced learning that is associated with increases in the capital stock.

The analytic and simulation results appear to favor the endogenous growth over the Solow-Swan model. Simulations show that the speed of adjustment toward equilibrium is substantially faster in a model of en-

dogenous growth. Moreover, an increase in learning-by-doing further shortens the adjustment period. The empirical results also validate the endogenous growth model, particularly those relating to the positive growth effects of public policies aimed at promoting greater openness of the trading system, high saving rates, and rapid growth in expenditures on human development, and those relating to the negative growth effects of rapid population growth and high ratios of fiscal deficits to GDP. Finally, the convergence property of the endogenous growth model has been confirmed (as has the convergence of the Solow-Swan model). However, the result on the saving rate–growth relationship is tenuous, in view of the short time interval (12 years) of the sample. Since the realized growth dynamics in the Solow-Swan model over this relatively short period would also show a positive relationship, the empirical results would hardly invalidate the Solow-Swan approach, pending additional research. Efforts are currently under way to use the very long time series (1950–85) from Summers and Heston (1988) in testing the equilibrium relationships among the growth rates of per capita real income, saving rates, population growth rates, and the growth and size of government. The 36 years spanned in these data would meet the adjustment time estimates of 14 to 35 years for equilibrium growth to be reached (but not the estimates of 42 to 106 years in a model without endogenous learning).

The policy implications are straightforward. Public policies that raise the capital-labor ratio magnify the effects on the growth rate of per capita income, owing to induced learning-by-doing associated with a rising capital stock. Policies that enhance the learning process also accelerate the speed of adjustment toward the balanced growth path. Examples of such policies include measures to raise saving and investment, to permit the steady expansion of the tradable sector, and to accelerate the growth of real expenditures on education and health. On the other hand, there are clear limits to the size of government in relation to GDP, because of the increasingly heavy costs of burgeoning deficits.

APPENDIX

Data Used in the Study

The data, except for foreign trade flows, are drawn from Orsmond (1990), which are based on the IMF's *Government Financial Statistics* and *International Financial Statistics*. Foreign trade flows are taken from the *World Economic Outlook* data base. The sample consists of observations averaged over the period 1975–86 for 36 developing countries.

PYG = Real per capita GDP growth rate (annual average);
KY = Gross investment divided by nominal GDP (annual average);
XC = Change in ratio of the sum of nominal exports and imports to nominal GDP between 1975 and 1986;
EG = Growth rate of government expenditures on education and health (annual average), deflated by GDP deflator for budget year;
SG = Growth rate of government expenditures on social security, housing, and recreation (annual average), deflated by GDP deflator for budget year;
DY = Nominal fiscal deficits divided by nominal GDP (annual average);
PG = Population growth rate (annual average);
GDP75 = Per capita income level in 1975 U.S. dollars;
DUM(*i*) = Dummy variable that assumes the value of 1 for region *i* and the value of 0 for other regions; *i* = Africa, Asia, Middle East, Western Hemisphere.

The countries in the sample are

Botswana
Burkina Faso
Cameroon
Chile
Costa Rica
Dominican Republic
Egypt
El Salvador
Ethiopia
Fiji
Guatemala
Indonesia
Islamic Republic of Iran
Kenya
Korea
Liberia
Mauritius
Mexico
Morocco
Myanmar
Nepal
Pakistan
Panama
Paraguay
Singapore
Sri Lanka
Tanzania
Thailand
Togo
Tunisia
Turkey
Uruguay
Venezuela
Yemen Arab Republic
Zambia
Zimbabwe

Country	*PYG*	*KY*	*XC*	*EG*	*SG*	*DY*	*PG*	*GDP75*
Botswana	7.4	29.8	3.5	17.5	38.7	−4.2	4.7	350.0
Korea	7.1	28.8	10.6	10.2	12.9	1.7	1.4	580.0
Singapore	5.7	40.3	35.6	12.8	12.8	−1.7	1.2	2,540.0
Yemen Arab Republic	3.7	27.9	−9.4	32.1	2.0	10.8	2.8	140.0
Pakistan	3.4	17.0	1.1	10.1	32.2	7.5	3.1	140.0
Cameroon	3.1	22.4	−7.6	0.0	13.3	0.5	3.0	310.0
Mali	3.9	24.1	11.1	9.0	6.1	4.1	2.1	360.0
Indonesia	4.0	24.4	−4.0	9.3	2.0	2.2	2.0	210.0
Paraguay	2.6	23.7	11.6	1.2	9.2	0.3	3.2	550.0
Myanmar	3.2	15.9	−0.3	7.0	10.0	−0.1	2.4	150.0
Sri Lanka	3.6	23.3	13.8	6.7	0.8	10.5	1.6	220.0
Tunisia	2.5	29.4	2.5	5.0	9.5	4.7	2.6	710.0

Country	PYG	KY	XC	EG	SG	DY	PG	GDP75
Kenya	0.8	20.9	−15.8	5.8	5.1	5.7	4.2	230.0
Panama	1.9	23.4	−26.7	4.1	9.1	7.6	2.6	1,030.0
Mauritius	3.1	24.4	−3.1	7.7	−0.5	8.5	1.3	300.0
Burkina Faso	2.7	22.9	4.6	6.7	18.7	0.3	1.6	100.0
Egypt	1.8	25.8	25.2	9.4	0.3	13.1	2.5	310.0
Turkey	2.1	20.2	15.4	0.7	2.7	4.3	2.0	830.0
Chile	2.2	14.7	8.0	3.4	7.3	−0.3	1.7	860.0
Morocco	1.3	25.2	−4.3	5.3	6.5	10.8	2.5	500.0
Nepal	0.9	17.7	10.7	11.3	14.3	4.6	2.8	110.0
Mexico	0.9	21.1	5.2	4.0	0.0	6.6	2.6	1,360.0
Ethiopia	−1.2	10.3	8.0	6.3	10.8	6.0	4.6	90.0
Dominican Republic	0.4	21.2	−8.9	2.0	1.6	2.1	2.8	670.0
Costa Rica	0.0	21.8	−38.9	1.5	9.5	3.3	3.1	950.0
Zimbabwe	−0.5	17.8	−29.8	10.9	5.0	7.9	2.9	570.0
Fiji	0.2	20.6	−14.3	4.8	13.0	3.8	2.0	1,030.0
Guatemala	−0.4	15.3	−16.5	13.9	−1.2	2.7	2.5	570.0
Togo	−0.9	29.1	0.4	5.9	6.6	11.3	2.9	260.0
Tanzania	−1.8	18.5	−16.0	−5.6	3.8	8.8	3.6	160.0
Venezuela	−1.4	26.4	−17.4	2.7	2.6	0.7	3.1	2,380.0
Uruguay	0.8	13.3	5.8	−1.8	3.8	2.6	0.6	1,370.0
Islamic Republic of Iran	−2.6	22.7	−58.9	3.5	10.1	5.8	3.8	1,449.7
Zambia	−3.2	18.6	9.0	−3.4	−0.5	13.7	3.5	550.0
Liberia	−3.3	22.4	−29.2	6.1	13.7	7.8	3.3	410.0
El Salvador	−2.1	15.9	−21.1	−3.4	0.3	2.8	1.9	440.0

REFERENCES

Arrow, Kenneth, "The Economic Implications of Learning by Doing," *Review of Economic Studies*, Vol. 29 (1962), pp. 155–73.

Bardhan, Pranab, and Sydney Lewis, "Models of Growth with Imported Inputs," *Economica*, Vol. 37 (1970), pp. 373–85.

Becker, Gary, Kevin Murphy, and Robert Tamura, "Human Capital, Fertility, and Economic Growth," *Journal of Political Economy*, Vol. 98 (1990), pp. S12–37.

Chen, Edward, *Hyper-Growth in Asian Economies: A Comparative Study of Hong Kong, Japan, Korea, Singapore, and Taiwan* (London and New York: Macmillan, 1979).

Conlisk, John, "A Modified Neo-Classical Growth Model with Endogenous Technical Change," *Southern Economic Journal*, Vol. 34 (1967), pp. 199–208.

Domar, Evsey, "Capital Expansion, Rate of Growth, and Employment," *Econometrica*, Vol. 14 (1946), pp. 137–47.

Edwards, Sebastian, "Trade Orientation, Distortions and Growth in Developing Countries," *Journal of Development Economics*, Vol. 39 (1992), pp. 31–57.

Feder, Gershon, "On Exports and Economic Growth," *Journal of Development Economics*, Vol. 12 (1983), pp. 59–73.

Grossman, Gene, and Elhanan Helpman, "Comparative Advantage and Long-Run Growth," *American Economic Review*, Vol. 80 (1990), pp. 796–815.

Harrod, Roy, "An Essay in Dynamic Theory," *Economic Journal*, Vol. 49 (1939), pp. 14–33.

Inada, Ken-Ichi, "On a Two-Sector Model of Economic Growth: Comments and Generalization," *Review of Economic Studies*, Vol. 30 (1963), pp. 119–27.

Keesing, Donald, "Outward-Looking Policies and Economic Development," *Economic Journal*, Vol. 77 (1967), pp. 303–20.

Khan, Mohsin, and Delano Villanueva, "Macroeconomic Policies and Long-Term Growth: A Conceptual and Empirical Review," IMF Working Paper 91/28 (Washington: International Monetary Fund, 1991).

Khang, Chulsoon, "Export-Led Economic Growth: The Case of Technology Transfer," *Economic Studies Quarterly*, Vol. 38 (1987), pp. 131–47.

Knight, Malcolm, Norman Loayza, and Delano Villanueva, "Testing the Neoclassical Theory of Economic Growth: A Panel Data Approach," *Staff Papers*, International Monetary Fund, Vol. 40 (September 1993), pp. 512–41.

Lucas, Robert, "On the Mechanics of Economic Development," *Journal of Monetary Economics*, Vol. 22 (1988), pp. 3–42.

Phelps, Edmund, *Golden Rules of Economic Growth* (New York: W.W. Norton, 1966).

Orsmond, David, "The Size of Government and Economic Growth: A Methodological Review" (unpublished; Washington: International Monetary Fund, 1990).

Ramanathan, Ramu, *Introduction to the Theory of Economic Growth* (Berlin: Springer-Verlag, 1982).

Rivera-Batiz, Luis, and Paul Romer, "International Trade with Endogenous Technological Change," NBER Working Paper No. 3594 (Washington: National Bureau of Economic Research, 1991).

Romer, Paul, "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, Vol. 94 (1986), pp. 1002–37.

Sato, Ryuzo, "Fiscal Policy in a Neo-Classical Growth Model—An Analysis of the Time Required for Equilibrium Adjustment," *Review of Economic Studies*, Vol. 30 (1963), pp. 16–23.

Solow, Robert, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, Vol. 70 (1956), pp. 65–94.

Summers, Robert, and Alan Heston, "A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries, 1950–85," *Review of Income and Wealth*, Vol. 34 (1988), pp. 1–25.

Swan, Trevor, "Economic Growth and Capital Accumulation," *Economic Record*, Vol. 32 (1956), pp. 334–62.

Thirlwall, A., "The Balance of Payments Constraint as an Explanation of International Growth Rate Differences," *Banca Nazionale del Lavoro Quarterly Review*, Vol. 128 (1979), pp. 45–53.

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

Collection Lags and the Optimal Inflation Tax

A Reconsideration

ALEX MOURMOURAS and JOSÉ A. TIJERINA*

The observation that inflation reduces real revenues when there are lags in tax collection has long been a strong argument against seigniorage. However, with the exception of Dixit, who used a general equilibrium model to reject this argument, the optimal taxation literature has not analyzed how collection lags affect desired tax structures. This paper reexamines the issue using an overlapping generations version of Dixit's model. It is shown that depending on the size of the expenditure ratio and the specification of the collection cost function, lags may increase, leave unchanged, or reduce the desired rate of inflation. [JEL E51, E62, H21]

A BASIC ISSUE in both public finance and monetary economics is the desirability of using inflationary finance to generate government revenue. In the monetary literature, the orthodox position is associated with Friedman's (1969) optimum quantity of money rule, which argues that the nominal rate of interest should be zero.[1] The public finance

* Alex Mourmouras is an Economist in the Fiscal Affairs Department. He holds a Ph.D. from the University of Minnesota. José A. Tijerina is a Professor of Economics at the Autonomous University of Nuevo Leon in Monterrey, Mexico. He holds a Ph.D. from the University of Cincinnati.
The authors would like to acknowledge helpful comments from Sheetal Chand, Avinash Dixit, Steve Russell, Vito Tanzi, and Carlos Végh. All errors are our own responsibility.

[1] Friedman argues that to allocate resources efficiently in a monetary economy, the social marginal benefit of money must be brought in line with its social marginal cost. Since fiat money is (almost) costless to produce, the nominal rate of interest should be set at zero—for example, by contracting the money supply at a rate equal to the real rate of interest. See Woodford (1990) for an extensive discussion of the optimum quantity of money rule.

literature has been dominated by Phelps (1973), who uses a model with money specified in the utility function to show that seigniorage can be part of a second-best tax system. Phelps' starting point is the observation that to implement Friedman's rule, lump-sum transfers must be feasible. If they are not, taxation of all commodities—including consumption and liquidity—may be required to raise government revenue.[2] Later contributions to the public finance literature criticize Phelps' treatment of liquidity as a separate commodity, focusing on fiscal inefficiencies to rationalize the use of seigniorage.[3] With the exception of Kimbrough (1986), who salvages Friedman's rule in a second-best setting, the inefficiencies assumed in this literature (such as positive foreign nominal rates of interest, collection costs, or a large underground economy) introduce distortions of their own for ordinary taxes. Resorting to the monetary financing of deficits helps reduce these inefficiencies and allows seigniorage to coexist with consumption (or income) taxes in optimal tax menus.

With the exception of Dixit (1991), however, the optimal inflation literature has not considered the implications of collection lags for the argument concerning seigniorage in second-best settings. As has long been emphasized by Tanzi (1978), inflationary finance may further weaken the public finances if high rates of inflation combine with long collection lags to erode the real value of ordinary fiscal revenues.[4] Tanzi's argument has been challenged recently by Dixit, who provides a welfare

[2] Kimbrough (1986), on the other hand, argues that Friedman's rule may be optimal even in second-best environments. His reasoning is based on the view (formalized in the shopping-time model) that fiat money is not a final good but rather an intermediate input in the transactions technology. A theorem on second-best taxation from Diamond and Mirrlees (1971) then applies. This theorem guarantees that if the production function exhibits constant returns to scale and all final goods are taxable, then intermediate inputs ought not to be taxed. For a discussion of the applicability of Friedman's rule in the shopping-time monetary model, see Végh and Guidotti (1993).

[3] See Frenkel (1987), Mourmouras (1991), and Végh (1989).

[4] Admittedly, the applicability of current abstract theories of inflation (especially of the steady-state variety) to macroeconomic stabilization problems is limited. For instance, the collection period (which is exogenous in our analysis as well as in Dixit's) is in reality an endogenous variable that depends, among other things, on the prevailing rate of inflation. In periods of high inflation, in particular, rational governments attempt to protect the real value of revenues by shortening the collection period. It is interesting to note that theorists allow income tax rates to be changed costlessly while exogenously holding fixed the collection interval, even though in practice the former is a "stickier" policy instrument than the latter. In particular, the frequency of tax payments can be changed simply by instructing the tax administration to issue new implementing orders, while a change in the tax rate must (in democratic systems) go through long legislative processes. See Tanzi (1992) on this point and on the other limitations of current normative tax theories.

analysis of inflation that incorporates collection costs and lags into a version of Végh's (1989) model. Dixit observes that rational governments will, in a general equilibrium environment, react to the presence of lags by adjusting all taxes, not just the rate of inflation. Since this changes prices and the real cost of collections, there is a richer menu of possibilities to consider than under a partial equilibrium framework. Dixit provides two examples that reverse the traditional argument. In the first, the length of the collection lags is irrelevant for the optimal choice of inflation, as if full interest were charged to compensate for the delay in tax payments. More interesting is the second case in which the presence of lags raises the excess burden of income taxes, thereby warranting greater reliance on seigniorage than when no lags exist.

While relying on the public finance approach to inflation, this paper reconsiders the optimal mix of inflation and costly income taxes for alternative specifications of the monetary model and the collection cost technology. In particular, an optimal-tax analysis is performed in the context of Samuelson's (1958) consumption loans model, in which, unlike the shopping-time model, the major distortion caused by money-financed deficits is that on intertemporal consumption allocations. Several interesting results emerge. First, regardless of the length of the collection lag, optimal inflation is proportional to the marginal cost of income tax collections, implying that price stability ought to be pursued whenever these collections are costless at the margin. Second, and in accordance with the Ramsey "inverse elasticity rule," the optimal rate of inflation is inversely related to both the marginal propensity to consume and the interest elasticity of real money balances. Third, the desired rate of inflation in a lag-ridden economy is lower than the desired rate in a lag-free economy if g, the share of government spending in GDP, is below a certain threshold. Thus, Dixit's unconditional rejection of the traditional presumption—that collection lags ought to reduce the optimal rate of inflation—is unwarranted on theoretical grounds. As a practical matter, the threshold value of g, which is crucial for the comparison, turns out to be a function of the marginal cost of collection (Table 1 and Figure 1).

These results owe to a combination of factors. First, in the present model the efficiency trade-off is between money-financed deficits, which lower real interest rates and distort intertemporal choice, and income taxes, which require real resources for collection. This leads naturally to Ramsey formulae that incorporate marginal collection costs as well as the interest and income elasticities of currency demand.[5] By contrast, in the

[5] This distortion, which is not always made explicit in welfare analyses of the costs of inflation, seems to be important in practice—particularly in developing

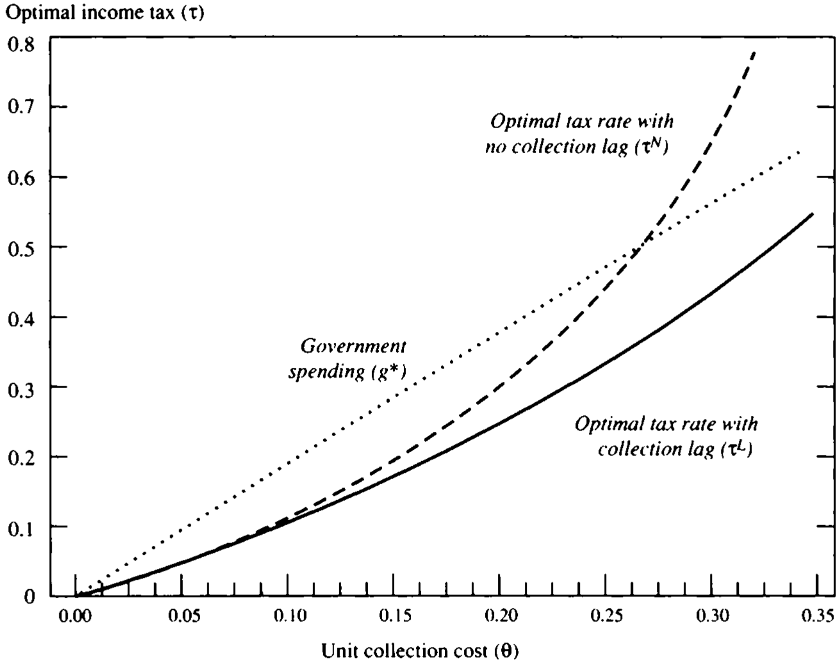
Table 1. *Critical Values of Government Spending (g) and Optimal Income Tax and Inflation (τ, π) as Functions of the Unit Cost of Collections (θ)*

Unit collection cost (θ)	Inflation		Income tax		Government spending (g^*)
	π^N	π^L	τ^L	τ^N	
0.00	0.0000	0.0000	0.0000	0.0000	0.0000
0.01	0.0204	0.0204	0.0101	0.0101	0.0199
0.02	0.0417	0.0417	0.0204	0.0204	0.0396
0.03	0.0638	0.0638	0.0309	0.0310	0.0591
0.04	0.0870	0.0870	0.0417	0.0418	0.0785
0.05	0.1111	0.1111	0.0526	0.0529	0.0976
0.06	0.1364	0.1364	0.0638	0.0644	0.1166
0.07	0.1628	0.1628	0.0753	0.0761	0.1355
0.08	0.1905	0.1905	0.0870	0.0883	0.1542
0.09	0.2195	0.2195	0.0989	0.1009	0.1727
0.10	0.2500	0.2500	0.1111	0.1139	0.1911
0.11	0.2821	0.2821	0.1236	0.1274	0.2094
0.12	0.3158	0.3158	0.1364	0.1415	0.2276
0.13	0.3514	0.3514	0.1494	0.1563	0.2456
0.14	0.3889	0.3889	0.1628	0.1717	0.2636
0.15	0.4286	0.4286	0.1765	0.1878	0.2815
0.16	0.4706	0.4706	0.1905	0.2048	0.2993
0.17	0.5152	0.5152	0.2048	0.2228	0.3170
0.18	0.5625	0.5625	0.2195	0.2417	0.3347
0.19	0.6129	0.6129	0.2346	0.2619	0.3524
0.20	0.6667	0.6667	0.2500	0.2833	0.3700
0.21	0.7241	0.7241	0.2658	0.3062	0.3876
0.22	0.7857	0.7857	0.2821	0.3308	0.4053
0.23	0.8519	0.8519	0.2987	0.3572	0.4229
0.24	0.9231	0.9231	0.3158	0.3857	0.4406
0.25	1.0000	1.0000	0.3333	0.4167	0.4583
0.26	1.0833	1.0833	0.3514	0.4503	0.4762
0.27	1.1739	1.1739	0.3699	0.4871	0.4941
0.28	1.2727	1.2727	0.3889	0.5275	0.5121
0.29	1.3810	1.3810	0.4085	0.5720	0.5303
0.30	1.5000	1.5000	0.4286	0.6214	0.5486
0.31	1.6316	1.6316	0.4493	0.6765	0.5671
0.32	1.7778	1.7778	0.4706	0.7383	0.5858
0.33	1.9412	1.9412	0.4925	0.8081	0.6047
0.34	2.1250	2.1250	0.5152	0.8873	0.6240
0.35	2.3333	2.3333	0.5385	0.9782	0.6435

Notes: (τ^N, π^N) = desired income tax and inflation pair with no collection lag; (τ^L, π^L) = desired income tax and inflation pair with collection lag; π is an absolute figure over a 20-year period; θ, τ, and g^* are absolute figures.

countries. In these countries, the ability of savers (especially small ones) to index asset returns through financial markets is hindered by prohibitive transaction costs and the primitive state of development of these markets. See Wallace (1980) on this point.

Figure 1.
Critical Value of Government Spending and Optimal Income Tax Rates



Notes: g* = critical value of the ratio of government spending to GDP as a function of the unit cost of income tax collections; τ^N = optimal income tax rate in absence of collection lags; and τ^L = optimal income tax rate in presence of collection lags.

shopping-time model used by Végh (1989) and Dixit (1991), a version of Irving Fisher's theory of interest is maintained, according to which the inflationary process does not affect real interest rates, the whole profile of which is taken to be exogenous (Dixit, p. 645). Second, although the introduction of payment lags reduces tax collections in real terms at a given positive rate of inflation, it also raises desired real currency balances, leading to complicated changes in tax bases and the optimal tax menu. Finally, unlike Dixit who maintains (p. 648) that "the nominal collection cost technology exactly keeps pace with inflation," this paper allows real collection costs to vary with inflation. This feature proves to be important in optimal tax calculations as well.

This paper, then, describes first a model of collection lags and inflation, as well as presents the public finance analysis and derives the Ramsey formula for an economy in which there are no (significant) lags in income

tax collections. Then, it calculates the optimal rate of inflation for two parametric examples corresponding, respectively, to Dixit's benchmark specification of constant marginal collection cost and the more realistic increasing marginal cost case. Next, the impact of a one-period collection lag on the optimal tax structure is analyzed, and the optimal rates of inflation are compared for the lag and lag-free environments. Finally, the paper examines the sensitivity of these results to changes in the specification of the collection cost function and proves a neutrality proposition for a plausible alternative functional form.

I. The Model

Consider the following simple version of the consumption loans model. The economy comprises an infinite sequence of generations who live two periods and overlap with one another. Each generation is the same size, which for simplicity is normalized to one. Perfect foresight is assumed. At each date $t = 1, 2, \ldots$, the representative young agent is endowed with $\bar{n}$ units of labor and a technology $f(n)$, which allows him or her to produce a single perishable consumption good. The production function $f(\cdot)$ satisfies $f(0) = 0, f' \geq 0, f'' \leq 0$. If real gross income is $w = f(\bar{n})$ and $\tau(t)$ is the proportional rate of taxation applied to such income, after-tax real income is $A(t) \equiv [1 - \tau(t)]w(t)$. A fiat currency issued by the government is the only asset in the economy. Agents are retired in the second period of life and must rely on accumulated currency balances to purchase the consumption good. If $p(t)$ denotes the price level at t, the rate of inflation between t and $t + 1$ is $\pi(t) \equiv [p(t+1) - p(t)]/p(t)$, while the real gross rate of return on currency balances is $R(t) \equiv 1/[1 + \pi(t)]$. The end-of-period economywide stock of currency is denoted $M(t)$. The entire pretransfer stock of currency at $t = 1$ is owned by the old agent belonging to generation 0.

Let $c_t(k)$ denote consumption in period k by an agent born at t, and let $m(t)$ denote his or her nominal currency holdings at the end of t. Given the current and expected future price levels, $p(t)$ and $p(t + 1)$ respectively, and $\tau(t)$, the competitive choice problem is to select a nonnegative vector $[c_t(t), c_t(t+1), m(t)]$ that maximizes lifetime utility

$$u[c_t(t), c_t(t+1)] \tag{1}$$

subject to

$$c_t(t) + m(t)/p(t) \leq [1 - \tau(t)]w \quad \text{and}$$
$$c_t(t+1) \leq m(t)/p(t+1). \tag{2}$$

In an interior solution (to be assumed throughout), the marginal rate of substitution between first- and second-period consumption, u_1/u_2, must be equal to the ratio of relative prices $p(t)/p(t+1) \equiv R(t)$. The condition $u_1 = Ru_2$ together with equation set (2) at equality may be used to derive consumer demand schedules $c_t(t) = c_1[A(t), R(t)]$ and $c_t(t+1) = c_2[A(t), R(t)]$, and a demand schedule for real currency balances $m(t)/p(t) = s[A(t), R(t)]$. In what follows, $\epsilon(s,R) \equiv \partial \ln(s)/\partial \ln(R)$ and $\epsilon(s,A) \equiv \partial \ln(s)/\partial \ln(A)$ denote, respectively, the interest and income elasticities of real currency demand, and $s_A \equiv \partial s/\partial A$ denotes the marginal propensity to save out of current income.

Government spending on public goods and services excluding collection costs is financed via a flat-rate income tax and seigniorage. Public goods enter individual utility functions in a separable manner that does not affect preference orderings of private goods. The sequence of government expenditures on public goods, $\{G(t), t = 1, 2, \dots\}$ may therefore be treated as exogenous. Income taxes require real resources for collection. Specifically, if τw is gross tax revenue in real terms, then net revenue is $\tau w - \phi(\tau w)$ where $\phi(\cdot)$ is a monotonic increasing function describing the resource cost of income tax collections faced by the government. The function ϕ satisfies $\phi(0) = 0, \phi' \geq 0, \phi'' \geq 0$. The government cash-flow constraint for period $t = 1, 2, \dots$ may be written:

$$G(t) = \tau(t)w(t) - \phi[\tau(t)w(t)] + [M(t) - M(t-1)]/p(t). \tag{3}$$

Given $\bar{n}$ and $M(0)$, a *perfect-foresight competitive equilibrium* is a set of sequences $\{c_t(t), c_{t-1}(t), s(t), M(t), p(t), \pi(t), R(t), G(t), \tau(t)\}$ that satisfy the conditions of individual optimization and are consistent with market clearing and the sequence of government budget constraints for all $t = 1, 2, \dots$. A *stationary equilibrium* is a set of scalars $(c_1, c_2, \pi, R, G, \tau)$ together with a value $c_0(1) = M(0)/p(1)$ and geometrically increasing levels of $p(t), M(t)$ for all $t = 1, 2, \dots$ satisfying the following:

$$u_2(A - s, Rs) = Ru_1(A - s, Rs) \tag{3a}$$

$$G = \tau w - \phi(\tau w) + (1 - R)s(A, R) \tag{3b}$$

$$G(1) = \tau(1)w(1) - \phi[\tau(1)w(1)] + [M(1) - M(0)]/p(1) \tag{3c}$$

$$M(1)/p(1) = s(A, R) \tag{3d}$$

$$M(t+1)/M(t) = p(t+1)/p(t) = 1/R. \tag{3e}$$

Note that the term $1 - R \equiv \pi/(1 + \pi)$ in equation (3b) is the effective rate of taxation on real currency balances (the "inflation tax rate").

II. Optimal Taxation

In this section, the model above is used to illustrate conditions under which inflationary finance can be part of a second-best tax menu. In particular, a Ramsey formula is derived that makes explicit the dependence of the desired rate of inflation on the specification of the collection cost function and such variables as the marginal propensity to consume and the income and interest elasticities of money demand. Basically, the assumption that income taxes require real resources for collection introduces the distortion necessary to allow seigniorage to be a part of the optimal government finance strategy. Thus, while higher rates of inflation reduce the real rate of interest and distort intertemporal choice, higher rates of income taxation entail higher collection costs. In an optimal position, marginal excess burdens of the two taxes are equalized. In some specifications, this involves trading off some distortion in the intertemporal margin to economize on the resource costs of income taxes.

Let the social welfare criterion be the steady-state utility of each generation $t = 1, 2, \ldots$, denoted $V(A, R) \equiv u(A - s, Rs)$. The authorities set the pair (R, τ) to maximize V subject to equation (3b) and the functional form of the real money demand function $s(A, R)$ dictated by private optimization. Letting μ denote the Lagrange multiplier associated with equation (3b) and assuming an interior solution, the first-order necessary conditions of this problem are equation (3b) and

$$u_1 = \mu[1 - \phi' - (1 - R)s_A], \tag{4a}$$

$$u_2 = \mu[1 - (1 - R)s_R/s]. \tag{4b}$$

Dividing equation (4a) by equation (4b) and using the identity $u_1 = Ru_2$ yields

$$R = \frac{1 - \phi' - (1 - R)s_A}{1 - (1 - R)s_R/s}. \tag{5}$$

Equation (5) may be solved for the *optimal inflation tax* $1 - R$ as a function of the marginal cost of collection ϕ' and the income and interest elasticities of real money demand. Cross-multiplying and rearranging terms lead to the following simple formula:

$$1 - R = \frac{\phi'}{1 - s_A + \epsilon(s, R)}. \tag{6}$$

The Ramsey formula (6) demonstrates, first, that *the desired inflation tax is proportional to the marginal cost of collection* ϕ'. Thus, regardless of the values of the income and interest elasticities of money demand, authorities should strive for price stability ($R = 1$ or $\pi = 0$) if tax collections are costless at the margin. The reason for this result is that in the present model agents can neither engage in untaxed home production nor value leisure directly. Since labor endowments are inelastically supplied in the taxed activity, costless flat-rate income taxes are equivalent to lump-sum taxes.[6] If costless income taxes are used to raise 100 percent of revenue it is possible for the economy to achieve its first-best allocation (Samuelson's golden rule). Given the assumption of zero population growth, this objective amounts to a zero net real interest rate target or, equivalently, the pursuit of a stable price level.

Equation (6) also suggests that the optimal inflation tax is *inversely* related to the marginal propensity to consume $1 - s_A$ and the interest elasticity of real money demand $\epsilon(s, R)$. A one-unit increase in the marginal propensity to consume lowers by one unit the real value of desired currency balances carried forward to the following period. As such, it reduces the base of the inflation tax and raises the excess burden of a given rate of inflation. Finally, and for entirely analogous reasons, the optimal inflation rate is negatively related to the interest elasticity of real currency demand.

III. Examples of Optimal Tax Menus

To gain additional insight into the nature of optimal taxes, two special cases are considered below that lead to closed-form solutions of the optimal (τ, π) pair. The first case corresponds to Végh's and Dixit's benchmark specifications of logarithmic utility and constant unit collection costs; the second corresponds to the more realistic case of increasing marginal collection costs.

Suppose first that consumer preferences are given by $u(c_1, c_2) = \log(c_1) + \beta \log(c_2)$, where $\beta > 0$ is the subjective time discount factor. Consumer optimization leads to a constant saving rate $[s = \beta/(1 + \beta)]$

[6]The presence of untaxed activities would alter this result. If workers value leisure or have employment opportunities in the underground economy, proportional income taxes would distort the margin of choice by consumers—the labor-leisure choice in the first case and the regular-underground employment in the second. In either case, efficient taxation would entail positive income and inflation taxes even in the absence of collection costs. Mourmouras (1991) provides an explicit analysis.

and interest-inelastic real currency demand $[\epsilon(s,R) = 0]$. The Ramsey formula (6) then becomes $1 - R = (1 + \beta)\phi'$. Assuming, in addition, that the marginal cost of collection is constant, say $\phi' = \theta_1 > 0$, then optimal inflation is (a) constant and independent of the level of government spending; and (b) inversely related to the subjective rate of time preference $1/\beta - 1$.

Given the constant desired value of $1 - R$, the optimal income tax rate can be computed from equation (3b). Letting $g \equiv G/w$ denote the share of government spending in national income, the optimal value of τ is

$$\tau = \frac{g - \beta\theta_1}{1 - (1 + \beta)\theta_1}. \tag{7}$$

According to equation (7), income taxes are positive provided that the unit cost of collection is not too large (in the sense of satisfying $g > \beta\theta_1$ and $1 > [1 + \beta]\theta_1$). As expected, the optimal income tax rate in this range is inversely related to the unit cost of collection. For example, if there is no discounting of the future ($\beta = 1$) and the unit cost of collection is 10 percent ($\theta_1 = 0.10$), then desired inflation is 25 percent per period ($\pi = 0.25$), regardless of the size of the government budget.[7] If, in addition, government spending net of collection costs is 40 percent of GDP ($g = 0.40$), then the income tax rate is $\tau = 0.375$. In this environment, where tax collections are 10 percent of gross budgetary revenue, or 3.75 percent of GDP, gross budget revenues are 37.5 percent of GDP, resulting in an optimal money-financed deficit of 6.25 percent of GDP. As a second example, if the unit cost of collection is only 5 percent ($\theta_1 = 0.05$) and $g = 0.40$, the optimal policy is calculated to be $(\pi, \tau) = (0.111, 0.389)$. These examples demonstrate that doubling the efficiency of the tax administration (that is, reducing unit collection costs by half) could reduce the desired rate of inflation by more than half. Figures 2 and 3 relate optimal policy to the fundamentals of the economy by drawing optimal (τ, π) pairs against g and θ_1, respectively. Table 2 shows optimal (τ, π) pairs for selected values of θ_1.

The conclusion that inflation is part of an *interior* optimal tax package is in sharp contrast with the results obtained by Frenkel (1987). This author studied a log-linear cash-in-advance model of labor-leisure choice and showed that if the marginal cost of income tax collections is constant, then the optimal policy involves financing government expenditures via inflation taxes *in their entirety*. This result stems directly from Frenkel's specification of money demand and his assumption that inflation is cost-

[7] Assuming the length of the period to be approximately 20 years, this translates to an annual rate of inflation of 1.1 percent.

Figure 2.
Optimal Inflation and Optimal Income Tax with Fixed Collection Cost (θ = 0.30)

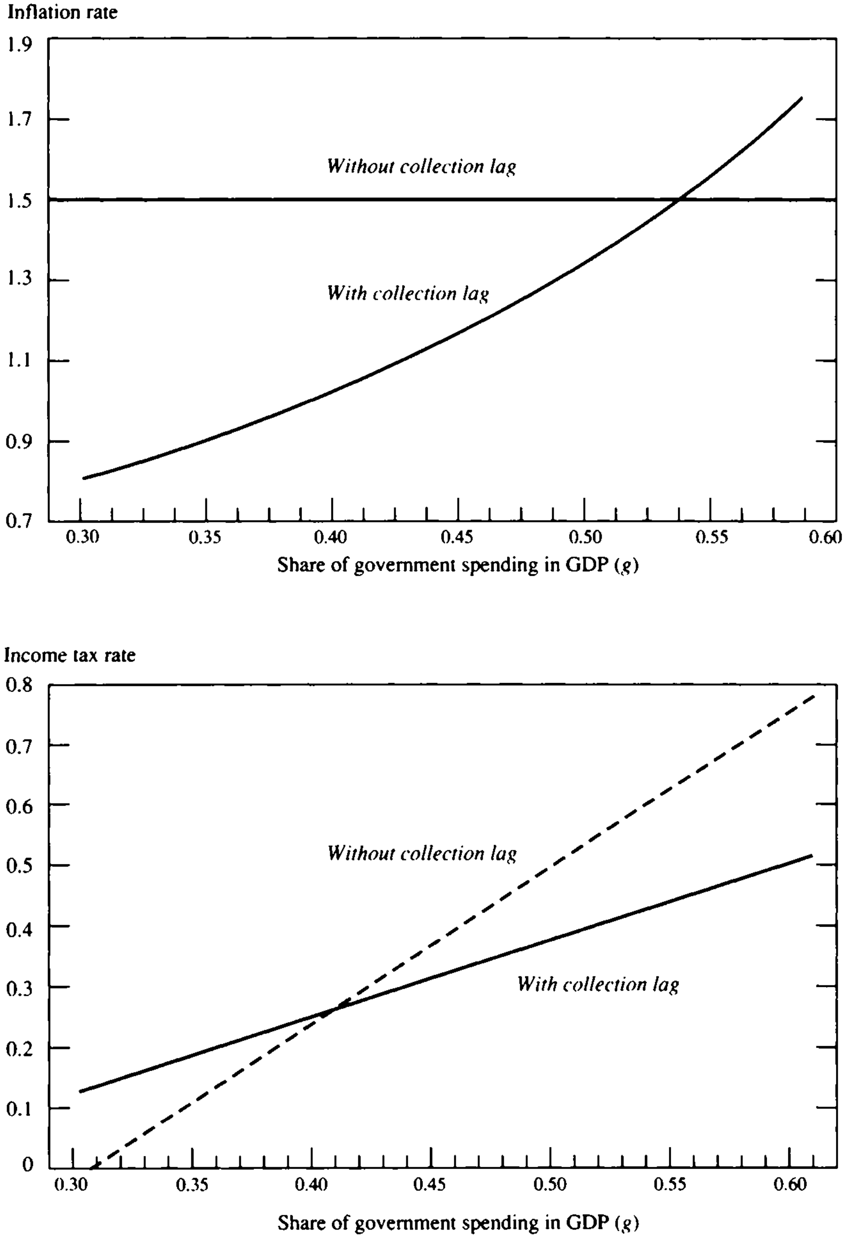
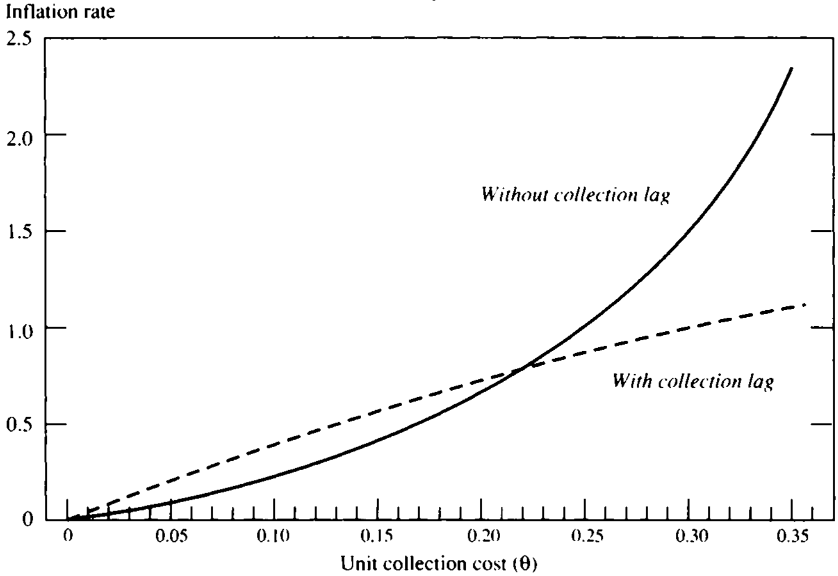


Figure 3.
Optimal Inflation and Optimal Income Tax with Fixed Government Spending (g = 0.40)



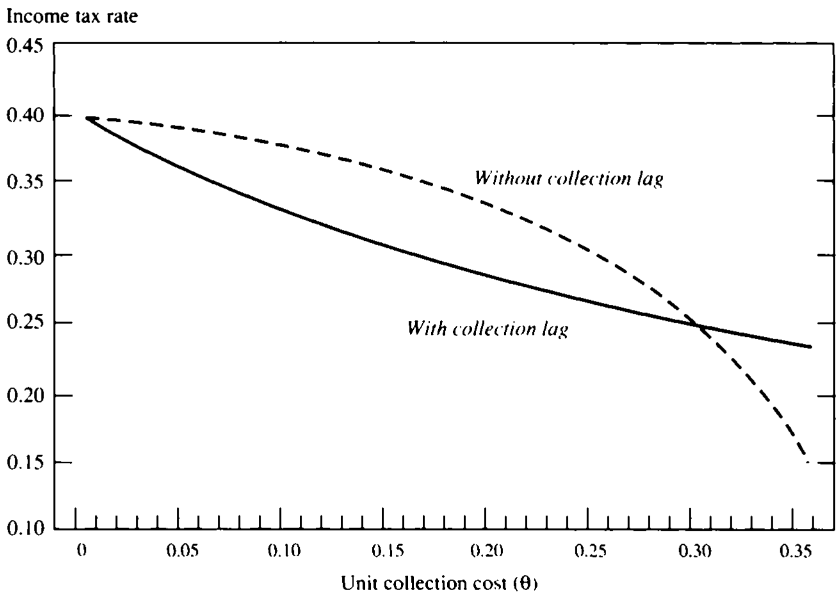


Table 2. *Size of Government Spending and Optimal Taxation: Selected Values of Desired (τ, π) Pairs as Functions of* g

Government spending (g)	Unit collection cost (θ)	Inflation π^N	Inflation π^L	Income tax τ^N	Income tax τ^L
30	30	4.69	2.93	0.00	13.16
31	30	4.69	2.99	2.50	14.34
32	30	4.69	3.04	5.00	15.52
33	30	4.69	3.10	7.50	16.70
34	30	4.69	3.16	10.00	17.88
35	30	4.69	3.21	12.50	19.07
36	30	4.69	3.27	15.00	20.25
37	30	4.69	3.34	17.50	21.43
38	30	4.69	3.40	20.00	22.62
39	30	4.69	3.46	22.50	23.81
40	30	4.69	3.53	25.00	25.00
41	30	4.69	3.59	27.50	26.19
42	30	4.69	3.66	30.00	27.38
43	30	4.69	3.73	32.50	28.58
44	30	4.69	3.80	35.00	29.77
45	30	4.69	3.87	37.50	30.97
46	30	4.69	3.95	40.00	32.17
47	30	4.69	4.02	42.50	33.37
48	30	4.69	4.10	45.00	34.57
49	30	4.69	4.18	47.50	35.78
50	30	4.69	4.26	50.00	36.98
51	30	4.69	4.35	52.50	38.19
52	30	4.69	4.43	55.00	39.40
53	30	4.69	4.52	57.50	40.61
54	30	4.69	4.61	60.00	41.82
55	30	4.69	4.70	62.50	43.03
56	30	4.69	4.80	65.00	44.25
57	30	4.69	4.89	67.50	45.46
58	30	4.69	4.99	70.00	46.68
59	30	4.69	5.10	72.50	47.90
60	30	4.69	5.20	75.00	49.13

Notes: g is expressed as a percent of GDP; π^N and π^L are expressed in annual rates in percent; τ^N, τ^L, and θ are expressed in percent.

less while income taxes require resource costs. In particular, Frenkel assumes demand for real currency balances to be proportional to income. Since the bases for the income and inflation taxes are then proportional, the costless inflation tax will always be preferred over the costly income tax. By contrast, in the overlapping generations model, the two tax bases are separate but related. On the one hand, income taxes are levied on

current income at the source; on the other hand, seigniorage subjects to tax that part of current income which is *not* consumed in the present period, creating a distortion of intertemporal choice. This distinction creates a nontrivial trade-off between the two types of taxes that is optimally exploited by the policymaker as reflected in equation (6).

In this section, the more realistic case of increasing marginal collection costs is taken up. Following Végh (1989) and Dixit (1991), suppose that the function ϕ is quadratic:

$$\phi(\tau w) = \theta_1(\tau w) + (\theta_2/2)(\tau w)^2, \qquad \theta_1 > 0 \quad \text{and} \quad \theta_2 > 0. \tag{8}$$

With the collection cost function as specified in equation (8) and logarithmic utility, it turns out that the optimal income and inflation tax rates are linearly related. Substituting $\phi' = \theta_1 + \theta_2(\tau w)$, $\epsilon(s, R) = 0$, and $1 - s_A = 1/(1 + \beta)$ into equation (6) yields the exact relation between τ and π:

$$1 - R \equiv \pi/(1 + \pi) = (1 + \beta)[\theta_1 + \theta_2 \tau w]. \tag{9}$$

To compute the reduced forms of τ and π, the right-hand side of equation (9) may be substituted into the government budget constraint (3b). After some algebra, the optimal value of τ is seen to be a solution to the following equation:

$$(\tfrac{1}{2} + \beta)\theta_2 w\tau^2 - [1 - (1 + \beta)\theta_1 + \beta\theta_2 w]\tau + g - \beta\theta_1 = 0. \tag{10}$$

The roots of equation (10) are real if $\Delta = [1 - (1 + \beta)\theta_1 + \beta\theta_2 w]^2 - 4(\tfrac{1}{2} + \beta)\theta_2 w(g - \beta\theta_1)$ is positive, a condition that will always hold if g is not "too large." The reduced form of τ is then given by the smaller of the two roots, namely

$$\tau = \frac{1 - (1 + \beta)\theta_1 + \beta\theta_2 w - \Delta^{1/2}}{2(\tfrac{1}{2} + \beta)\theta_2 w}. \tag{11}$$

There are two differences between (9)–(11) and (6)–(7). First, in equation (9) the greater fiscal inefficiency raises the optimal inflation tax (note that $1 - R$ as given in equation (9) is uniformly greater than $(1 + \beta)\theta_1$ in the range $\tau \epsilon [0, 1]$). Second, equation (9) implies that higher values of τ (or, equivalently, g) will be associated with higher values of π. (That is, $\partial\tau/\partial g > 0$, as can be established by differentiating equation (10) with respect to g. From equation (9), it then follows that $\partial(1 - R)/\partial g > 0$, implying that the two tax instruments are normal.) By comparison, in the benchmark specification, π is constant and independent of g.

IV. Collection Lags and Optimal Inflation

In order to assess the impact of collection lags on the optimal tax menu, the basic model is now extended to incorporate a one-period delay in income tax payments. As before, a stream of government spending $[G(t)]$ is financed by income taxes $[\tau(t)]$ and currency issues $[M(t) - M(t-1)]$ for $t = 1, 2, \ldots$. Let the nominal income tax liability accrued at t, $\tau(t)p(t)w(t)$, become due at $t+1$. The payment lag allows workers (who also own the firms in the economy) to use these balances for an additional period and, assuming positive inflation, to reduce their tax liability in real terms. To the initial conditions $\bar{n}$ and $M(0)$ must now be added the nominal tax liability in the initial period, namely $\tau(0)p(0)w(0)$. (The real value of this nominal revenue is endogenous, as it depends on $p(1)$, the price level at time period 1.)

The delay in tax payments requires an elaboration of the time at which collection costs are incurred. Given the payment lag, nominal collections at t are $p(t-1)\tau(t-1)w(t-1)$. The real value of these collections in period t is $p(t-1)\tau(t-1)w(t-1)/p(t) = R(t-1)\tau(t-1)w(t-1)$, which is reduced by higher inflation. Two plausible formulations are considered. In this section, it is first assumed that real collection costs at t, $\phi(t)$, are a function of the real value of revenues collected at t. In other words, $\phi(t) = \phi[R(t-1)\tau(t-1)w(t-1)]$. This implies that the real cost of collecting a given nominal liability is *lowered* by inflation. While this is a plausible formulation, it must be emphasized that it is entirely ad hoc.[8] An alternative is to write the collection cost function as $\phi(t) = \phi[\tau(t-1)w(t-1)]$, implying that costs of collection are completely indexed for inflation.[9] This exercise is undertaken in a later section.

Formally, the private optimization problem is now to maximize $u[c_t(t), c_t(t+1)]$ subject to equations (12)–(13) below:

$$p(t)c_t(t) + m(t) \leq p(t)w(t), \tag{12}$$

$$p(t+1)c_t(t+1) \leq m(t) - \tau(t)p(t)w(t). \tag{13}$$

[8] Note that this formulation is consistent with the view that *nominal* collection costs at t, say $\Phi(t)$, are proportional to the nominal value of the liability to be collected, $\tau(t-1)p(t-1)w(t-1)$. Dividing through $\Phi[\tau(t-1)p(t-1)w(t-1)] = \theta_1\tau(t-1)p(t-1)w(t-1)$ by $p(t)$ shows that real collection costs, $\phi(t) = \Phi(t)/p(t)$, are equal to $\theta_1\tau(t-1)p(t-1)w(t-1)/p(t)$, or that $\phi(t) = \theta_1 R(t-1)\tau(t-1)w(t-1)$. This specification corresponds to Végh's and Dixit's benchmark cases of constant real marginal collection costs.

[9] The third alternative, namely when real collection costs are raised by higher inflation, is left as an exercise for the interested reader.

Using the notation developed earlier, equations (12)–(13) become

$$c_t(t) + s(t) \leq w(t) \tag{12'}$$

and

$$c_t(t+1) \leq R(t)s(t) - R(t)\tau(t)w(t). \tag{13'}$$

In equation (12′), $s(t)$ is an agent's *gross* real currency balance at the end of t; in equation (13′), the term $R(t)\tau(t)w(t)$ is his or her effective real tax liability. Combining (12′) and (13′), the agent's consumption set is

$$c_t(t) + c_t(t+1)/R(t) \leq A(t) \equiv [1 - \tau(t)]w(t). \tag{14}$$

As before, the solution to this problem is a pair of consumer demand schedules $c_1(A, R)$ and $c_2(A, R)$ and a real currency demand schedule $s = w - c_1$. It will be observed that given $R(t)$ and $\tau(t)$, individual consumption sets are not affected by payment lags. Anticipating the tax liability to be incurred one period ahead, agents react to the change in the timing of taxes by altering their financial decisions. The resulting increase in real currency holdings broadens the base of the inflation tax, thereby altering the government's choice set and the outcome of the optimal tax calculation.

Turning to the government budget constraint, since payment of taxes accrued at $t - 1$ is not made until t, we have that

$$\begin{aligned} p(t)G(t) = p(t-1)\tau(t-1)w(t-1) \\ - p(t)\phi[p(t-1)\tau(t-1)w(t-1)/p(t)] \\ + M(t) - M(t-1). \end{aligned} \tag{15}$$

Dividing equation (15) by $p(t)$ yields

$$\begin{aligned} G(t) = R(t-1)\tau(t-1)w(t-1) \\ - \phi[R(t-1)\tau(t-1)w(t-1)] \\ + [M(t) - M(t-1)]/p(t). \end{aligned} \tag{16}$$

In equation (16), the Tanzi effect is reflected in the term $R\tau w = \tau w / (1 + \pi)$ in the right-hand side: given the statutory rate τ, the real value of *gross* income tax collections is lowered by higher inflation. In addition, and provided that ϕ' is less than unity, the real value of *net* income tax collections, $R\tau w - \phi(R\tau w)$, is raised by lower rates of inflation. The latter statement can be established by differentiating $R\tau w - \phi(R\tau w)$ with respect to R, holding w and τ constant. This yields

$$(\partial/\partial R)[R\tau w - \phi(R\tau w)] = \tau w(1 - \phi') > 0 \qquad \text{if } \phi' < 1. \tag{17}$$

Stationary equilibria are pairs (R, τ) satisfying

$$G = R\tau w - \phi(R\tau w) + (1 - R)s(A, R). \tag{18}$$

Given a feasible choice of (R, τ), equilibrium sequences for prices and money are completely specified once the initial price level and currency issue, $p(1)$ and $M(1)$ respectively, are determined. Given values for G and the initial conditions $[p(0), \tau(0), w(0), M(0)]$, $p(1)$ and $M(1)$ may be calculated using the government budget constraint for the initial period $t = 1$,

$$G = p(0)\tau(0)w(0)/p(1) + M(1)/p(1) - M(0)/p(1), \tag{19}$$

and the stationary value of the real stock of currency $M(1)/p(1) = s(A, R)$. In principle, equations (19) and (3c) can be used to compare the initial price levels in the lag-free and lag-ridden economies. This comparison, which would establish the change in contemporaneous prices resulting from shorter payment lags, is beyond the steady-state comparisons undertaken in this paper and is left as an exercise for the interested reader.

The presence of payment lags introduces two mutually opposing forces that in principle have an ambiguous net effect on total "revenue" (seigniorage plus income taxes). On the one hand, owing to the Tanzi effect, higher inflation lowers the real value of revenue from a given rate of income tax. On the other hand, the increase in real money demand associated with the presence of collection lags makes a *given* rate of inflation more "productive." As suggested by Dixit, to ascertain how payment lags affect the desired rate of inflation in the face of these forces, the optimal tax problem must be recalculated in its entirety.

Formally, given the private decision rules $c_1(A, R)$, $c_2(A, R)$, and $s = w - c_1$, the government selects (τ, R) to maximize $u(w - s, Rs - R\tau w)$ subject to

$$G = R\tau w + (1 - R)s - \phi(R\tau w). \tag{20}$$

In the Appendix, it is shown that the optimal rate of inflation is given by

$$\pi = \frac{\phi'}{1 - s_A + \epsilon(s, R) + [\epsilon(s, A)\tau/(1 - \tau) - 1]}. \tag{21}$$

The Ramsey formula (21) retains the basic characteristics of equation (6). First, desired inflation continues to be proportional to the marginal cost of collection, and price stability should still be pursued whenever $\phi' = 0$. Second, the optimal inflation rate is inversely related to the marginal

propensity to consume and the interest elasticity of money demand. To ascertain the relative magnitude of desired inflation in the lag-free and lag-ridden environments, equations (6) and (21) may be compared directly. However, since τ is endogenous and the terms in the right-hand sides of (6) and (21) are functions, a simple comparison cannot be established without resorting to additional assumptions about preferences.

Sharper results are possible for the benchmark case considered by Dixit and Végh in which the functions in (6) and (21) are constants. When this comparison is undertaken it turns out that a theoretical case can be made for the orthodox presumption, at least for levels of government spending below a certain threshold.[10] Assuming $u(c_1, c_2) = \log(c_1) + \log(c_2)$ and $\phi' = \theta > 0$, optimal inflation π is easily calculated to be

$$\pi = 2\theta(1 + \tau)/(1 - \tau). \tag{22}$$

The reduced forms of τ and π can be computed analytically as follows. The real money demand function $s = w(1 + \tau)/2$ may be substituted into the stationary form of the government budget constraint (equation (18)). After some rearranging, the following relation between τ and R is established:

$$\tau = \frac{2g - (1 - R)}{2R(1 - \theta) + (1 - R)}. \tag{23}$$

Equation (23) can also be written in a form that is more convenient for subsequent substitutions:

$$\frac{1 + \tau}{1 - \tau} = \frac{1 - \theta + (1 + \pi)g}{1 - \theta + \pi - (1 + \pi)g}. \tag{23'}$$

Substituting the right-hand side of (23′) into (22) yields the following equation:

$$\pi = 2\theta \frac{1 - \theta + (1 + \pi)g}{1 - \theta + \pi - (1 + \pi)g}. \tag{24}$$

Equation (24) is seen to be equivalent to the quadratic equation

$$(1 - g)\pi^2 - (\theta + g - 1 + 2\theta g)\pi - 2\theta(1 - \theta + g) = 0. \tag{25}$$

[10] Analysis of the more realistic case of *quadratic* collection costs yields similar results. The proof of this assertion for the quadratic case involves some tedious algebra—finding the roots of a fourth-order polynomial equation—and is available upon request.

The roots of equation (25) are real if $\Delta = (\theta + g - 1 + 2\theta g)^2 + 8(1 - g)\theta(1 - \theta + g)$ is positive, in which case the reduced form of π, denoted π^L, is given by

$$\pi^L = \frac{(\theta + g - 1 + 2\theta g) + \Delta^{1/2}}{2(1 - g)}. \tag{26}$$

Given π^L from equation (26), the reduced form for τ^L may then be computed from equation (22). The pairs (τ^L, π^L) satisfying equations (26) and (22) are drawn against g in Figure 2 and against θ in Figure 3. It may be observed that even in the benchmark specification, the introduction of a payment lag makes desired inflation rise with g. This outcome may be contrasted with the results of Section II, where the optimal rate of inflation in the benchmark case was shown to be independent of g.[11] Given the definition $R \equiv 1/(1 + \pi)$, equation (6) describing the desired rate of inflation π^N may be written

$$\pi^N = 2\theta/(1 - 2\theta). \tag{27}$$

In the top panel of Figure 2, equation (26) is the upward-sloping curve, while equation (27) is the horizontal line. For a given value of θ, the relative size of inflation in the two environments depends on the size of g. In particular, there is a threshold value of g, say $g^* > 0$, which solves $\pi^L(g^*) = \pi^N(g^*)$ given that $\pi^L(g) < \pi^N(g)$ whenever $0 < g < g^*$ and that $\pi^L(g) > \pi^N(g)$ whenever $g > g^*$. In accordance with the orthodox position, the desired rate of inflation is lower in the presence of collection lags for all values of g not exceeding the threshold. The numerical value of g^* depends on the unit cost of collection and can be calculated analytically. Setting $\pi^L = \pi^N = 2\theta/(1 - 2\theta)$ in equation (26) yields the following expression for $g^*(\theta)$:

$$g^*(\theta) = \theta(2\theta^2 - 3\theta + 2)/(1 - \theta). \tag{28}$$

Figure 1 draws $g^*(\theta)$ and the associated $\tau(\theta)$ functions for the lag-free and lag-ridden economies. Table 1 provides illustrative calculations of g^* and the associated (τ, π) pairs for selected values of θ. Notice that g^* is strictly increasing in θ, with $g^*(0) = 0$ and $g^*(0.5) = 1$. It can be shown that for values of θ greater than approximately 0.27, $g^*(\theta)$ is strictly convex, implying that as the degree of fiscal inefficiency grows, the critical value of g grows at an increasing rate. Moreover, g^* approaches unity as θ approaches 0.5 from below.

[11] If marginal collection costs are increasing, inflation continues to be an increasing function of g.

V. An Invariance Proposition

In this section, we explore the implications of altering the specification of the collection cost technology. This is important because, as emphasized by Dixit (1991), no deep theory exists to explain the nature of these costs. It will be recalled that in the preceding section real collection costs were assumed to be a stable and increasing function of real *realized* tax revenues, $\phi = \phi(R\tau w)$, implying that, for given statutory tax rates and nominal collections, real collection costs are lowered by higher inflation. In this section, the function ϕ is written as $\phi = \phi(\tau w)$, so that collection costs are assumed to be a function of *accrued* real revenues—or, that nominal costs of collection rise in proportion with prices. The main result is a neutrality proposition establishing that the optimal (τ, R) pair is independent of the collection lag in the benchmark specification of constant marginal collection costs and logarithmic utility.

Formally, $R \geq 0$, and τ are selected to maximize social welfare

$$U = \log(c_1) + \log(c_2) \tag{29}$$

subject to the resource constraint (30) and private sector demand and supply schedules (31)–(34):

$$G = (1 - R)s + R\tau w - \phi(\tau w) \tag{30}$$

$$c_1 = \tfrac{1}{2}w(1 - \tau) \tag{31}$$

$$c_2 = \tfrac{1}{2}w(1 - \tau)R \tag{32}$$

$$s = \tfrac{1}{2}w(1 + \tau). \tag{33}$$

This problem is equivalent to selecting $R \geq 0$ and τ to maximize

$$U = 2\log(1 - \tau) + \log(R) \tag{34}$$

subject to

$$G = (1 - R)\tfrac{1}{2}w(1 + \tau) + R\tau w - \phi(\tau w). \tag{35}$$

In the constant marginal cost case, the problem above can be solved directly by substitution. Assuming $\phi(\tau w) = \theta\tau w$ for some $\theta > 0$, equation (35) may be used to write τ as a function of R and θ:

$$\tau = (2g + R - 1)/(R + 1 - 2\theta). \tag{36}$$

Since equation (36) implies

$$1 - \tau = 2(1 - \theta - g)/(R + 1 - 2\theta), \tag{37}$$

the objective function, equation (34), can be written as a function of a single variable R:

$$U = 2\log[2(1-\theta-g)/(R+1-2\theta)] + \log(R). \quad (38)$$

Maximizing U with respect to $R \geq 0$ is equivalent to maximizing

$$\Omega(R) \equiv \log(R) - 2\log(R+1-2\theta). \quad (39)$$

The first and second derivatives of Ω may be written

$$\Omega' = 1/R - 2/(R+1-2\theta) \quad (40)$$

and

$$\Omega'' = -1/R^2 + 2/(R+1-2\theta)^2. \quad (41)$$

From equation (40), the optimal value of R is $1-2\theta$, implying that the optimal inflation rate does not change if lags are removed.

Returning to the general specification of the collection cost function, the problem of maximizing equation (29) subject to equations (30)–(33) may be solved by forming the Lagrangian

$$L = 2\log(1-\tau) + \log(R) + \lambda[(1-R)\tfrac{1}{2}w(1+\tau) + R\tau w - \phi(\tau w) - G]. \quad (42)$$

The first-order necessary conditions for this problem are

$$1/R = \lambda w[(1+\tau)/2 - \tau], \quad (43)$$

$$2/(1-\tau) = \lambda w[(1-R)/2 + R - \phi']. \quad (44)$$

Dividing equation (43) by (44) yields $1 - R = 2\phi'$, which is identical to the formula for the inflation tax rate derived under the assumption that there were no collection lags ($\beta = 1$, $s_A = 0.5$, and $\epsilon(s, R) = 0$ in equation (6)). To prove that collection lags are neutral, it remains to show that the reduced form for τ is unchanged as well. This is first established for the linear marginal collection cost case: substituting equation (33) and $1 - R = 2[\theta_1 + \theta_2\tau w]$ into equation (30) and solving for τ yield the same formula as equation (10). Upon substitution, the optimal value of τ must solve

$$g = \theta_1 + \theta_2\tau w)(1+\tau) + (1 - 2\theta_1 - 2\theta_2\tau w)\tau - \theta_1\tau - (\theta_2/2)w\tau^2, \quad (45)$$

which, after a few steps of algebra, is equivalent to

$$0 = (3/2)\theta_2 w\tau^2 + (1 - 2\theta_1 + \theta_2 w)\tau - (g - \theta_1). \quad (46)$$

Inspection of equation (46) reveals that it is equivalent to equation (10) for $\beta = 1$. Finally, to establish the equivalence for the constant marginal cost case, simply note that equation (46) with $\theta_2 = 0$ implies equation (7). This establishes the claim that the presence of collection lags leaves unaltered the optimal (R, τ) pair in both the linear and quadratic collection cost specifications.

A note of caution is in order when interpreting this invariance proposition. While the proposition does suggest that the optimal inflation tax *rate* $1 - R$ is invariant to the lag, it does not suggest invariance of inflation tax *revenue*, where the latter is defined as the product of the inflation tax rate and the real currency stock $(1 - R)s(A, R)$. Analogously, this proposition does not suggest invariance of income tax revenue. Clearly, with inelastic labor supply, a positive rate of inflation, and an unchanged rate of labor income taxation, the introduction of collection lags lowers the effective yield of labor taxes. According to the proposition, all revenue losses are made up by higher real seigniorage earnings. In the presence of an unchanged rate of inflation, the additional seigniorage earnings are possible by the higher stock of real balances accumulated to meet future income tax liabilities.

VI. Conclusions

This paper has analyzed how the presence of collection lags affects the optimal inflation rate in an overlapping generations version of the Frenkel-Végh model of costly income taxation. Ramsey pairs (τ, π) were derived from first principles, and the traditional argument, that inflation ought to reduce the optimal rate of inflation, was confirmed for cases where the level of government spending in GDP was below a threshold level. Dixit's conclusions were confirmed for tax rates above the threshold. This demonstration casts some doubt on the claim that, as a general principle, the Tanzi hypothesis is inconsistent with optimal tax theory.

APPENDIX

The Appendix derives the optimal tax formula applicable in Section IV. The following notation is used:

w = real pretax income;

τ = income tax rate;

$A = (1 - \tau)w$ = after-tax income;

c_1 = consumption in first period of life;

c_2 = consumption in second period of life;

G = rate of real government purchases;

$\phi(\cdot)$ = collection cost function;

$\phi'(\cdot)$ = marginal cost of income tax collections;

π = rate of inflation;

$R = 1/(1 + \pi)$ = gross real yield of currency;

$1 - R$ = rate of inflation tax.

Given the private decision rules $c_1(A, R)$, $c_2(A, R)$, and $s = w - c_1$, the pair (τ, R) is selected to maximize indirect utility $u(w - s, Rs - R\tau w)$ subject to

$$G = R\tau w + (1 - R)s - \phi(R\tau w). \tag{A1}$$

The Lagrangian expression for this problem is

$$L = u(w - s, Rs - R\tau w) + \lambda[R\tau w + (1 - R)s - \phi(R\tau w) - G]. \tag{A2}$$

The first-order necessary conditions for an interior solution are

$$\tau\colon u_1(-s_A A_\tau) + u_2 R(s_A A_\tau - w) + \lambda[wR(1 - \phi') + (1 - R)s_A A_\tau] = 0, \tag{A3}$$

and

$$R\colon u_1(-s_R) + u_2(s + RS_R - \tau w) + \lambda[\tau w(1 - \phi') - s + (1 - R)s_R] = 0. \tag{A4}$$

Substituting $A_\tau = -w$ and $u_1 = Ru_2$ into equations (A3)–(A4) yields

$$u_2 R = \lambda[R(1 - \phi') - (1 - R)s_A], \tag{A5}$$

and

$$u_2(s - \tau w) = -\lambda[\tau w(1 - \phi') - s + (1 - R)s_R]. \tag{A6}$$

Equations (A5)–(A6), in turn, imply that

$$\frac{-R}{s - \tau w} = \frac{R(1 - \phi') - (1 - R)s_A}{\tau w(1 - \phi') - s + (1 - R)s_R}. \tag{A7}$$

Next, cross-multiplication in equation (A7) yields

$$-R\tau w(1 - \phi') + Rs - R(1 - R)s_R = sR(1 - \phi') - s(1 - R)s_A - \tau wR(1 - \phi') + \tau w(1 - R)s_A. \tag{A8}$$

After some simplification, equation (A8) can be written

$$-R(1 - R)s_R = -sR\phi' - s(1 - R)s_A + \tau w(1 - R)s_A \Rightarrow \tag{A9}$$

$$(1 - R)Rs_R/s = R\phi' + (1 - R)s_A - (1 - R)s_A \tau w/s. \tag{A10}$$

Note that the expression Rs_R/s on the left-hand side of equation (A10) is the interest elasticity of real currency demand $\epsilon(s, R)$, and the term $s_A \tau w/s$ equals $\epsilon(s, A)\tau/(1 - \tau)$. Equation (A10) may be rewritten as follows:

$$(1 - R)\epsilon(s, R) = R\phi' + (1 - R)s_A - (1 - R)\epsilon(s, A)\tau/(1 - \tau). \tag{A11}$$

Collecting the $(1 - R)$ terms in equation (A11) and rearranging yield

$$(1 - R)[\epsilon(s,R) - s_A + \epsilon(s,A)\tau/(1 - \tau)] = R\phi'. \quad \text{(A12)}$$

Since the factor $(1 - R)/R$ equals the rate of inflation π, equation (A12) yields the Ramsey rule for inflation in the presence of collection lags:

$$\pi = \frac{\phi'}{\epsilon(s,R) - s_A + \epsilon(s,A)\tau/(1 - \tau)}. \quad \text{(A13)}$$

Equation (A13) corresponds to Ramsey formula (21).

REFERENCES

Diamond, Peter, and James Mirrlees, "Optimal Taxation and Public Production: I. Production Efficiency, and II. Tax Rules," *American Economic Review*, Vol. 61 (1971), pp. 8–27 and 261–78.

Dixit, Avinash, "The Optimal Mix of Inflationary Finance and Commodity Taxation with Collection Lags," *Staff Papers*, International Monetary Fund, Vol. 38 (September 1991), pp. 643–54.

Frenkel, Jacob, "Comments on Zvi Hercowitz and Efraim Sadka: Optimal Currency Substitution Policy and Public Finance," in *Economic Policy in Theory and Practice*, ed. by Assaf Razin and Efraim Sadka (London: Macmillan, 1987).

Friedman, Benjamin M., "The Optimum Quantity of Money," in *Optimum Quantity of Money and Other Essays* (Chicago: Aldine, 1969).

Kimbrough, Kent S., "The Optimum Quantity of Money Rule in the Theory of Public Finance," *Journal of Monetary Economics*, Vol. 18 (1986), pp. 277–84.

Mourmouras, Alex, "Optimal Inflation and Income Taxes in a Model with an Untaxed Sector and a Black Market in Foreign Exchange" (unpublished; University of Cincinnati, 1991).

Phelps, Edmund S., "Inflation in the Theory of Public Finance," *Swedish Journal of Economics* (1973), pp. 867–82.

Samuelson, Paul A., "A Pure Consumption-Loan Model of Interest With or Without the Social Contrivance of Money," *Journal of Political Economy*, Vol. 66 (1958), pp. 467–82.

Tanzi, Vito, "Inflation, Real Tax Revenue, and the Case for Inflationary Finance: Theory with an Application to Argentina," *Staff Papers*, International Monetary Fund, Vol. 25 (September 1978), pp. 417–51.

———, "Theory and Policy: A Comment on Dixit and on Current Tax Theory," *Staff Papers*, International Monetary Fund, Vol. 39 (December 1992), pp. 957–66.

Végh, Carlos A., "Government Spending and Inflationary Finance: A Public Finance Approach," *Staff Papers*, International Monetary Fund, Vol. 36 (September 1989), pp. 657–77.

———, and Pablo Guidotti, "The Optimal Inflation Tax When Money Reduces

Transactions Costs: A Reconsideration," *Journal of Monetary Economics*, Vol. 31 (1993), pp. 189–205.

Wallace, Neil, "The Overlapping Generations Model of Fiat Money," in *Models of Monetary Economies*, ed. by John H. Kareken and Neil Wallace (Minneapolis: Federal Reserve Bank of Minneapolis, 1980).

Woodford, Michael, "The Optimum Quantity of Money," in *Handbook of Monetary Economics*, Vol. 2, ed. by Benjamin M. Friedman and Frank H. Hahn (Amsterdam: North-Holland, 1990).

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

The Parallel Market Premium

Is It a Reliable Indicator of Real Exchange Rate Misalignment in Developing Countries?

PETER J. MONTIEL and JONATHAN D. OSTRY*

It is often argued that the parallel market premium is a useful indicator of real exchange rate misalignment in developing countries. The empirical evidence, however, does not suggest a robust correlation between these two endogenous variables that is independent of the nature of economic shocks and various structural relationships in the economy. This paper analyzes the reliability of the parallel market premium as an indicator of real exchange rate misalignment. It suggests that one should exercise caution in drawing inferences about the sign and magnitude of real exchange rate misalignment from the premium. [JEL F31, F41, O11]

REAL EXCHANGE RATE misalignment—the sustained departure of the actual real exchange rate from its equilibrium value—has been a recurring policy problem in many developing countries. Overvaluation of the real exchange rate has had undesirable effects on net exports and growth in some countries, while undervaluation has created problems for monetary control and inflation in others. Getting the real exchange rate

* Peter J. Montiel is the Danforth-Lewis Professor of Economics at Oberlin College and was formerly a Deputy Division Chief in the IMF's Research Department. Work on this project was undertaken in part while Montiel was a Visiting Scholar in the Research Department. He is a graduate of Yale University and the Massachusetts Institute of Technology.
Jonathan D. Ostry, an economist in the Research Department, holds a doctorate from the University of Chicago, as well as degrees from the London School of Economics, Oxford University, and Queen's University.
The authors would like to thank Mohsin Khan and Peter Wickham for useful comments on a previous draft, and Carlos Végh for helpful discussions on this topic.

"right" in a world in which the fundamental structural determinants of the equilibrium real exchange rate are constantly changing remains one of the most important goals of economic policy in developing countries.

The problems associated with real exchange rate misalignment have led policymakers in developing countries to try to reduce the degree of overvaluation or undervaluation of their currencies by appropriate policies for the nominal exchange rate. The idea behind such policies is that the transition from, for example, a situation of overvaluation to one of real exchange rate equilibrium can be long and drawn out, particularly if institutional factors—such as wage- and price-setting behavior—are unfavorable. Under these circumstances, a change in the nominal exchange rate (a devaluation) can help to reestablish equilibrium more quickly and thereby mitigate some of the costs associated with the transition.

A well-known problem associated with the implementation of such policies is that the extent of real exchange rate misalignment—and in some cases even its sign—may be difficult to gauge. This is because information about the extent of misalignment requires knowledge of the level of the equilibrium real exchange rate, which depends both on structural factors (including trade and industrial policies, the degree of capital mobility, and the terms of trade) and on macroeconomic factors, such as the level and composition of government spending and taxation as well as the international macroeconomic environment. The problem of real exchange rate misalignment thus lies at the heart of both macroeconomic and structural policy in developing countries.

In practice, the problem of estimating the extent of real exchange rate misalignment has been addressed in a variety of ways. One approach has been to determine some base period in which the actual and equilibrium real exchange rates were equal and then attempt to determine the extent by which the equilibrium real rate has changed as a result of changes in its fundamental structural determinants, so that a comparison can be made with the path of the actual real exchange rate.[1] This approach, however, has not been used as frequently as one might think, probably because the effects of exogenous and policy-induced shocks on the equilibrium real exchange rate depend on structural relationships in the economy about which policymakers are likely to have insufficient information.

An alternative approach, which is simpler and more direct, involves using information from the parallel market to gauge the extent of real

[1] For an attempt at such an exercise, see Khan and Ostry (1992).

exchange rate misalignment.[2] The existence of a premium on foreign exchange in the free market is taken to indicate an excess demand for foreign exchange at the official exchange rate, which in turn is interpreted as arising from an overvaluation of the domestic currency at the prevailing official exchange rate.

This intuition is buttressed by models in the literature in which the premium and the degree of misalignment both respond endogenously to some shock, such as an increase in the stock of credit. For example, models in the currency-substitution tradition (Calvo and Rodríguez (1977)) have been used by Edwards (1989) and Kamin (1993) to analyze the effects of unsustainable financial policies on the parallel market premium and the divergence of the real exchange rate from its long-run equilibrium value. The robust finding from these models is that, along the adjustment path, overvalued real exchange rates are associated with high premia. One interpretation of this relationship views the equilibrium exchange rate as a weighted average of the free rate and the official rate. Consequently, in many developing countries, exchange rate policy is designed to reduce the gap between the two rates by depreciating the official exchange rate (see, for example, Aghevli, Khan, and Montiel (1991)). In effect, the misalignment of the official rate is contained by targeting the premium at a reduced level.[3]

This view of the relationship between the premium and real exchange rate misalignment has some empirical support. For example, studies of a number of devaluation episodes (see Edwards (1989) and Kamin (1993)) find that the parallel market premium often rises very rapidly in the period immediately preceding a major devaluation, and then falls off just after the devaluation. A positive correlation between the premium and the extent of underlying real exchange rate overvaluation is suggested by such cases.

Nonetheless, from an analytical standpoint, the case for treating the size of the parallel market premium as an indicator of the magnitude of real exchange rate misalignment seems far from obvious. Both the premium on foreign exchange in the free market and the real exchange rate in the official market are endogenous variables with complex macroeconomic roles. As such, the correlation between them should depend on the sources of shocks impinging on the economy. Moreover, the parallel market premium is an asset price, which can be expected to exhibit much greater volatility than the official real exchange rate, in particular by

[2] On the use of the premium as an indicator of real exchange rate misalignment in developing countries, see Quirk and others (1987).

[3] For an analysis of the Bolivian case, see Kharas and Pinto (1989).

responding to transitory shocks that leave the equilibrium real exchange rate unaffected. The very different time series properties of the two variables raise some doubts about the reliability of the premium as an indicator of real exchange rate misalignment.

Empirically, parallel market premia tend to be a good deal more variable than real exchange rates in developing countries and can easily reach levels of several hundred percent without discernible changes in the underlying extent of real exchange rate misalignment. Furthermore, the sign of the correlation between the parallel market premium and the official real exchange rate seems to vary from country to country and time period to time period.[4] For example, in the study cited previously, Kamin (1993) found that in about a third of the devaluation episodes he studied, the premium actually fell before an exchange rate correction. This certainly calls into question the presumption that there is a robust correlation between real exchange rate misalignment and the parallel market premium that is independent of the nature of the underlying shocks and the structure of the economy.

The purpose of this paper is to explore these issues in the context of a fully optimizing model of a developing country that simultaneously determines the degree of misalignment of the real exchange rate and the premium in the parallel market. The paper's objective is to assess the reliability of the premium as an indicator of exchange rate misalignment by answering the following questions: Letting e and $\bar{e}$ denote respectively the actual and equilibrium real exchange rates, and letting b denote (one plus) the premium in the free exchange market, can we identify a structural relationship $e - \bar{e} = f(b - 1)$ that would provide *quantitative* information about the degree of real exchange rate overvaluation from observations of the premium? If not, can we at least establish that the sign of $(b - 1)$ is the same as the sign of $(e - \bar{e})$, so that the premium serves as a *qualitative* indicator of misalignment? If neither of these can be established in a standard model, then the former cannot be taken as a reliable indicator of misalignment.

This paper is organized as follows. First, we set up the model by describing the consumer's optimization problem and then discussing the equilibrium and solution of the model. We then proceed to demonstrate the central point of the paper, that adjustment in an economy (modeled along fairly standard lines) to even a simple shock (consisting in this case of a permanent productivity shock) can be rather complex. We show, in particular, that while the adjustment of the real exchange rate to its new

[4] See, for example, the discussion in Edwards (1989), chapter 4.

long-run equilibrium is monotonic, that of the parallel market premium is not, implying that neither the sign nor the magnitude of the premium is informative about the extent of misalignment in this case. We end by summarizing the main conclusions.

I. The Household's Problem

Consider a small open economy in which agents derive utility from the consumption of traded and nontraded goods, denoted respectively by c_T and c_N. The instantaneous utility function u is thus given by

$$u(t) = u[c_T(t), c_N(t)]. \tag{1}$$

Consumption is subject to a cash-in-advance constraint of the form

$$m(t) = \alpha c(t), \tag{2}$$

where $m = M/P$ is the stock of money deflated by the true consumption-based price index P; c is the real value of aggregate consumption, which is related to the consumption of traded and nontraded goods by $c = e^{\beta}(c_T + c_N/e)$, where e is the price of tradables relative to nontradables (the real exchange rate) and $0 < \beta < 1$ is a parameter whose economic interpretation will be given below; and $\alpha > 0$. Apart from domestic money, household financial wealth includes foreign currency denominated securities, which are denoted by F.

A "dual" exchange rate regime is assumed, involving a fixed "official" exchange rate s and a floating "unofficial," or "parallel," market exchange rate v. All commercial transactions are assumed to take place at the official rate s, while all other (financial) transactions take place at the market-determined rate v. We let b denote the ratio of the financial to the commercial exchange rate, so that $b = v/s$ denotes one plus the premium in the parallel market. Thus, the real value (in terms of traded goods) of household financial wealth, denoted by a, may be written as

$$a = me^{-\beta} + bF. \tag{3}$$

It is assumed that agents can engage in cross-transactions between official and unofficial markets. These leakages, denoted by L, carry a cost $T(L)$, where $T(0) = 0$, $T'(0) = 0$, and $T''(0) > 0$.[5] Thus, the increase in resources available to consumers by engaging in "fraudulent" activity

[5] The first optimizing model of dual rates with leakages is Bhandari and Végh (1990). The main differences between their paper and ours relate to the goods structure and the economy's access to world capital markets.

(leakages) must be traded off against the costs associated with such activity, which are captured by the $T(L)$ function. Taking into account these leakages, the instantaneous budget constraint of the representative household may be written as

$$\dot{a} = y + (r^* + \hat{b})bF - c_T - c_N/e - T(L) + (b - 1)L, \quad (4)$$

where a dot above a variable denotes a time derivative and a circumflex denotes a proportional rate of change—that is, $\hat{x} = \dot{x}/x$. Also, y denotes the real value (in terms of traded goods) of output of both traded and nontraded goods;[6] r^* is the world interest rate; and the last two terms in equation (4) respectively denote the costs associated with fraudulent transactions and the benefits associated with such activity. Thus, the change in real household wealth (per unit of time) is simply equal to income (which consists of income from production (y) plus interest earnings on foreign securities (r^*bF) plus capital gains ($\hat{b}bF$)) less consumption, plus the *net* benefits associated with fraudulent activity.

The problem of the representative household is to choose values for its consumption of traded and nontraded goods (c_T and c_N), foreign securities (F), and the amount of leakage into the free market (L), to maximize the discounted sum of utility given by

$$\int_0^\infty u[c_T(t), c_N(t)] \exp(-\delta t)\, dt, \quad (5)$$

subject to the budget constraint given by equation (4) and a transversality condition that requires the present value of household net worth to converge to a nonnegative number. In addition to the budget constraint (4) and the transversality condition, the optimality conditions for this problem are given by

$$u_1[c_T(t), c_N(t)] = \lambda[1 + \alpha(r^* + \hat{b})], \quad (6)$$

$$u_2[c_T(t), c_N(t)] = \lambda[1 + \alpha(r^* + \hat{b})]/e, \quad (7)$$

$$T'(L) = b - 1, \quad (8)$$

$$\dot{\lambda} = \lambda[\delta - (r^* + \hat{b})]. \quad (9)$$

The interpretation of these conditions is standard. The first two equations require that the marginal utility of consumption of each of the two goods be equated with the product of the marginal utility of wealth, λ, and the price of the good. The price itself involves two components. First, there is the market price (unity in the case of traded goods and the

[6] The production side of the economy is described in the following section.

reciprocal of the real exchange rate e in the case of home goods). Second, because of the cash-in-advance constraint, there is the opportunity cost of holding money, which is α (the factor of proportionality between consumption and money) times the real interest rate, $r^* + \hat{b}$. Clearly, the domestic real interest rate is equal to the world rate plus a term (the rate of change in the parallel market premium) that captures the capital gains or losses from holding foreign currency denominated assets.

The third optimality condition (equation (8)) relates the amount of leakage to the level of the parallel market premium. Given that $T(L)$ is a convex function, the condition states that the higher the premium $b - 1$, the greater the amount of leakage between the official and parallel markets. Finally, equation (9) is the standard condition relating the path of the co-state variable to the difference between the subjective rate of time preference and the domestic real interest rate.

We proceed under the simplifying assumption that the instantaneous utility function in equation (1) is logarithmic:

$$u[c_T(t), c_N(t)] = (1 - \beta) \ln c_T(t) + \beta \ln c_N(t). \qquad (1')$$

Denoting aggregate consumption in terms of traded goods by $Z = c_T + c_N/e$, we have from equation (1′) and the first two optimality conditions that the consumption of traded and nontraded goods is given by

$$c_T(t) = (1 - \beta)Z(t), \quad \text{and} \quad c_N(t) = \beta e(t)Z(t). \qquad (10)$$

Thus, consumption of the two goods depends on the level of total expenditure Z and the real exchange rate e.[7] We can also solve for the marginal utility of wealth in terms of aggregate expenditure and the domestic interest rate:

$$\lambda = 1/\{Z[1 + \alpha(r^* + \hat{b})]\}. \qquad (11)$$

II. Equilibrium and Solution of the Model

We begin by discussing some issues that affect the dynamics of the aggregate economy but that are exogenous from the point of view of the individual household and hence were ignored in the previous section. First, it should be noted that for the economy as a whole, the legal stock of foreign securities is fixed and the actual stock can therefore be altered only through leakages. Recalling that interest earnings on foreign secu-

[7] Clearly, β is the expenditure share of nontraded goods.

rities are repatriated through the parallel market (since the official market is, by assumption, only for commercial transactions), we thus have

$$\dot{F} = L + r^*F, \tag{12}$$

where the second term represents interest earnings on the existing stock of foreign securities.

Second, the country is assumed to face an upward-sloping supply of foreign loans—that is, the cost of funds to domestic private residents depends on the actual stock of foreign securities they hold. Specifically, it is assumed that $r^* = r^*(F)$, with $r^*(F)$ being a decreasing function that satisfies $r^*(0) = \delta$, where δ is the rate of time preference. Thus, when the country is neither a net creditor nor a net debtor ($F = 0$), the rate of interest its residents face is equal to the rate of time preference δ, but the rate of interest faced by the domestic residents rises with $-F$—that is, with the extent of the country's net debtor position. This makes r^* a decreasing function of F.[8]

Third, in the previous section, the household took the production side of the economy as given. To close the model, however, the conditions of production must be specified. We adopt a simple approach here, wherein there is no investment or other dynamic problem that firms need to solve and production decisions maximize period-by-period profits. It is also assumed that all prices are flexible so that the economy operates continually at full employment.

A convenient way of modeling the supply side is to assume that the economy possesses a concave transformation frontier between traded and nontraded goods.[9] It operates at the point of tangency between the frontier and a straight line with slope equal to (minus) the reciprocal of the real exchange rate. This implies supply functions of the form:

$$y_T = y_T(e;\theta), \qquad \partial y_T/\partial e > 0, \qquad \partial y_T/\partial \theta > 0;$$

$$y_N = y_N(e), \qquad \partial y_N/\partial e < 0, \tag{13}$$

where θ is a productivity parameter in the traded goods sector to be discussed below.[10] Using the definition of output in terms of traded goods y, together with the envelope condition, also implies

[8] Essentially, this assumption pins down the stock of foreign securities in the steady state.

[9] For a similar modeling of the supply side, see Lizondo and Montiel (1991).

[10] To simplify the analysis, it is assumed that the output (productivity) shock affects only the tradables sector. This implies that the productivity shock does not enter into equation (16) below, which relates the equilibrium real exchange rate to the level of total expenditure.

$$y(e;\theta) = y_T(e;\theta) + y_N(e)/e, \qquad y_1 \equiv \partial y/\partial e > 0,$$
$$y_2 \equiv \partial y/\partial\theta > 0. \tag{14}$$

Finally, we can use this simple specification of production together with the demand functions given in equation (10) to solve for the equilibrium real exchange rate in terms of the level of total expenditure Z. Specifically, since in equilibrium the nontraded goods market must clear, we have

$$y_N(e) = \beta e Z. \tag{15}$$

Differentiating equation (15) we can solve for the equilibrium real exchange rate in terms of total expenditure:

$$e = e(Z), \qquad e'(Z) < 0. \tag{16}$$

Clearly, an increase in total expenditure creates excess demand for home goods, thereby requiring an appreciation of the real exchange rate (a decrease in e) to restore equilibrium in the home goods market.

With these three issues in hand, the specification of the model is complete. Notice that the analytical set-up is a fairly standard one based on familiar components, so that the co-movements between the premium and the real exchange rate in response to shocks should be of particular interest.

We shall derive four dynamic equations, two of which refer to the "jumping" variables b and the domestic real interest rate $r^* + \hat{b}$, which we denote as R, and two of which involve the predetermined variables Z and F. We begin by deriving the dynamics of private consumption expenditure Z. Equation (11) links Z to the marginal utility of wealth. Proceeding in standard fashion, replace $r^* + \hat{b}$ by R in equation (11), differentiate with respect to time, and substitute from equation (9). The result is

$$\dot{Z} = \{R - \delta - [\alpha R/(1 + \alpha R)]\hat{R}\}Z, \tag{17}$$

which is the familiar condition relating the slope of the time path of consumption to the difference between the consumption real interest rate and the discount rate.[11]

[11] In any period, the effective price of consumption is equal to $1 + \alpha R$, which is the sum of its market price (unity) and the opportunity cost (αR) of holding the α units of money needed to purchase a unit of the good. The consumption rate of interest, which is the true intertemporal price of consumption, is equal to the interest rate R less the rate of change of the effective price of consumption, $[\alpha R/(1 + \alpha R)]\hat{R}$.

Substituting the cash-in-advance constraint into the expression for household wealth (recalling that $c = e^{\beta} Z$) gives the following alternative expression for wealth:

$$a = \alpha Z + bF. \tag{18}$$

Differentiating equation (18) with respect to time, equating the resulting expression for $\dot{a}$ to the one given in equation (4), and using equation (12) to eliminate $\dot{F}$ give the following alternative expression for $\dot{Z}$:

$$\dot{Z} = (1/\alpha)[y - Z - T(L) - L]. \tag{19}$$

Inverting the optimality condition in equation (8), we can write

$$L = H(b - 1), \tag{20}$$

where $H'(b - 1) = 1/T''(L) > 0$. More intuitively, equation (20) states that the amount of leakage is an increasing function of the parallel market premium $b - 1$. Substituting equation (20) into equation (19) then gives

$$\dot{Z} = (1/\alpha)\{y - Z - T[H(b - 1)] - H(b - 1)\}. \tag{21}$$

Equating the right-hand sides of equations (17) and (21) provides a dynamic equation for the domestic real interest rate:

$$\dot{R} = -[(1 + \alpha R)/(\alpha^2 Z)]\{y - Z - T[H(b - 1)] - H(b - 1)\} + (R - \delta)(1 + \alpha R)/\alpha. \tag{22}$$

The dynamic system that characterizes the economy comprises the following four equations:

$$\dot{b} = [R - r^*(F)]b, \tag{23}$$

$$\dot{Z} = (1/\alpha)\{y - Z - T[H(b - 1)] - H(b - 1)\}, \tag{24}$$

$$\dot{R} = -[(1 + \alpha R)/(\alpha^2 Z)]\{y - Z - T[H(b - 1)] - H(b - 1)\} + (R - \delta)(1 + \alpha R)/\alpha, \tag{25}$$

$$\dot{F} = H(b - 1) + r^*(F)F. \tag{26}$$

Equation (23) simply follows from the definition of the real interest rate; equations (24) and (25) were given previously as equations (21) and (22) and are rewritten here for convenience; equation (26) incorporates the leakage function (20) and the dependence of the external interest rate facing domestic residents on the stock of foreign assets into the equation for $\dot{F}$ given previously as equation (12). The state variables are Z and F, while the non-predetermined variables are b and R.

Imposing the steady-state conditions $\dot{b} = \dot{Z} = \dot{R} = \dot{F} = 0$, we can solve for the steady-state values (where an overbar denotes a steady state):

$$\bar{b} = 1, \tag{27}$$

$$\bar{Z} = y[e(\bar{Z});\theta], \tag{28}$$

$$\bar{R} = \delta, \tag{29}$$

$$\bar{F} = 0. \tag{30}$$

The logic of the steady state is seen by observing that if the net stock of foreign securities reaches a constant value that differs from zero, this implies that the constant value reached by the premium $b - 1$ must also differ from zero (equation (26)) and that the value of the world interest rate must differ from the rate of time preference (by definition of the relationship between F and $r^*(F)$). This would make the last term in equation (25) nonzero, which in turn implies that the term in curly brackets in that equation must also be nonzero. Since this term also appears on the right-hand side of equation (24), it would not be possible for expenditure Z to reach a constant value, which contradicts the definition of the steady state. Therefore, the steady-state values must be those given by equations (27)–(30).

Our interest is in analyzing the co-movements of the premium and the real exchange rate in this economy when the economy is hit by an exogenous shock. For that purpose, we consider a simple shock, in the form of a permanent improvement in productivity. In the remainder of this section, we provide a formal solution for the response dynamics, which we then analyze in more detail in the section that follows. To analyze the dynamics in the vicinity of the new steady state, we linearize the system of equations (23)–(26) around the steady state:

$$\begin{bmatrix} \dot{b} \\ \dot{R} \\ \dot{Z} \\ \dot{F} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & b_4 \\ R_1 & R_2 & R_3 & 0 \\ Z_1 & 0 & Z_3 & 0 \\ F_1 & 0 & 0 & F_4 \end{bmatrix} \begin{bmatrix} db \\ dR \\ dZ \\ dF \end{bmatrix} + \begin{bmatrix} 0 \\ R_5 \\ Z_5 \\ 0 \end{bmatrix} d\theta. \tag{31}$$

The partial derivatives in matrix system (31) are given by

$$b_4 = -r^{*\prime}(F) > 0,$$

$$R_1 = [(1 + \alpha R)/(\alpha^2 Z)]H' > 0,$$

$$R_2 = (1 + \alpha R)/\alpha > 0,$$

$$R_3 = [(1 + \alpha R)/(\alpha^2 Z)](1 - y_1 e') > 0,$$

$$R_5 = -[(1 + \alpha R)/(\alpha^2 Z)]y_2 < 0,$$

$$Z_1 = -H'/\alpha < 0,$$

$$Z_3 = -(1 - y_1 e')/\alpha < 0,$$

$$Z_5 = y_2/\alpha > 0,$$

$$F_1 = H' > 0,$$

$$F_4 = \delta > 0.$$

The trace of the matrix on the right-hand side of (31) is given by

$$\text{Trace} = 2\delta + (y_1 e')/\alpha > 0,$$

while the determinant is given by

$$\text{Determinant} = -r^{*\prime}H'(1 + \alpha\delta)(1 - y_1 e')/\alpha^2 > 0.$$

The fact that the trace is positive implies that not all roots of the system can be negative, and the fact that the determinant is positive implies that the number of positive roots is either two or four. Since there are two predetermined variables and two non-predetermined variables, the former case (two positive roots) implies saddle-point stability of the equilibrium, whereas the case of four positive roots implies global instability.

To determine which of these two cases is relevant, we examine the characteristic polynomial in more detail. Defining this polynomial by $P(\lambda)$, we have by definition

$$P(\lambda) = \det\begin{bmatrix} -\lambda & 1 & 0 & b_4 \\ R_1 & R_2 - \lambda & R_3 & 0 \\ Z_1 & 0 & Z_3 - \lambda & 0 \\ F_1 & 0 & 0 & F_4 - \lambda \end{bmatrix}. \tag{32}$$

Expanding the expression for the polynomial, one obtains

$$P(\lambda) = \gamma_0 + \gamma_1\lambda + \gamma_2\lambda^2 + \gamma_3\lambda^3 + \gamma_4\lambda^4. \tag{33}$$

Upon substitution for the coefficients as defined below matrix system (31), one obtains

$$\gamma_0 = -r^{*\prime}H'(1 + \alpha\delta)(1 - y_1 e')/\alpha^2 > 0,$$

$$\gamma_1 = r^{*\prime}H'[\delta + (y_1 e')/\alpha] - (\delta/\alpha)(1 + \alpha\delta)[(1 - y_1 e')/\alpha + H'/(\alpha Z)] < 0,$$

$$\gamma_2 = \delta[\delta + (y_1 e')/\alpha] - [(1 + \alpha\delta)/(\alpha^2 Z)][H' + (y_1 e' - 1)^2/\alpha] + r^{*\prime}H' \lesseqgtr 0,$$

$$\gamma_3 = 2\delta + (y_1 e')/\alpha > 0,$$

$$\gamma_4 = 1.$$

By definition, γ_0 is equal to the product of all the roots (the determinant), γ_1 is equal to the sum of all products of three of the roots, γ_2 is equal to the sum of all products of two of the roots, and γ_3 is equal to the sum of all products of one of the roots (or the trace). Because γ_1 is negative, it must be the case that at least one of the roots of the system is negative. This rules out the case in which all roots are positive and implies that the system is saddle-point stable, possessing exactly two positive roots and two negative roots.

The general solution of the model may therefore be written as

$$\begin{bmatrix} b \\ R \\ Z \\ F \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{bmatrix} \begin{bmatrix} \omega_1 \exp(\lambda_1 t) \\ \omega_2 \exp(\lambda_2 t) \\ \omega_3 \exp(\lambda_3 t) \\ \omega_4 \exp(\lambda_4 t) \end{bmatrix}, \tag{34}$$

where the elements of each eigenvector may be solved from

$$0 = \begin{bmatrix} -\lambda_i & 1 & 0 & b_4 \\ R_1 & R_2 - \lambda_i & R_3 & 0 \\ Z_1 & 0 & Z_3 - \lambda_i & 0 \\ F_1 & 0 & 0 & F_4 - \lambda_i \end{bmatrix} \begin{bmatrix} h_{1i} \\ h_{2i} \\ h_{3i} \\ h_{4i} \end{bmatrix}, \tag{35}$$

for each of the four eigenvalues $\lambda_i, i = 1 - 4$. If we let λ_3 and λ_4 be the two positive eigenvalues, then convergence of the dynamic system requires $\omega_3 = \omega_4 = 0$. To solve for the remaining two scalars, we use the initial conditions, which imply

$$\omega_1 = h_{42}/(h_{31}h_{42} - h_{32}h_{41})(Z_0 - \bar{Z}) - h_{32}/(h_{31}h_{42} - h_{32}h_{41})(F_0 - \bar{F}), \tag{36}$$

$$\omega_2 = h_{31}/(h_{31}h_{42} - h_{32}h_{41})(Z_0 - \bar{Z}) - h_{41}/(h_{31}h_{42} - h_{32}h_{41})(F_0 - \bar{F}). \tag{37}$$

We can now write the solution of the system as

$$b - \bar{b} = \Delta^{-1}\{[h_{11}h_{42} \exp(\lambda_1 t) + h_{12}h_{31} \exp(\lambda_2 t)](Z_0 - \bar{Z}) - [h_{11}h_{32} \exp(\lambda_1 t) + h_{12}h_{41} \exp(\lambda_2 t)](F_0 - \bar{F})\}, \tag{38}$$

$$R - \bar{R} = \Delta^{-1}\{[h_{21}h_{42} \exp(\lambda_1 t) + h_{22}h_{31} \exp(\lambda_2 t)](Z_0 - \bar{Z}) - [h_{21}h_{32} \exp(\lambda_1 t) + h_{22}h_{41} \exp(\lambda_2 t)](F_0 - \bar{F})\}, \tag{39}$$

$$Z - \bar{Z} = \Delta^{-1}\{[h_{31}h_{42} \exp(\lambda_1 t) + h_{32}h_{31} \exp(\lambda_2 t)](Z_0 - \bar{Z}) - [h_{31}h_{32} \exp(\lambda_1 t) + h_{32}h_{41} \exp(\lambda_2 t)](F_0 - \bar{F})\}, \tag{40}$$

$$F - \bar{F} = \Delta^{-1}\{[h_{41}h_{42}\exp(\lambda_1 t) + h_{42}h_{31}\exp(\lambda_2 t)](Z_0 - \bar{Z}) - [h_{41}h_{32}\exp(\lambda_1 t) + h_{42}h_{41}\exp(\lambda_2 t)](F_0 - \bar{F})\}, \quad (41)$$

where $\Delta = h_{31}h_{42} - h_{32}h_{41}$.

III. Adjustment to a Productivity Shock

In this section, we analyze the properties of this solution, focusing specifically on the behavior of the premium and the real exchange rate. To begin, it is worth noting that the discussion that follows would not be altered in any important way if the source of the shock were an improvement in the (exogenous) terms of trade faced by the country[12] or simply some productivity-enhancing policy reform. Alternatively, one may prefer to substitute an endowment structure for the simple transformation function described in the previous section. In this case, the shock would simply be an increase in the endowment of the tradable good. Finally, as mentioned previously, we confine the shock to the tradables sector mainly to simplify the analysis of its effect on the equilibrium real exchange rate (equation (16)). It may also be noted that the model can be used to analyze a variety of other disturbances, but this is not necessary for the purpose at hand, which is simply to illustrate the potential problems that may arise in using the parallel market premium to draw inferences about real exchange rate misalignment.

Consider then the effect of a permanent favorable productivity shock. As far as the steady state is concerned, all variables ($\bar{b}$, $\bar{R}$, and $\bar{F}$) remain constant at their original steady-state values (given by equations (27), (29), and (30), respectively), except for expenditure, for which equation (28) implies that a positive (negative) shock causes the steady-state level of expenditure to rise (fall). From equation (16), therefore, the long-run value of the equilibrium real exchange rate decreases for a positive shock: there is an equilibrium real appreciation in the long run. Notice that the steady-state premium does not change in this model. This feature permits policymakers to feel confident about the equilibrium value of the premium after a shock—and thus about whether the observed premium is above or below its long-run value—without having to solve the model.

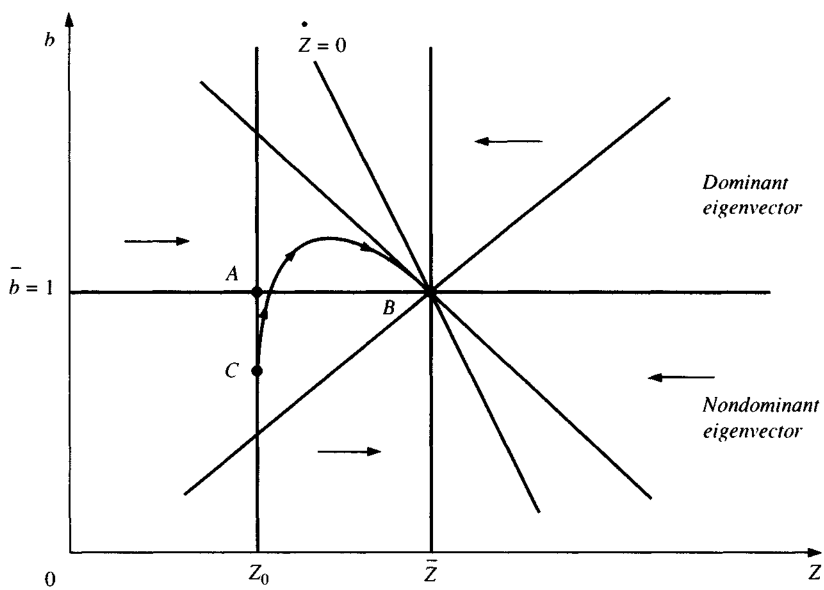
[12] In this case, we could think of the model as comprising an exportables and a nontradables sector on the supply side and importables and nontradables on the demand side. Also, it should be noted that the analysis of a *negative* shock would be completely symmetric to that presented below for a positive shock.

The issue, then, is whether this conveys information about the position of the real exchange rate relative to its long-run value.

To determine the out-of–steady state dynamics, it is useful to turn to Figure 1, which plots the level of expenditure Z on the horizontal axis and the level of (one plus) the parallel market premium b on the vertical axis. Imposing the steady-state condition $\dot{Z} = 0$ in equation (31) yields a negatively sloped locus in b-Z space, with the slope given by $-(1 - y_1 e')/H'$. Since the steady-state premium must be zero, the intersection of this locus with the horizontal line at $b = 1$ determines the steady-state value of expenditure. Z_0 denotes the initial steady-state level of expenditure while $\bar{Z}$ denotes the higher post–productivity shock level. The initial steady state is located at point A, while the new steady state is located at point B. The parallel market premium is zero ($b = 1$), in both steady states, while the position of the $\dot{Z} = 0$ locus is shifted to the right as higher spending is induced by the higher (post-shock) level of productivity. The arrows of motion relative to the new $\dot{Z} = 0$ locus in the figure follow from the fact that $Z_3 < 0$, as established below equation (31).

Consider now the dynamics of the adjustment of b and Z from A to B. Since expenditure is linked to the stock of money by the cash-in-advance constraint, and since the money stock can only change over time

Figure 1. *Dynamics of Consumption and the Premium*



through the process of hoarding, Z cannot jump on impact. This means that the instantaneous equilibrium of the economy must be somewhere along the perpendicular rising from point Z_0. To determine the paths followed by b and Z from an initial point on this perpendicular to the new steady state at B, we proceed in several steps. First, we derive slopes of the dominant and nondominant eigenvectors (that is, the eigenvectors associated with the smallest and largest of the negative eigenvalues, λ_1 and λ_2, respectively). These loci have the property that the paths of b and Z cannot cross them, since if the economy is ever on one of these vectors, it can be shown to remain on it as it converges to the new steady state. From equations (38) and (40), the slope of the dominant eigenvector is equal to h_{11}/h_{31}, while the slope of the nondominant eigenvector is equal to h_{12}/h_{32}. To determine the signs of the slopes, we use equation (35), which implies

$$\text{slope of dominant eigenvector} = (\lambda_1 - Z_3)/Z_1 > 0, \tag{42}$$

$$\text{slope of nondominant eigenvector} = (\lambda_2 - Z_3)/Z_1 < 0, \tag{43}$$

where the signs in equations (42) and (43) follow from evaluating the polynomial $P(\lambda)$ in equation (32) at $\lambda = Z_3$.[13] Unless the initial values of the state variables Z and F place the economy precisely on the dominant eigenvector, the economy will converge to the steady state at B along the nondominant eigenvector drawn in Figure 1, since the influence of the largest negative root must gradually vanish over time.

It remains only to determine the sign of the jump in b resulting from the productivity disturbance. The Appendix establishes that, in fact, the premium and therefore b must make a discrete jump downward on impact, to a point such as C in Figure 1. This means that the adjustment path must follow the dark arrowed path in the figure: expenditure rises throughout the adjustment, but the premium falls on impact and then rises temporarily before falling back toward zero in the final stages of the adjustment.[14]

Intuitively, the favorable productivity shock increases income, which

[13] Specifically from equation (32), it can be shown that as λ goes to $-\infty$ then lim $P(\lambda) > 0$ and $P(Z_3) < 0$. From these two facts, it follows that $\lambda_1 < Z_3 < \lambda_2 < 0$. The signs of the slopes in equations (42) and (43) follow directly.

[14] It is interesting to note that the qualitative path followed by the economy in response to a real shock (an improvement in productivity) is the same as that followed in response to a nominal shock (for example, an exogenous change in the stock of money). Figure 1 is again illustrative. Suppose the *initial* steady state is B in Figure 1 and that there is an exogenous decrease in the stock of money. Since the model exhibits neutrality in the steady state, the new long-run equilibrium will be identical to the initial equilibrium: the new steady state is also point

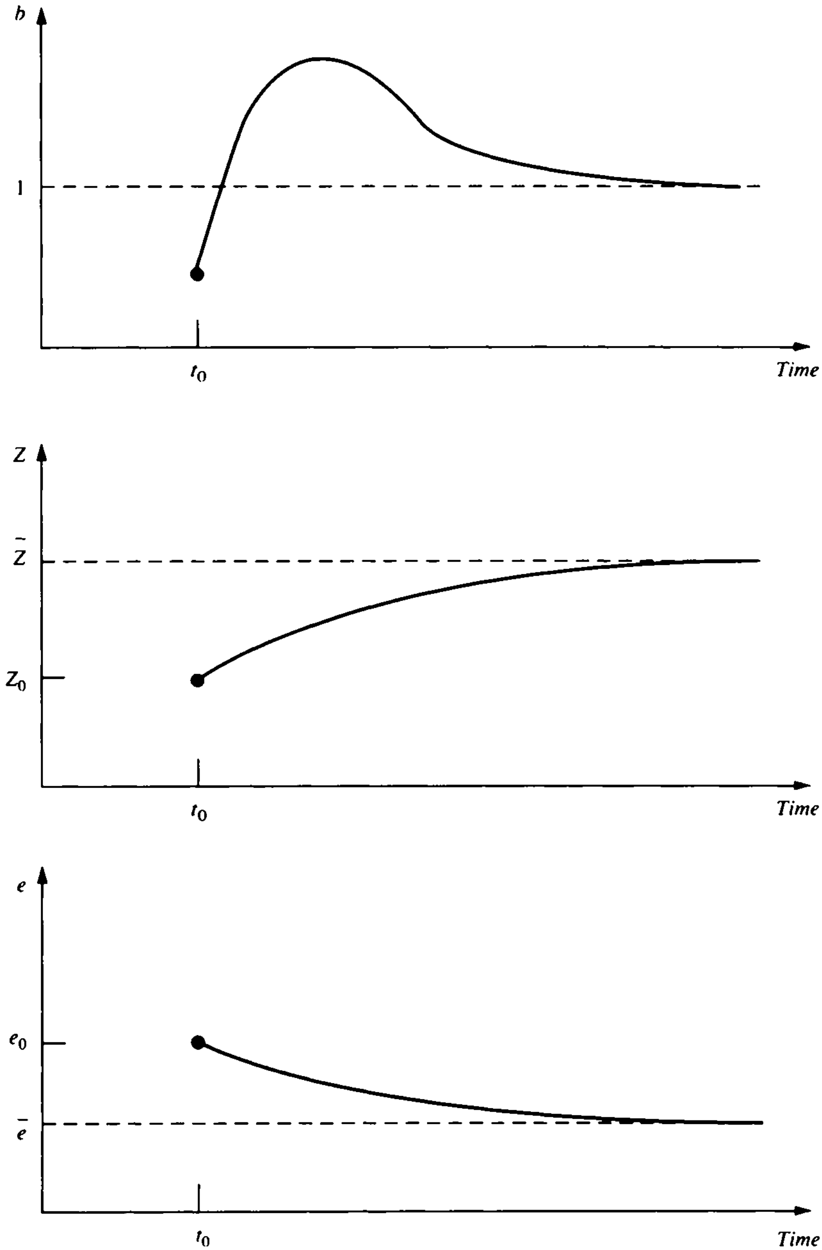
is consistent with a permanent increase in household expenditure. However, the adjustment in spending cannot be instantaneous, because an increase in spending requires an increase in the money stock, and the latter can only occur gradually through the process of hoarding. Initially, then, spending remains below permanent income, and the associated saving results in the gradual accumulation of money, permitting spending to rise over time. This accounts for the monotonic adjustment of Z. At the same time, the path of b cannot be monotonic because, if it were, the stock of foreign exchange could not remain unchanged from one steady state to the next, as required by equation (30). With monotonic adjustment in b, foreign exchange balances would either be drawn down or accumulated continuously.

The contrast between the behavior of b and Z along the adjustment path has important implications for the use of the premium as an indicator of real exchange rate misalignment. Figure 2 plots the time paths of the various variables resulting from the adjustment of the economy portrayed in the first figure. As can be seen, the monotonic adjustment of expenditure implies that the adjustment of the real exchange rate is also monotonic by equation (16), while the path of the parallel market premium necessarily overshoots. This means that, throughout the adjustment, the currency is undervalued in real terms (obviously if the shock had been negative, the problem would have been real exchange rate overvaluation). However, the parallel market premium is initially negative (foreign currency sells at a discount in the unofficial market), then rises to a positive value, before falling again toward zero. Thus, only the first part of the adjustment process coincides with the conventional view that undervaluation (overvaluation) should be associated with parallel market discounts (premia). Clearly, the parallel market premium cannot by itself provide reliable information about even the sign of real exchange rate misalignment at a given point in time, let alone its magnitude.[15]

B in Figure 1. On impact, though, a decrease in the nominal money stock implies a decrease in expenditure in terms of traded goods (that is, a decrease in Z), since the official exchange rate is fixed and the cash-in-advance constraint must hold. Thus, on impact, the economy must move to a point along the perpendicular rising from Z_0 in Figure 1. But then, as shown in the Appendix, the dynamic path followed by the economy must be as drawn in Figure 1—in particular, the premium must jump down on impact. Thus, the economy begins at point B, jumps to point C, and then follows the arrowed curve back to point B. Clearly, therefore, a nominal shock is capable of generating the same type of nonmonotonicity in the adjustment of the premium as the productivity shock does.

[15] An alternative to looking at the premium at a point in time is to look at its *average* behavior over some period of time. Whether this yields a more reliable indicator depends on how long the economy spends in each phase of the adjustment path and on how long after the shock an assessment is made.

Figure 2. *Adjustment to a Favorable Productivity Shock*



IV. Conclusion

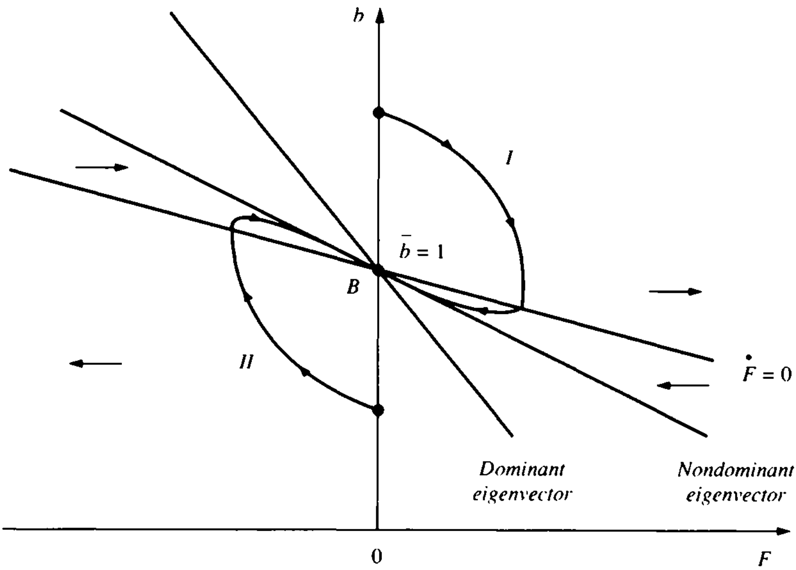
Because most developing countries maintain a managed official exchange rate, policymakers in such countries commonly have to make judgments about the extent of real exchange rate misalignment. The degree of misalignment is very difficult to measure, since the equilibrium real exchange rate is unobservable and depends on a range of structural and macroeconomic factors. Thus, economists looking at this issue have frequently used the parallel market premium—about whose equilibrium value they have more confidence—to draw inferences regarding the extent of real exchange rate misalignment.

The conventional wisdom is that there is a robust correlation between the extent of overvaluation or undervaluation of a currency and the parallel market premium, a view that appeals to intuition, as well as having both some analytical and empirical support. Against this view, however, one must set the observation that, since the degree of misalignment and the parallel market premium are both endogenous real variables, both the nature of shocks and the economic structure are likely to play a role in determining the sign and magnitude of co-movements between these variables. Moreover, available empirical evidence suggests that the premium is far more variable than any likely deviations of actual exchange rates from their equilibrium values, reaching several hundred percent with no ex post evidence of comparable magnitudes of real exchange rate misalignment.

This paper investigated, using a fully optimizing model of a small open developing country, how the parallel market premium and the real exchange rate jointly respond to an illustrative shock, taken here to be a permanent productivity disturbance. The analysis suggested that the informational content of the premium may be limited, since in response to a shock the premium was both positive and negative at various times along the adjustment path, while the degree of overvaluation of the currency was always positive for a negative shock and always negative for a positive shock.

The obvious policy implication of the analysis is that by itself the premium is an unreliable indicator of the sign and magnitude of real exchange rate misalignment. The premium may provide useful information about the relationship of the real exchange rate to its equilibrium level—for example, the magnitude of the initial jump in the premium will be correlated with that of the initial extent of real exchange rate misalignment—but it is dangerous to draw inferences about deviations of the actual exchange rate from the equilibrium real exchange rate based on observations of the premium at a given moment in time. In the model in

Figure 3. *Dynamics of Foreign Assets and the Premium*



this paper, other variables, such as the behavior of the trade balance, would provide more reliable information on the extent of real exchange rate misalignment.

APPENDIX

The purpose of this appendix is to show that in response to a positive productivity shock, the premium must jump down on impact and that the adjustment of the economy to such a shock must be as presented in Figure 1. Figure 3 plots the stock of foreign securities F on the horizontal axis and the parallel market premium (plus unity) b on the vertical axis. The slope of the $\dot{F} = 0$ locus (which is drawn relative to the new post-shock steady-state equilibrium at B) is equal to $-F_4/F_1$, which is negative (equation (31)).[16] From equations (38) and (41), we can derive the slopes of the dominant and nondominant eigenvectors:

$$\text{slope of dominant eigenvector} = (\lambda_1 - F_4)/F_1 < 0, \tag{A1}$$

$$\text{slope of nondominant eigenvector} = (\lambda_2 - F_4)/F_1 < 0, \tag{A2}$$

where, by the fact that $\lambda_1 < \lambda_2 < 0$, the dominant eigenvector is more steeply sloped than the nondominant eigenvector, and both are more steeply sloped than

[16]Because the new and old steady-state values of b and F are identical, the new and old steady states in Figure 3 coincide and are given by point B.

the $\dot{F} = 0$ locus, as drawn in Figure 3. As mentioned previously, unless the initial steady-state values of the two state variables place the economy on the dominant eigenvector, the path followed by the economy in Figure 3 eventually converges to the slope of the nondominant eigenvector.

The question, as in Figure 1, is where the economy jumps on impact along the line perpendicular to point $F = 0$, which is the new (and old) steady-state value of F. There are two possibilities, labeled I and II in Figure 3. Under path I, the premium jumps up on impact, and, given the arrows of motion drawn relative to the $\dot{F} = 0$ locus, the premium must approach its steady-state level from below. Under path II, the premium jumps down on impact and, given the arrows of motion, must approach the steady state at B from above.

It may be recalled from Figure 1 that whether b jumps up or down on impact it must approach the steady state from above. But this rules out path I in Figure 3 since, in this case, b approaches the steady state from below. Therefore, b must jump down on impact, and the economy must follow the path indicated in Figure 1 (with the dynamics of the individual variables as drawn in Figure 2) and likewise path II in Figure 3.

REFERENCES

Aghevli, Bijan B., Mohsin S. Khan, and Peter J. Montiel, *Exchange Rate Policy in Developing Countries: Some Analytical Issues*, Occasional Paper No. 78 (Washington: International Monetary Fund, 1991).

Bhandari, Jagdeep S., and Carlos A. Végh, "Dual Exchange Markets Under Incomplete Separation: An Optimizing Model," *Staff Papers*, International Monetary Fund, Vol. 37 (March 1990), pp. 146–67.

Calvo, Guillermo A., and Carlos A. Rodríguez, "A Model of Exchange Rate Determination Under Currency Substitution and Rational Expectations," *Journal of Political Economy*, Vol. 85 (1977), pp. 617–25.

Edwards, Sebastian, *Real Exchange Rates, Devaluation, and Adjustment: Exchange Rate Policy in the Developing Countries* (Cambridge, Mass.: MIT Press, 1989).

Kamin, Steven B., "Devaluation, Exchange Controls, and Black Markets for Foreign Exchange in Developing Countries," *Journal of Development Economics*, Vol. 40 (1993), pp. 151–69.

Khan, Mohsin S., and Jonathan D. Ostry, "Response of the Equilibrium Real Exchange Rate to Real Disturbances in Developing Countries," *World Development*, Vol. 20 (1992), pp. 1325–34.

Kharas, Homi, and Brian Pinto, "Exchange Rate Rules, Black Market Premia, and Fiscal Deficits: The Bolivian Hyperinflation," *Review of Economic Studies*, Vol. 56 (1989), pp. 435–48.

Lizondo, J. Saul, and Peter J. Montiel, "Fiscal Policy and the Dynamics of Devaluation for a Small Country with Optimizing Agents" (unpublished; Washington: International Monetary Fund, 1991).

Quirk, Peter J., and others, *Floating Exchange Rates in Developing Countries: Experience with Auction and Interbank Markets*, Occasional Paper No. 53 (Washington: International Monetary Fund, 1987).

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

Exchange Rate Strategies and Fiscal Performance in Sub-Saharan Africa

KARIM NASHASHIBI and STEFANIA BAZZONI*

This paper investigates the relationship between fiscal performance and movements in the exchange rates, the terms of trade, and other macroeconomic aggregates in 28 sub-Saharan African countries over the 1980–91 period. It finds that the tax base in most of these countries is heavily dependent on imports and import substitutes. Consequently, an overvaluation of the exchange rate in countries that adopt a fixed exchange rate strategy (CFA franc zone countries), when the purpose of the strategy is to restore the real exchange rate to its equilibrium through fiscal contraction, can result in a widening of the fiscal deficit. Those countries with a variable exchange rate strategy fail to attain price stability but are able to enhance their fiscal balance, competitiveness, and growth. [JEL E31, E42, E6, F3, H2, H5, H6]

SUB-SAHARAN AFRICAN countries emerged from the 1970s with large and unsustainable fiscal deficits, stemming from the increase in government spending following the two oil shocks. The commodity price boom, associated with these shocks, increased government revenues and fueled government spending in a broad range of activities, including investment in public enterprises and marketing boards. Because such

* Karim Nashashibi is Assistant Director of the Middle East Department. He holds a Ph.D. from the University of California at Berkeley. Stefania Bazzoni, a doctoral candidate at the Institut Universitaire de Hautes Études Internationales (Geneva), contributed to this paper as a summer intern in the Fiscal Affairs Department in 1992. The paper also benefited from the comments of Tamim Bayoumi, Carlo Cottarelli, Pierre Dhonte, Christian François, Alejandro García, Robert Hagemann, Daniel Hewitt, Malcolm Knight, and Vito Tanzi. Tarja Papavassiliou and Eric Sidgwick provided valuable research support, and Derek Bills was helpful in generating the charts.

spending entailed a large expansion in government employment and other recurrent expenditures that are difficult to reverse, it resulted in a structural imbalance between revenue and expenditure, which has persisted in most African countries throughout the 1980s. This internal imbalance was exacerbated by a sharp deterioration in the external environment, characterized by a protracted decline in terms of trade, an increase in real interest rates, and a sharp decline in the availability of external bank credit.

To alleviate both the internal and external imbalances, sub-Saharan African countries pursued two different strategies. Countries in the western and central African monetary unions (the CFA franc zone, in which CFA stands for Communauté Financière Africaine), retained the "internal" adjustment path predicated on maintaining a fixed exchange rate parity with the French franc to ensure fiscal discipline and low inflation.[1] Most other sub-Saharan countries addressed the decline in the terms of trade using a variable exchange rate strategy.[2] This "external" adjustment strategy allowed the current account to adjust directly, although at the price of substantially higher inflation.

With a pegged exchange rate, a real devaluation can be achieved through deflationary policies, provided price and wage flexibility prevail. On the other hand, when the tax base is highly dependent on international trade, an overvaluation of the real exchange rate would tend to undermine tax revenue and the attainment of fiscal balance (Tanzi (1989)). With a floating or a managed exchange rate, it could be argued that the inflation generated by the lack of fiscal and monetary discipline would also undermine government revenue—owing to lags in tax collection and adjustments of public utility prices—causing growing fiscal deficits, an unfavorable investment environment, and slow growth (Tanzi (1977)).

This paper argues that for economies in which the tax base is highly dependent on imports and import substitutes, an exchange rate that is adjusted toward its equilibrium level is a critical element in improving fiscal performance. To the extent that an overvaluation of the exchange rate undermines the tax base, internal adjustment may result in a widening of the fiscal deficit, when its very purpose is to restore the real exchange rate to its equilibrium level through fiscal contraction. Hence,

[1] The 11 fixed-rate countries in the sample are all members of the CFA franc zone: Benin, Burkina Faso, Cameroon, Central African Republic, Congo, Côte d'Ivoire, Gabon, Mali, Niger, Senegal, and Togo.

[2] The 17 variable-rate countries are Botswana, Burundi, The Gambia, Ghana, Kenya, Madagascar, Malawi, Mauritania, Mauritius, Nigeria, Rwanda, Sierra Leone, Tanzania, Uganda, Zaire, Zambia, and Zimbabwe.

presenting the internal and external adjustment strategies as policy alternatives is not meaningful. Rather, these two strategies should be complementary.

Having established a fiscal data base for 28 sub-Saharan countries for the 1980–91 period, the paper assesses the implementation of the two alternative strategies by comparing the fiscal performance of the two groups of countries during the 1980–91 period.

I. Initial Macroeconomic Setting

Government participation and intervention in sub-Saharan African economies have been substantial since independence, reflecting the prevailing view that the investment necessary for development and long-term growth would not be forthcoming through private investment alone.

The 1973 oil shock, associated with a boom in other commodity prices, resulted in a structural shift in the public expenditure pattern in sub-Saharan Africa, seemingly with little concern about the fiscal consequences of a possible sharp decline in future commodity prices. Total public expenditure in sub-Saharan Africa grew from 19 percent of GDP in 1970 to 29 percent of GDP in 1980 (Schelizi, Ghandi, and Ehdaie (1985)). This upward trend was common to virtually the entire region with the exception of Tanzania and Burkina Faso. Moreover, the capital formation component of government expenditures, which was around 5 percent of GDP, expanded very rapidly after 1974; by 1980, it reached 9.3 percent of GDP.[3] Beyond a noteworthy improvement in infrastructure, public investment focused on capital-intensive manufacturing aimed at import substitution and sheltered from outside competition by protectionist policies. Much of this investment resulted in large losses for public enterprises and banks and a fall in domestic savings. Other public investment in such areas as health, education, and social infrastructure were accompanied by an increase in public employment, which introduced a secular upward shift in recurrent government spending that has proved difficult to reverse.

On the revenue side, sub-Saharan African countries raised their tax ratios mostly through an expansion in the volume of imports. By the late 1970s, tax revenues were averaging around 17 percent of GDP, and total government revenue was comparable with that of other developing countries (Tanzi (1981)). Despite the improvement in resource mobilization,

[3] In terms of GDP, the average investment rate in the 1970s was 20.6 percent in sub-Saharan Africa and 21.6 percent in south Asia.

tax revenue was reaching certain limits imposed by the tax structure, in particular by the high dependence on taxes on international trade.[4] Yet attempting to expand the tax base to domestic transactions, without considerable strengthening of tax administration, was difficult to achieve in the short term.

The expansion of government spending, coupled with difficulties in raising additional revenue, generated fiscal imbalances, which were financed primarily by external borrowing at a time when banks were seeking to recycle the oil money and when "sovereign risk" was perceived as minimal. By 1980, the average government deficit, excluding grants, in our sample of 28 countries was 9 percent of GDP, and external debt was ten times higher than in 1970, reaching 35 percent of GDP. Thus, despite the sharp increase in commodity prices in the 1970s, large fiscal imbalances had already appeared as a result of excessive government spending. These imbalances—coupled with the large accumulation of external debt—were inconsistent with the stable macroeconomic environment needed to encourage a strong private sector and sustained growth.

Hence, prior to the deterioration in the terms of trade, sub-Saharan African countries entered the 1980s with a need for lower fiscal deficits in order to correct internal and external imbalances and reduce the crowding out of the private sector. There was also a need to improve public finance management so as to reduce nonproductive expenditures and alleviate the distortions introduced through trade restrictions and high customs tariffs aimed at protecting the manufacturing base.

Initial Imbalances

During the 1980s, different macroeconomic performances between the two groups of countries may be explained by several factors: a different starting point, a different intensity in the external shocks sustained, and a different adjustment strategy. Although both groups had a tendency to overspend during the 1970s, the fixed-rate countries started the decade with lower budget deficit-to-GDP ratios (excluding grants) than did the variable-rate countries (7.6 percent of GDP versus 10.4 percent of GDP). The external current account deficit was also more favorable for the

[4] This dependence is defined as taxes on imports and exports, nontax revenue on imports and exports, plus excise taxes, sales taxes, and fees levied on imports as a percentage of total tax receipts.

fixed-rate countries (6 percent of GDP) than for the variable-rate countries (10 percent of GDP).

On the other hand, the impact of the increase in the real interest rate and the structure of external debt in the two groups were roughly similar. Between 1980 and 1983, the average ratio of total external debt to GDP was 45 percent for the fixed-rate countries and about 47 percent for the variable-rate countries. The proportion of nonconcessional debt in total outstanding debt in 1980 was 22 percent and 24 percent, respectively, for the fixed- and variable-rate countries.

These figures suggest that, on average, the sharp increase in real interest rates at the beginning of the 1980s affected the two groups of countries in a similar fashion. However, between 1985 and 1991, fixed-rate countries experienced a much sharped deterioration (31 percent) in the terms of trade than variable-rate countries (17 percent).[5] This difference comes primarily from export price developments, because movements in import unit values for the two groups of countries increased in a similar fashion. Moreover, this deterioration was more broadly based in the fixed-rate group, where only one country—Senegal—experienced an improvement in its terms of trade since 1985; in the variable-rate group, 5 countries (out of 17) experienced such an improvement (Figure 1).

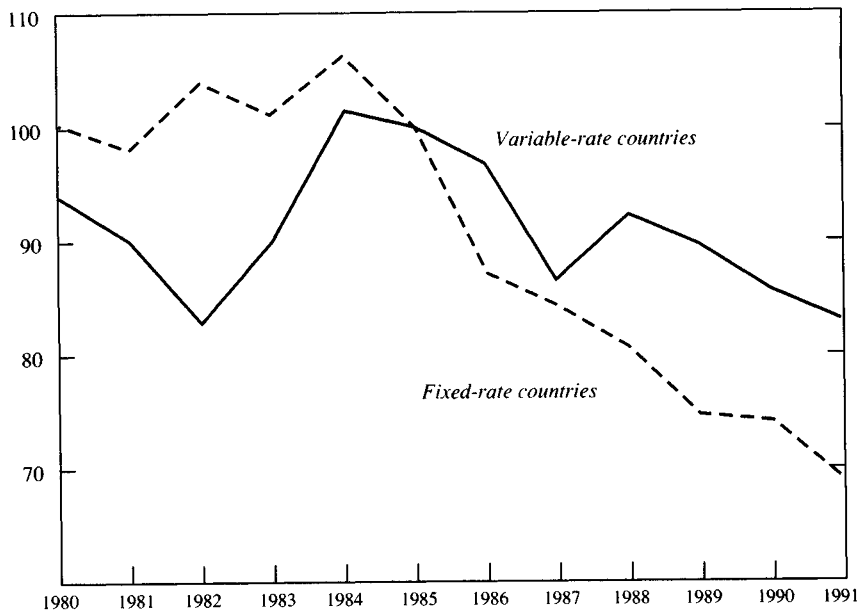
To summarize, although the fixed-rate countries started the 1980s with much more favorable fiscal and current account positions, they have experienced a sharper decline in terms of trade since the mid-1980s than the variable-rate countries. These factors, coupled with fairly close levels of concessional aid, relative to GDP, would suggest that the overall need for adjustment in the two groups of countries was similar. Thus, an explanation of the divergence in the fiscal (and other macroeconomic) performances of the two groups lies more with policy strategy and its implementation than with differences in initial imbalances or external shocks.

Two Alternative Strategies

An adjustment strategy predicated on a nominal exchange rate anchor has been followed by the 13 members of the West and Central African Monetary unions. Until very recently, their common currency—the CFA franc—has been pegged to the French franc at the rate of CFAF 50 =

[5] If terms of trade are weighted by each country's share of trade, the figures are 23 percent for the variable-rate countries and 37 percent for the fixed-rate countries.

Figure 1.
Sub-Saharan Africa: Terms of Trade, 1980–91
(1985=100)



Sources: World Economic Outlook data base and *International Financial Statistics*.

FF 1.[6] For about three decades the monetary unions and the peg to the French franc served these countries relatively well. Convertibility of their currency, guaranteed by the French Treasury, and mobility of capital throughout the zone, helped provide these countries with monetary stability, low inflation, and an environment conducive to investment and growth (Boughton (1991)).

One of the main benefits of a fixed exchange rate regime for a small, open economy is that it provides a convenient nominal anchor around which a consistent set of macroeconomic policies can be formulated. However, such an anchor can only hold if both fiscal and monetary discipline are achieved. Such discipline was intended by the institutional arrangements of the two monetary unions, which provided for the inde-

[6] On January 12, 1994, the CFA franc was devalued from FF 1 = 50 FCFA to FF 1 = 100 FCFA.

pendence of the two central banks and which limited the extension of credit to member country governments to the equivalent of 20 percent of the previous year's fiscal revenue.

To adhere to a common external standard, members of the zone give up the use of the exchange rate as a policy instrument, a cost that depends on the nature of the external shocks. In the event of a temporary shock, this exchange regime can provide sufficient flexibility through the use of external reserves for efficient stabilization. However, a more protracted external shock, such as the deterioration of the terms of trade since 1979, may require a change in relative prices. If a nominal exchange rate adjustment is ruled out, a real exchange rate depreciation has to be brought about through a reduction in nominal prices and wages.[7]

But, if some degree of wage rigidity exists and no strong productivity gains are made in favor of the traded goods sector, a substantial loss of output and unemployment would be incurred. What is interesting in the fixed-rate countries is that the nominal exchange rate anchor is not determined using a basket of currencies reflecting major export destinations or sources of imports but is pegged to the French franc, which is governed by French macroeconomic objectives, including for the past several years the maintenance of a parity with the German mark.[8] Therefore, since the mid-1980s, the fixed-rate countries have suffered two external shocks: a major terms of trade deterioration and an appreciation of the French franc relative to the currencies of the fixed-rate countries' trading partners, which has been translated into a real appreciation of the CFA franc.

To summarize, saddled with structural fiscal deficits from the outset of the 1980s, the fixed-rate countries have had a single instrument—fiscal policy—to achieve two objectives: (i) to bring down the fiscal deficit to a level that is consistent with the targeted inflation rate set by the two central banks, and (ii) to engineer a change in the relative prices of traded and nontraded goods in response to the decline in terms of trade, so as to restore external competitiveness. Theoretically, a reduction of the fiscal deficit would also reduce the real value of the CFA franc.

[7] Another adjustment instrument that has been used by some sub-Saharan African countries is commercial policy. Theoretically, a real devaluation can be approximated through a uniform surcharge on imports and a uniform subsidy on exports (see Devarajan and De Melo (1987)). However, this approach is difficult to implement, partly because of the need to impose subsidies on the value added rather than on the gross value of exports and partly because of substantial unrecorded border trade from non-CFA countries, particularly Nigeria.

[8] Although trade with France amounts to a sizable share of CFA franc zone imports and exports, prices of major exports (coffee, cocoa) and imports (petroleum products) are denominated in U.S. dollars.

In contrast to the fixed-rate countries, the 17 variable-rate countries included in this study are a more heterogeneous group. Nigeria and Zaire, the largest countries in the group, have been pursuing expansionary fiscal and monetary policies resulting in high inflation, albeit with occasional attempts at stabilization. Ghana, Madagascar, Tanzania, Uganda, and Zambia have undergone successive adjustment programs and, with the assistance of the IMF and World Bank, have adopted structural reforms to move away from heavy government interference in the economy. Other countries, such as The Gambia, Malawi, and Zimbabwe, have established a relatively liberal environment stemming from a strong private sector. All the variable-rate countries have allowed the exchange rate to seek an equilibrium, but a number of countries only gradually liberalized their exchange and trade policies during the 1980s.[9]

While this group of countries used restrictive fiscal and monetary policies to reduce internal imbalances, the fiscal and monetary discipline instilled by an independent central bank was lacking, resulting in relatively high inflation and frequent currency depreciation.

II. Fiscal Performance in the 1980s

Aggregate data on fiscal performance in sub-Saharan countries suggest that little progress was made in reducing fiscal imbalance over the decade. When external grants are excluded, the average deficit remains virtually constant throughout the 1980s at about 8 percent of GDP. Further scrutiny reveals, however, that the aggregate data average out the contrasting performance of the two groups of countries.

Fiscal Deficits

The overall budget deficit (excluding grants) for the fixed-rate countries in 1980–81 averaged 7.6 percent of GDP, or substantially less than the deficit levels for the variable-rate countries (10.4 percent).[10] By the end of the decade, the fiscal imbalance in the first group widened to 8.2 percent, while the second group reduced its fiscal imbalance to 6.2 percent of GDP (Table 1 and Figure 2). Most of the deterioration for the fixed-rate countries occurred between 1985 and 1989; moreover, most of

[9] For instance, The Gambia did not drop the peg to the pound sterling until 1986.

[10] Simple annual averages across each group have been derived for the fiscal aggregates.

Table 1. *Sub-Saharan Africa: Summary of Fiscal Performance*
(In percentage of GDP)

	1980–81	1985–86	1990–91	Change from 1980–81 to 1990–91
Overall deficit, excluding grants				
Entire region	−9.4	−7.8	−8.3	1.1
Fixed-rate countries	−7.6	−8.0	−8.2	−0.6
Fixed-rate countries, except Cameroon, Congo, and Gabon	−10.6	−8.9	−8.9	1.7
Variable-rate countries	−10.4	−6.5	−6.5	3.9
Variable-rate countries, except Botswana and Nigeria	−11.0	−8.3	−7.4	3.6
Total revenue				
Entire region	18.9	18.9	18.9	0.0
Fixed-rate countries	20.3	21.4	17.1	−3.2
Fixed-rate countries, except Cameroon, Congo, and Gabon	17.4	18.3	15.3	−2.1
Variable-rate countries	17.9	19.5	20.2	2.4
Variable-rate countries, except Botswana and Nigeria	16.5	17.7	18.5	1.9
Total expenditure				
Entire region	28.6	28.1	27.0	−1.6
Fixed-rate countries	27.9	29.4	25.3	−2.6
Fixed-rate countries, except Cameroon, Congo, and Gabon	27.9	27.2	24.1	−3.8
Variable-rate countries	28.2	26.0	26.7	−1.5
Variable-rate countries, except Botswana and Nigeria	27.6	26.1	25.8	−1.7

Source: National authorities and IMF staff estimates.

the improvement in the variable-rate countries became more difficult to sustain after 1985. Nevertheless, there was some fiscal retrenchment for both groups during the 1989–91 period.

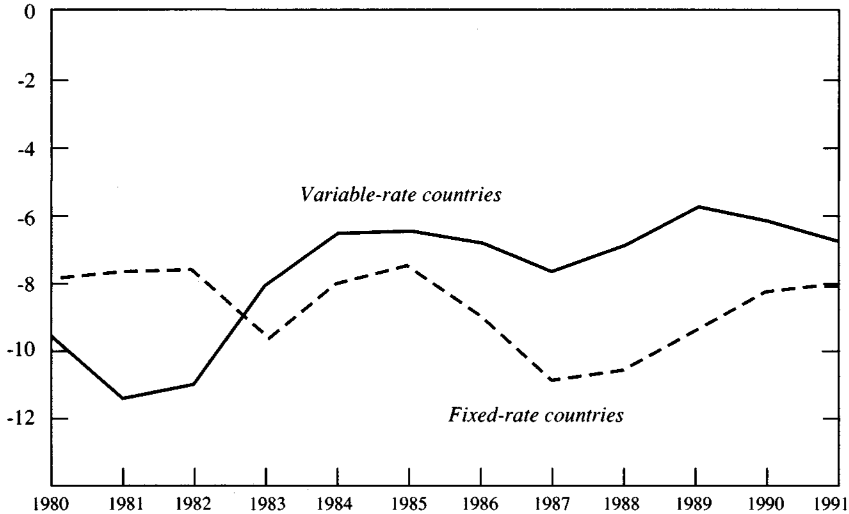
When interest costs are excluded and the focus is shifted to the deficit measure, which can be directly affected by domestic policy, the difference in performance becomes even more pronounced. Interest costs have been roughly the same for the two groups of countries, doubling from 2 percent of GDP to 4 percent of GDP over the decade. In the fixed-rate countries, there was only a small reduction in the primary deficit—from 5.8 percent of GDP to 4.3 percent of GDP. Only two, Gabon and Senegal, succeeded in generating a primary surplus by 1991, although in the case of Gabon it was mostly the result of higher oil exports and prices. In contrast, the variable-rate countries have experienced a more pronounced and sustained improvement in their primary balance, with the deficit falling from 9 percent of GDP in 1980–81 to 1.3 percent of GDP in 1990–91 (Figure 3).

Looking at the current fiscal position (defined as the difference between total current revenue excluding grants and current expenditure) sheds further light on the government's contribution to domestic savings.[11] Evidence suggests that after some improvements during the first half of the decade, owing to a depreciation in the real exchange rate and an upturn in commodity prices, fixed-rate countries have not been able to generate public sector savings during the second half of the decade or contribute additional resources to the development of the private sector by reducing the government's indebtedness to the banking system. The current fiscal balance was in substantial surplus in 1980–81 (3.5 percent of GDP) but declined sharply after 1985 and turned negative in 1987 (Figure 4). The turnaround in the current fiscal balance over the decade (about 4.5 percentage points) is much sharper than the deterioration in the overall fiscal balance, pointing to the major causes of fiscal imbalance for the fixed-rate countries—recurrent revenue and expenditure rather than expanding investment. Nevertheless, because of the improvement in their fiscal performance in 1990 and 1991, Burkina Faso, Gabon, Mali, and Senegal managed to generate a current account surplus.

In contrast to fixed-rate performance, the current balance was negative in variable-rate countries at the beginning of the 1980s but improved steadily after 1982, generating a surplus by the end of the decade for about half of the countries in that group. On the whole, the improvement

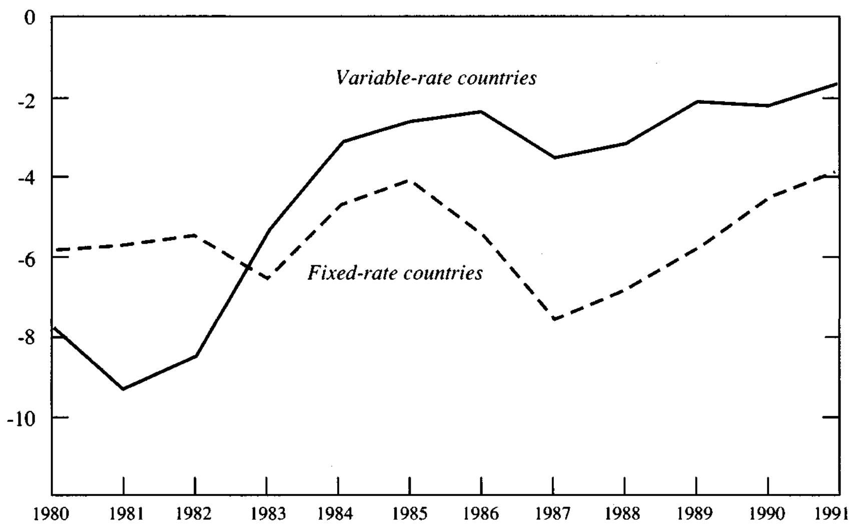
[11] The main limitation of this concept is the implicit assumption that all investment is good and all current expenditure (including health, education, operations, and maintenance) is consumption and does not contribute to growth.

Figure 2.
Sub-Saharan Africa: Overall Fiscal Deficit, Excluding Grants, 1980–91
(As percent of GDP)



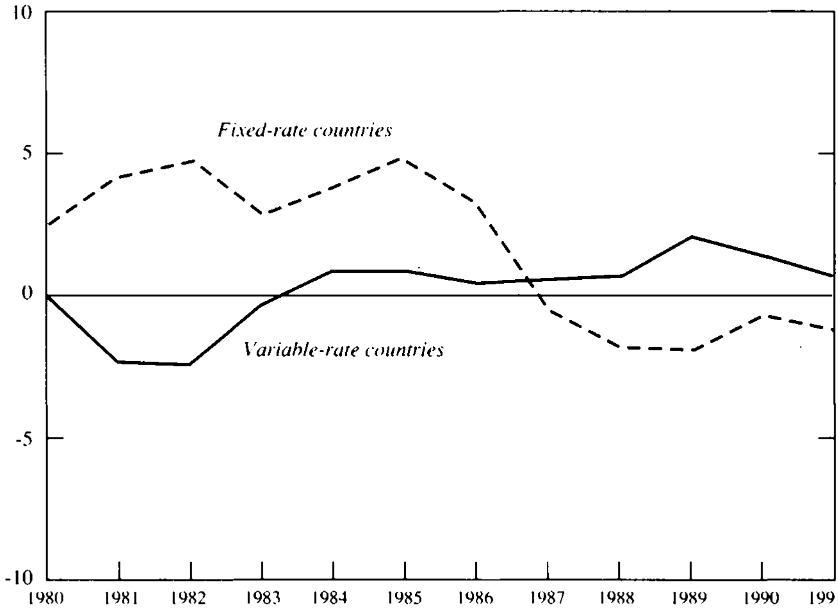
Sources: National authorities and IMF staff estimates.

Figure 3.
Sub-Saharan Africa: Primary Balance, Excluding Grants, 1980–91
(As percent of GDP)



Sources: National authorities and IMF staff estimates.

Figure 4.
Sub-Saharan Africa: Current Balance, Excluding Grants, 1980–91
(As percent of GDP)

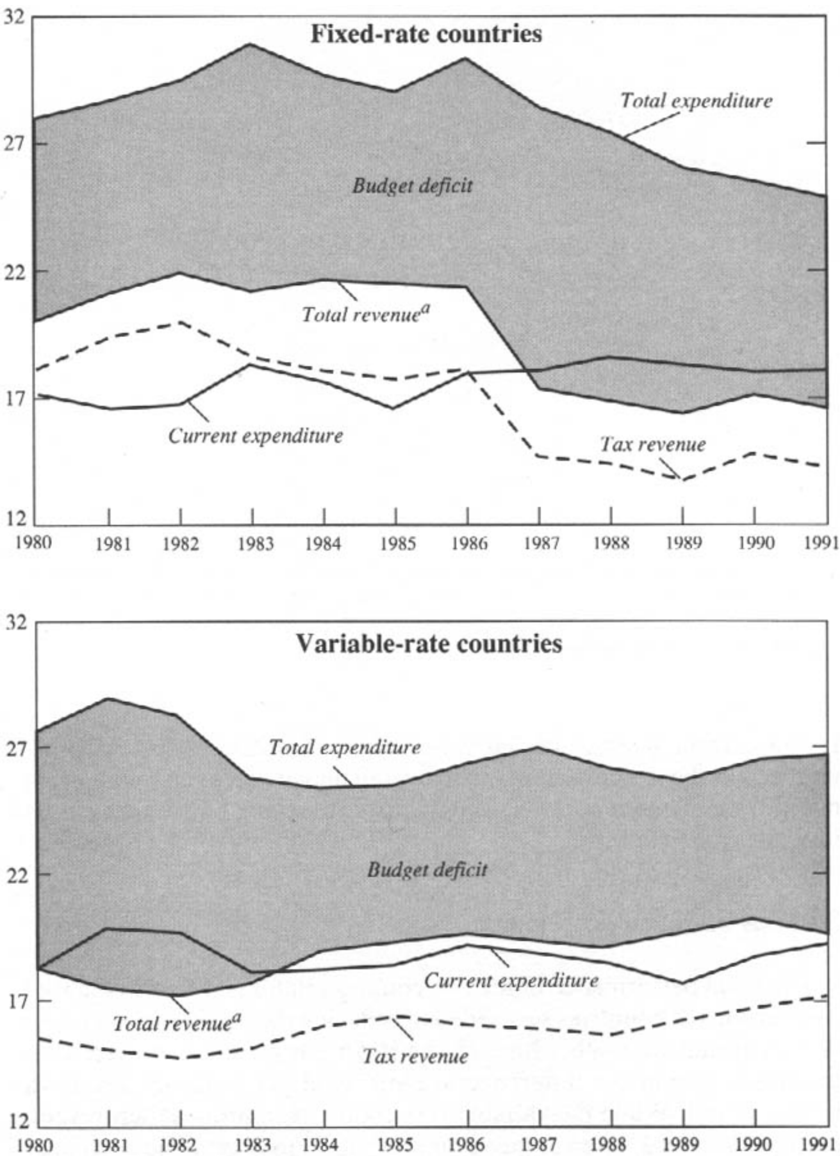


Sources: National authorities and IMF estimates.

in the current balance of the variable-rate countries was not as pronounced as the improvement in the overall deficit. This suggests a broad-based retrenchment in public expenditures touching both recurrent and investment expenditure.

Sources of Fiscal Adjustment

The fiscal performance of the two country groups is shown in Figure 5. Both groups of countries succeeded in reducing their government expenditures during the 1980s, although the effort was greater in the fixed-rate countries. The major difference in their fiscal performance lies on the revenue front. While the variable-rate countries improved their revenue performance by 2.4 percentage points of GDP, total revenue declined in the fixed-rate countries by 3.2 percentage points of GDP, with the result that, by the end of the 1980s, their revenue ratio stood well below that



Sources: National authorities and IMF staff estimates.
[a] Excluding external grants.

of the variable-rate countries. The differences in revenue performance were mostly due to tax revenue,[12] which declined in fixed-rate countries by 21 percent (from 18.7 percent of GDP to 14.7 percent of GDP) and increased in the variable-rate countries by 11 percent (from 15.4 percent of GDP to 17.1 percent of GDP).

A breakdown of the various tax categories for both groups of countries is revealing (Table 2). The largest and most broad-based decline in tax revenue in the fixed-rate countries stems from a fall in the taxes on international trade by 2 percentage points of GDP, or a 40 percent decline from its level in 1980–81 (Figure 6). In the variable-rate countries, this tax fluctuated between 5 percent and 6 percent of GDP during the decade, with virtually no change by the end of the decade (5.3 percent of GDP).[13]

Taxes on goods and services in fixed-rate countries have declined throughout the decade (Figure 7). This decline occurred despite concerted efforts at expanding the taxation of domestic transactions by introducing the value-added tax (VAT) and broadening the sales taxes to include services and the distribution sector. Interestingly, Côte d'Ivoire and Cameroon made gains in this tax category, despite the overall downward trend, by broadening their tax bases. In Côte d'Ivoire, taxes on goods and services were maintained at 6.5 percent of GDP over the entire 1987–91 period, despite declines in GDP and per capita income. In the variable-rate countries, this is the tax category in which the most progress was achieved, with the tax ratio rising from 3.8 percent of GDP in 1980 to 5.4 percent of GDP in 1991. The main reason for this progress is the shift of some taxation of international trade to taxation of domestic sales, although the main base of indirect taxation remains imports. When the two tax categories are taken together, the fixed-rate countries have lost 2.3 percentage points of GDP, while the variable-rate countries have gained 0.8 percentage point.

Income taxes declined in fixed-rate countries by 2.2 percentage points of GDP over the decade, mostly reflecting lower taxes from oil-exporting enterprises following the reverse oil shock in 1986–87. If these oil exporters are excluded, the decline is much lower (0.9 percent of GDP). Nevertheless, at 3.4 percent of GDP in 1990–91, this ratio is particularly

[12] Nontax revenue, which accounts for about 2 percent of GDP in both country groups, hardly changed in the variable-rate countries and increased by 0.5 percent of GDP in fixed-rate countries.

[13] These results, as well as those pertaining to taxes on goods and services, should be interpreted cautiously since tax reforms have tended to lower tariffs and shift taxation to domestic transactions. However, this pattern of tax reform has affected both groups of countries in a similar fashion.

Table 2. *Sub-Saharan Africa: Tax Revenue Performance*
(In percentage of GDP)

	1980–81	1985–86	1990–91	Change from 1980–81 to 1990–91
Total revenue				
Fixed-rate countries	20.3	21.4	17.1	−3.2
Fixed-rate countries, except Cameroon, Congo, and Gabon	17.4	18.3	15.3	−2.1
Variable-rate countries	17.9	19.5	20.2	2.4
Variable-rate countries, except Botswana and Nigeria	16.5	17.7	18.5	1.9
Tax revenue				
Fixed-rate countries	18.7	18.1	14.7	−4.0
Fixed-rate countries, except Cameroon, Congo, and Gabon	15.8	14.1	12.7	−3.1
Variable-rate countries	15.4	16.4	17.1	1.7
Variable-rate countries, except Botswana and Nigeria	14.4	16.1	16.4	2.1
Taxes on international trade				
Fixed-rate countries	7.0	6.1	5.1	−2.0
Fixed-rate countries, except Cameroon, Congo, and Gabon	7.6	6.6	5.6	−2.0
Variable-rate countries	5.3	5.7	5.3	0.0
Variable-rate countries, except Botswana and Nigeria	4.8	5.9	5.2	0.4

Taxes on goods and services				
Fixed-rate countries	3.1	2.8	2.8	−0.3
Fixed-rate countries, except Cameroon, Congo, and Gabon	3.4	2.9	2.7	−0.7
Variable-rate countries	4.1	4.2	5.1	1.0
Variable-rate countries, except Botswana and Nigeria	4.6	4.6	5.8	1.2
Income taxes				
Fixed-rate countries	8.1	8.5	5.9	−2.2
Fixed-rate countries, except Cameroon, Congo, and Gabon	4.3	4.0	3.4	−0.9
Variable-rate countries	5.9	5.7	6.1	0.2
Variable-rate countries, except Botswana and Nigeria	4.8	4.6	4.8	0.0
Other taxes				
Fixed-rate countries	0.4	1.2	1.1	0.7
Fixed-rate countries, except Cameroon, Congo, and Gabon	0.4	1.2	1.2	0.8
Variable-rate countries	0.4	0.7	0.7	0.3
Variable-rate countries, except Botswana and Nigeria	0.5	0.8	0.8	0.3
Nontax revenue				
Fixed-rate countries	1.7	2.9	2.2	0.5
Fixed-rate countries, except Cameroon, Congo, and Gabon	1.7	3.6	2.4	0.7
Variable-rate countries	1.9	1.6	1.9	−0.1
Variable-rate countries, except Botswana and Nigeria	2.0	1.6	1.9	−0.2

Source: National authorities and IMF staff estimates.

Figure 6.
Sub-Saharan Africa: Taxes on International Trade, 1980–91
(As percent of GDP)

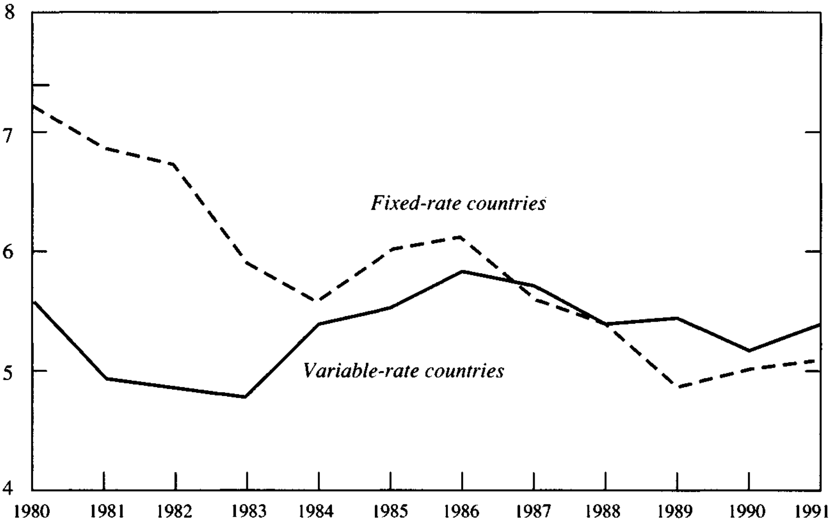
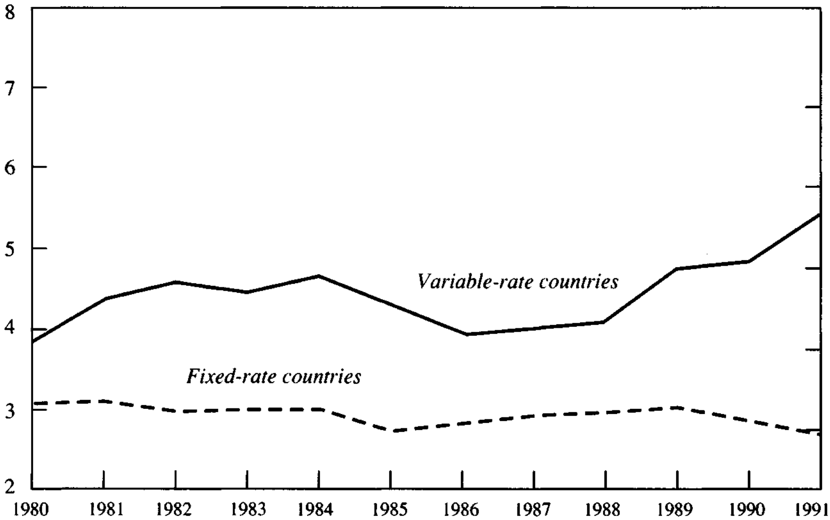


Figure 7.
Sub-Saharan Africa: Taxes on Goods and Services, 1980–91
(As percent of GDP)



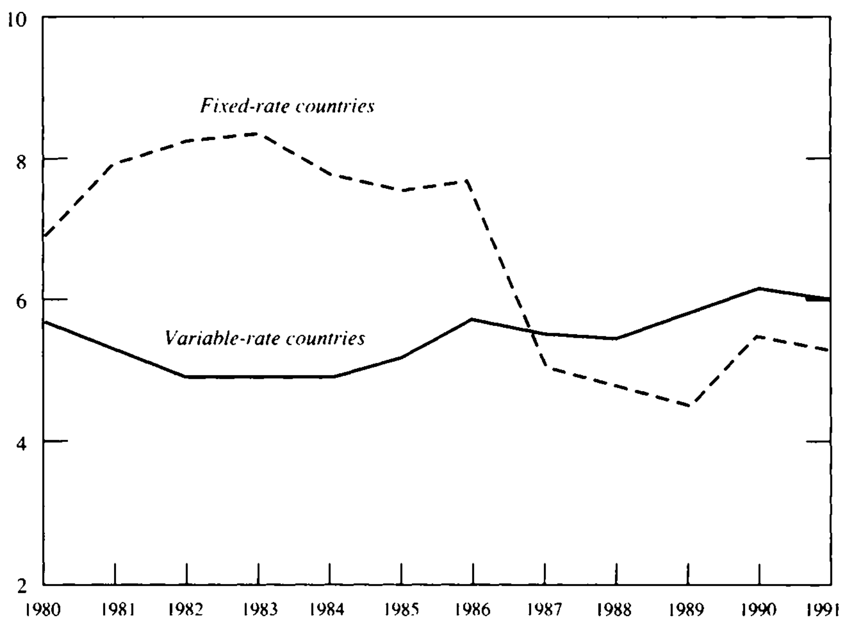
Sources: National authorities and IMF staff estimates.

low relative to other countries with similar income levels. In the variable-rate countries, excluding Botswana and Nigeria, direct taxes remained remarkably steady during the 1980s (about 4.8 percent of GDP). However, when these two countries are included (Figure 8), there is a gain in direct taxes, reflecting higher oil profits in Nigeria that stemmed partly from the sharp real depreciation of the naira, which more than offset the decline in world oil prices.

III. Determinants of Tax Revenue

The stark divergence in fiscal performance between the variable-rate and fixed-rate countries in sub-Saharan Africa raises a number of questions. What are the major factors contributing to the deterioration in revenue performance in the fixed-rate countries, and to the relative success of the variable-rate countries? Why are the major gains and losses

Figure 8.
Sub-Saharan Africa: Income Taxes, 1980–91
(As percent of GDP)



Sources: National authorities and IMF staff estimates.

Table 3. *Tax Base in Sub-Saharan Countries*

	Côte d'Ivoire 1990		Cameroon 1989/90		Mali 1990		Kenya 1989/90		Tanzania 1986/87	
	Millions of CFA francs	Percent of total receipts	Millions of CFA francs	Percent of total receipts	Millions of CFA francs	Percent of total receipts	Billions of shillings	Percent of total receipts	Billions of shillings	Percent of total receipts
Taxes on imports and traded goods sector	*508.9*	*83.2*	*390.2*	*78.8*	*72.1*	*87.3*	*33.6*	*87.7*	*26.0*	*85.8*
International trade	*266.7*	*43.5*	*73.4*	*14.8*	*60.6*	*66.7*	*12.3*	*32.1*	*7.7*	*25.4*
Exports	8.0	1.3	4.7	0.1	0.1	—	0.1	—	—	—
Imports	171.3	28.0	68.7	13.9	37.6	41.5	6.9	18.0	4.3	14.1
Sales tax on imports	65.8	10.8	—	—	12.0	13.2	5.3	13.8	3.7	12.2
Stabilization funds	21.6	3.5	—	—	10.9	12.0	—	—	—	—
Taxes on domestic transactions	*144.7*	*23.6*	*95.8*	*19.4*	*11.1*	*12.2*	*11.0*	*28.7*	*12.7*	*41.9*
Sales taxes	48.2[a]	7.8	45.4	9.2	9.6	10.6	7.6	19.8	12.4	40.9
Excises	90.0	14.7	35.4	7.1	1.0	1.1	3.0	7.8	0.3	1.0
Vehicle registration and fees	6.5	1.0	5.0	1.0	0.5	—	0.4	—	—	—
Other	—	—	10.0	2.0	—	—	—	—	—	—
Direct taxes on factors	*97.5*	*15.9*	*221.0*	*44.6*	*7.4*	*8.2*	*10.3*	*26.8*	*5.6*	*18.5*
Corporate income	32.5	5.3	38.5	9.9	5.9	6.5	7.7	20.1	4.3	14.2
Individual income	65.0	10.6	37.5	9.6	1.5	0.2	2.6	6.8	1.3	4.2
Mineral receipts	—	—	145.0	37.2	—	—	—	—	—	—
Taxes on nontraded goods sector	*102.8*	*16.8*	*104.7*	*21.1*	*11.5*	*12.6*	*4.7*	*12.3*	*4.3*	*14.2*
Domestic trade	*46.9*	*7.7*	*69.0*	*13.9*	*7.3*	*8.0*	*2.5*	*6.5*	*1.2*	*4.0*
Licenses and registration	17.3	2.8	16.6	3.3	2.0	2.2	1.0	2.6	0.2	0.1
Services	29.6[b]	4.8	18.7	3.8	5.3	5.8	1.5	3.9	1.0	3.3
Factors	*55.9*	*9.1*	*35.7*	*7.2*	*4.2*	*4.6*	*2.2*	*5.7*	*3.1*	*10.2*
Labor	44.2[c]	7.2	9.4	1.9	2.1	2.3	1.7	4.4	2.1	6.9
Property and capital	11.7	1.9	24.3	4.9	2.1	2.3	0.5	1.3	1.0	3.3
Total receipts	*611.7*	*10.0*	*494.9*	*10.0*	*90.6*	*100.0*	*38.3*	*100.0*	*30.3*	*100.0*

Source: Data from country authorities and IMF staff estimates.

[a] Of total sales tax proceeds of CFAF 58 million, CFAF 10 million were collected from the informal sector.

[b] Includes water charges of CFAF 18 million.

[c] Of which, CFAF 34.2 million is tax that is withheld from government employees.

concentrated on indirect taxes? To what extent have changes in the real exchange rate affected tax revenue and government expenditures in both groups of countries? Why were the fixed-rate countries unable to reduce their expenditures and conform to the prescriptions of the nominal anchor and the monetary policies of their two central banks? Before addressing these questions, it is important to establish the tax base profile of a typical sub-Saharan country.

Tax Base

An analysis of the major components of the tax base in selected African countries—Cameroon, Côte d'Ivoire, Kenya, Mali, and Tanzania—reveals that imports and the formal segment of the traded goods sector constitute the overwhelming share of the tax base in sub-Saharan Africa (Table 3). In this sample, imports, exports, and the formal import substitution sector would account for 78 percent to 85 percent of tax revenue. As was mentioned earlier, "nontax revenue" mostly reflects implicit taxation of exports (particularly oil and other natural-resource exporters) and to a lesser extent taxation of imports through price stabilization funds.[14]

Imports constitute the largest segment of the tax base.[15] They are either taxed directly through customs or indirectly through sales taxes and excises. Broad-based sales taxes often yield more revenue from imports than from domestic transactions. In Mali, the VAT on imports produced more revenue in 1991 (12 percent of total revenue) than the VAT on domestic transactions (8 percent). In Côte d'Ivoire, it yields 16 percent of tax revenue as opposed to 14 percent on domestic transactions. In The Gambia, the sales tax on imports generates 32 percent of total revenue, while the sales tax on domestic transactions accounts for only 8 percent. Those taxes on goods and services that are not imposed on imports are mostly levied on import substitutes manufactured domestically. They are concentrated on large manufacturing enterprises because they offer proper bookkeeping and well-defined tax collection points, which tax administrations can monitor easily, as opposed to the fragmented and

[14] In some countries, central-bank profits constitute a sizable share of nontax revenues.

[15] Explicit export taxes, which were small in the early 1980s, have been further reduced through tax reform, which has eliminated them in a number of countries. However, when implicit export taxes are added, export taxes can become significant—between 1 percent and 2 percent of GDP.

elusive retail sector.[16] On the other hand, small-scale manufacturing (furniture, leather products, and cottage industries) tends to escape the tax net. Income taxes are also mostly levied on large enterprises in the import substitution sector, either as corporate income tax or as an individual income tax withheld by these enterprises from the salaries and wages of their employees.

The nontraded goods sector (mostly government, services, domestic trade, and subsistence agriculture) contributes only 10–15 percent of total tax revenue. Government employees, utilities, and banking are the major contributors of tax revenue in this sector, mostly through direct taxes. Sales taxes on domestic services are limited to major utilities, the banking sector, and the largest hotels. Most other services and domestic trade are in the informal sector and escape taxation. Of course, the informal sector contributes tax revenue (customs duties, sales taxes, and excises) through its own spending on consumption goods but the tax collection handles are again either imports or import substitutes.

The bulk of the nontraded goods sector—subsistence agriculture, small-scale manufacturing, and all the services provided by the informal sector—are outside the tax net. Also virtually excluded because of weak compliance and tax administration are the relatively wealthy self-employed professionals (doctors, lawyers, and traders) and property owners. Some registration and license fees are imposed on property owners but, here again, an imported product—motor vehicles—accounts for the bulk of license fees.

Having established that imports, exports, and import substitution in the formal sector constitute the bulk of the tax base in sub-Saharan African countries, we proceed to analyze the major factors that can affect these activities: (i) the terms of trade, (ii) the exchange rate, (iii) inflation, and (iv) efficiency factors—associated with the trade regime, price policy, and tax structure and administration.

Terms of Trade

Declines in terms of trade have a direct price effect on the tax base by affecting the value of imports and exports and an indirect effect by reducing income and changing relative prices.[17] While in fixed-rate coun-

[16] Among the largest contributors of excise and sales taxes are the manufacturers of mass-market goods, such as petroleum products, textiles, shoes, beverages, plastics, cigarettes, fertilizers, cement, and processed foods.

[17] When real GDP is adjusted for the decline in the terms of trade, growth of national income can be substantially weaker than the growth of GDP. Thus, in

tries the decline in the terms of trade can be attributed equally to a decline in export prices and an increase in import prices, in most variable-rate countries the increase in import prices is the dominant factor. In contrast to a decline in export prices, which has a negative effect on government revenues by reducing nontax revenue and direct taxes, an increase in import prices has a positive effect on tax revenue by raising the domestic currency value of imports and import substitutes. While the income and price effects may reduce the volume of imports by shifting some of the demand for imports to cheaper import substitutes, the value of imports in relation to GDP may still rise. Depending on the relative impacts of the decline in export prices and the increase in import prices, a deterioration in the terms of trade can have a *positive* effect on the tax base. If explicit and implicit export taxation tend to be low relative to import taxation, as in the sub-Saharan countries, the positive price effect may clearly dominate.

In fixed-rate countries, export unit values in U.S. dollars fell an average 22 percent over the 1980s, while in variable-rate countries they fell by only 5 percent, although countries reliant on coffee exports (Burundi, Rwanda, and Uganda) experienced a much sharper decline. While these reductions have mostly affected nontax revenues—and have indirectly affected tax revenues through their negative income effect—both groups of countries have had virtually the same rate of increase in import unit values—about 20 percent—over the decade. Thus, while the price effect of the terms of trade decline appears equally favorable for the tax base in both country groups, the negative income effect is more severe in the fixed-rate countries.

Exchange Rate

Changes in the nominal exchange rate will affect both the unit value of imports and the GDP deflator. A depreciation of the exchange rate would tend to raise import unit values relative to the GDP deflator, and vice versa. It will also tend to increase the unit values of import substitutes

Côte d'Ivoire, the cumulative growth of real GDP between 1980 and 1991 was 5.3 percent but, when adjusted for terms of trade changes, real national income fell by 20.6 percent. This may be an extreme example since Côte d'Ivoire, with its exports concentrated on cocoa and coffee, has experienced a 56 percent decline in terms of trade between 1980 and 1991. However, other countries, particularly Cameroon, Gabon, Nigeria, Tanzania, and Uganda, have experienced even larger declines, although the effect on national income depends on the relative shares of imports and exports in GDP.

and exports relative to nontraded goods. Therefore, a depreciation of the domestic currency tends to increase tax revenue from the traded goods sector in relation to GDP. This price effect may have a negative impact on the volume of imports but a positive impact on domestic production of import substitutes and on tax revenue—both direct and indirect. As already noted, the tax base is heavily weighted toward both imports and import substitutes. In addition, by raising producer prices of exportables, real depreciation would moderate the negative impact of a decline in international commodity prices on the income of the export sector and its demand for tradables.

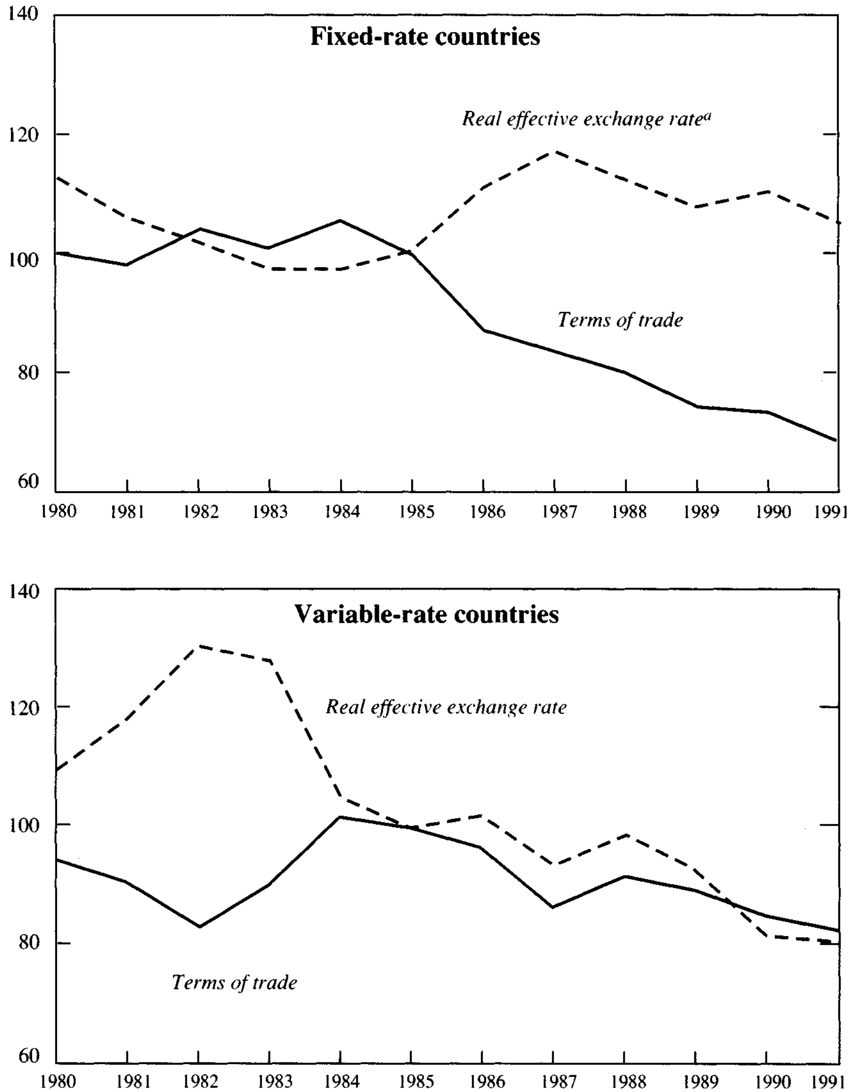
While changes in the terms of trade or in the nominal exchange rate may have a positive effect on the tax base, what really matters is how these changes are transmitted to the domestic economy through changes in the real exchange rate and other macroeconomic policies. The interplay between terms of trade developments and the nominal exchange rate will determine critical components of the tax base: import unit value, import unit value relative to the GDP deflator, import volume, import composition, the ratio of import duties to imports (the effective import duty rate), and income generated by the traded goods sector.

Import Unit Value and the GDP Deflator

The evolution of the relationship between the terms of trade and the real exchange rate in fixed-rate and variable-rate countries is shown in Figure 9. In the fixed-rate countries, while the terms of trade remained roughly stable in the first half of the 1980s, the CFA franc depreciated substantially vis-à-vis the U.S. dollar, bringing down the real effective exchange rate by about 13 percent, with the result that the import unit value in CFA francs actually increased. However, there was a substantial decline in import volume in the fixed-rate countries (20 percent), particularly when oil exporters are excluded (44 percent). Consequently, tax revenues for the fixed-rate countries remained roughly stable during that period but registered a decline when oil producers were excluded (Table 4).

After 1985, the terms of trade took a turn for the worse, declining by about 30 percent, mostly as a result of a fall in export prices. The 1987 reverse oil shock caused a marked deterioration in the fiscal revenues of Cameroon, the Congo, and Gabon. Yet, the real exchange rate moved in the opposite direction by appreciating sharply in 1986 and 1987 and maintaining much of the appreciation for the remainder of the decade (top panel of Figure 9). For the fixed-rate countries, this acted as a second external shock since this appreciation was caused by the strengthening of the French franc vis-à-vis the U.S. dollar, and indirectly by the strength-

Figure 9.
Sub-Saharan Africa: Real Effective Exchange Rates and Terms of Trade, 1980–91
(1985=100)



Sources: World Economic Outlook data base and IMF staff estimates.
[a] Foreign currency is in per unit of domestic currency; trade weights are adjusted for informal trade with The Gambia, Ghana, Nigeria, and Zaire.

Table 4. *External Sector Determinants of Tax Revenue*

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
						Index, 1985 = 100						
Terms of trade												
Fixed-rate countries	100	98	104	101	106	100	87	84	80	75	74	69
Variable-rate countries	94	90	83	90	101	100	97	86	92	90	85	83
Real exchange rate												
Fixed-rate countries	113	106	101	98	97	100	111	117	113	108	111	107
Variable-rate countries	109	117	130	127	104	100	101	93	98	93	82	81
Import value in local currency												
Fixed-rate countries	58	69	76	88	99	100	84	83	85	86	85	87
Variable-rate countries	52	54	60	64	81	100	154	285	420	727	850	920
Import volume												
Fixed-rate countries	115	113	105	93	87	100	99	91	91	82	84	81
Variable-rate countries	125	120	110	199	98	100	104	102	113	116	124	123
Unit value/GDP deflator												
Fixed-rate countries	100	102	99	104	102	100	94	94	96	99	93	95
Variable-rate countries	118	101	97	86	94	100	133	202	255	365	539	655
						In percent						
Import value/GDP												
Fixed-rate countries	28	30	27	26	25	26	24	21	22	21	19	18
Variable-rate countries	28	28	26	22	25	24	27	28	28	29	31	33
Customs receipts/import value												
Fixed-rate countries	23	22	23	22	21	21	24	24	23	22	21	21
Variable-rate countries	15	16	18	18	17	17	17	17	17	17	18	18

Source: Data provided by national authorities, IMF staff estimates, and World Economic Outlook data base.

ening of the German mark. By 1990–91, the two largest economies in the franc zone, Côte d'Ivoire and Cameroon, incurred the highest real appreciation (36 percent and 32 percent, respectively) followed by Benin (22 percent). This currency appreciation has been a major contributor to the decline in import unit values in the fixed-rate countries, when measured in CFA francs, and in the difficulties experienced by the import substitution sector, particularly the formal manufacturing sector. Interestingly enough, while imports became relatively cheaper because of the appreciation of the CFA franc, the volume of imports actually declined—as discussed below—because of the income effects in the export and import substitution sectors.

Another factor contributing to the decline in import unit values in both domestic and foreign currencies was a change in the composition of recorded imports, typically induced by recessionary conditions, toward basic necessities and lower value items. For instance, food imports have increased as a share of total recorded imports in fixed-rate countries, and there has been a tendency to replace higher priced imports by cheaper imported substitutes (for example, used cars instead of new ones). In Cameroon, products bearing high customs duties fell from 35 percent of imports in fiscal year 1985/86 to 28 percent of imports in 1990/91.[18] This may also reflect the tendency since 1985 for a larger proportion of such imports—by Benin, Cameroon, and Niger, in particular—to have originated unrecorded from Nigeria and other non-CFA neighboring countries. This drives down the average effective duty rate on imports since basic necessities, raw materials, and intermediate products bear lower customs duties (and sales taxes) than consumer goods.

Thus, the appreciation of the CFA franc combined with changes in the composition of recorded imports has reduced import unit value in CFA francs by about 15 percent in the fixed-rate countries during the second half of the 1980s. Despite a substantial increase in import prices in foreign currency, the ratio of import unit values to the GDP deflator—which was roughly stable during the 1980–85 period—has also declined by about 7 percent in the second half of the 1980s (Table 3). However, the GDP deflator understates the price relationship between nontraded goods and imports—as well as import substitutes—to the extent that it also includes export prices. Indeed, a fall in export prices raises the ratio of import unit value to the GDP deflator by depressing the denominator, as was the case in Côte d'Ivoire and Gabon, where export prices fell more than import

[18] These products include cigarettes and tobacco, beverages, furniture, textiles, plastics, glass products, aluminum, vehicles, pesticides, lubricants, electrical machinery, and electrical appliances.

unit values. On the other hand, countries that did not experience a sharp decline in relative export prices—Mali, Senegal, and Togo—have had a large decline in the price of imports relative to the GDP deflator, ranging from 20 percent for Togo to 45 percent for Senegal.

In the variable-rate countries, the real exchange rate followed terms of trade developments more closely (bottom panel of Figure 9). During the first half of the decade, it appreciated by 13 percent, parallel with a similar improvement in the terms of trade, causing import unit values to decline. Since 1985, while the terms of trade deteriorated by 17 percent, the average real depreciation reached 20 percent, and there was no evidence of a change in the composition of imports. Thus, while the ratio of import unit values to the GDP deflator was stable during the first half of the decade, it increased by an average 70 percent during the second half of the 1980s. The largest increases occurred in countries where currencies depreciated rapidly—Madagascar, Nigeria, Uganda, and Tanzania.

Import Volume and the Import-to-GDP Ratio

The divergent trajectories of the real exchange rate and terms of trade in the fixed-rate countries have severely undermined the competitiveness of the traded goods sector and growth performance. Cocoa and coffee producer prices were reduced by more than half in nominal terms in Cameroon, Côte d'Ivoire, and Togo, falling in some cases below production costs. Export proceeds from mineral products (oil in Cameroon, Gabon, and Congo, uranium in Niger, phosphates in Togo) declined sharply with an immediate impact on government revenue from lower royalties and income taxes.

The second half of the 1980s also coincided with import liberalization, which lowered customs tariffs and relaxed quantitative restrictions in fixed-rate countries. While more imports may have contributed to customs duties receipts, they nevertheless exposed the import substitution sector to stiffer competition at a time when domestic demand was falling because of exchange rate appreciation. Moreover, both Nigeria and Zaire sharply depreciated their currencies in real terms, which heightened the competitiveness of their traded goods sectors, causing substantial market penetration in the fixed-rate countries. Overall, the growth of real GDP in these countries declined from 3 percent during the first half of the 1980s to 1.2 percent during the second half, with a concomitant decline in per capita income (Table 5). In contrast, the annual growth rate in variable-rate countries rose from 2.1 percent to 3.6 percent during 1986–91.

Table 5. *Macroeconomic Indicators*
(Average annual rate in percent)

	1980–85	1986–91	1980–91
Real GDP growth rate			
Fixed-rate countries	3.0	1.2	2.1
Variable-rate countries	2.1	3.6	2.9
Inflation rate			
Fixed-rate countries	9.4	1.6	5.5
Variable-rate countries	23.7	33.3	28.5
Monetary expansion			
Fixed-rate countries	14.7	2.6	8.7
Variable-rate countries	21.4	30.8	26.1
Gross capital formation/GDP			
Fixed-rate countries	22.6	17.6	20.1
Variable-rate countries	19.6	18.6	19.1
External current account/GDP			
Fixed-rate countries	−6.1	−5.9	−6.0
Variable-rate countries	−6.9	−4.5	−5.7
Imports/GDP			
Fixed-rate countries	27.2	20.1	23.6
Variable-rate countries	25.6	29.2	27.4
Debt/GDP			
Fixed-rate countries	52.1	74.3	63.2
Variable-rate countries	54.2	107.2	80.8
Debt service/exports			
Fixed-rate countries	23.7	25.1	24.4
Variable-rate countries	25.4	39.9	32.7

Source: World Economic Outlook data base and IMF staff estimates.

Finally, the decline in the competitiveness of the traded goods sector in fixed-rate countries has resulted in an expansion of the informal sector, further undermining the tax base. Under an overvalued exchange rate, the incentives to operate in the informal sector become particularly strong. On the supply side, moving into the informal sector improves a producer's competitiveness by avoiding the costs associated with the formal sector's regulatory framework, import duties and other taxes, license fees, and labor regulations. On the demand side, the consumption pattern shifts toward cheaper substitutes, mostly unrecorded imports from Nigeria and Zaire, offered by the informal trading sector. One indicator of the expansion of the informal sector is the export of CFA francs to neighboring countries—Nigeria, Zaire, Ghana, and The Gambia—in exchange for goods. These currency notes are then remitted to Europe (mainly through London) and repurchased by the central bank

in France.[19] Such repurchases, which can be interpreted as the net trade balance of the informal sector with more competitive neighboring countries, increased from less than FF 5 billion in 1985 to a peak of FF 12 billion in 1992.[20]

In the 1970s, the volume of imports rose steadily in sub-Saharan Africa in the wake of the commodity price boom and external borrowing, making up for any occasional decline in import unit values. But with the collapse of export prices and the need to undertake fiscal retrenchment, the demand for imports—both public and private—declined. Despite the favorable price effect of the CFA franc appreciation, recorded imports in the fixed-rate countries declined throughout the 1980s (particularly during the second half of the decade), falling by 1991 to 70 percent of the import volume of 1980. In the variable-rate countries, import volume also declined during the first half of the decade by 23 percent but increased during the second half by a similar magnitude, with the result that, by 1991, import volume was about the same as in 1980. As a percentage of GDP, import volume declined for both groups of countries during the first half of the decade. However, during the second half, variable-rate countries maintained their import shares, while fixed-rate countries incurred a 20 percent decline.

The combination of the changes in import prices in local currency, import volume, and the composition of imports determines a critical component of the tax base, namely the value of recorded imports in relation to GDP. In the fixed-rate countries, recorded imports declined from 29 percent of GDP in 1980–81 to 19 percent of GDP in 1990–91, with the downward trend accelerating after 1985 (Table 4). This shrinkage in the tax base by 10 percentage points of GDP on the import side and a concomitant decline in the production level of the formal import substitution sector—instead of a resurgence in import substitution under adequate exchange rate protection—go a long way in explaining the weakening of tax revenue in the fixed-rate countries.

Variable-rate countries also experienced a reduction in their imports in relation to GDP during the first half of the decade, very much in line with the fixed-rate countries (from 28 percent to 25.5 percent of GDP). However, with the real exchange rate depreciation that took place after 1985, imports increased to 32 percent of GDP by 1990–91. Thus,

[19] These flows should be distinguished from capital movements which are effected mostly through the banking system.

[20] See Secrétariat du comité monétaire de la zone franc (several issues). On August 1 and 2, 1993, in reaction to a sharp increase in the outflow of their currency, the 13 CFA franc zone countries suspended the repurchase of their currency notes, in effect partially suspending the convertibility of the CFA franc.

variable-rate countries were able to broaden this critical tax base—the imports to GDP ratio—over the 1980s.[21]

Efficiency Factors

In addition to the macroeconomic factors outlined above, tax revenue has been affected by discretionary government policy and the institutional changes carried out over the past decade under structural adjustment programs supported by the IMF and the World Bank. In most sub-Saharan African countries, there has been an attempt to improve overall efficiency by allowing the market to determine prices and dictate most investment decisions, rather than following the failed industrial strategy of the 1970s. In the fiscal area, this approach has entailed broadening the tax base, deregulating prices, lowering external tariffs, liberalizing imports, shifting the taxation system toward domestic transactions, and reforming the income tax. It is difficult to measure the overall impact of these reforms on tax revenue, but they tend to reduce distortions and create an enabling environment for investment and higher growth.

Governments have also resorted to ad hoc fiscal measures to relieve the economic pressures experienced in the 1980s. Additional taxes (such as an import surcharge or solidarity tax) or increases in tax rates have frequently been introduced after a widening of the fiscal deficit. Such pressure has been manifest in fixed-rate countries that needed to mobilize additional resources to finance their budget deficits. A case in point in Côte d'Ivoire, which in 1987 and 1988 increased its import tariffs by 10 percent, increased excise taxes on tobacco, and increased stamp duties, license fees, and vehicle registration fees. Indeed, some doubts may be raised on the quality of adjustment in this case, since raising government revenues by increasing tax rates, given the concomitant narrowing of the tax base, cannot be sustained and may follow a Laffer curve scenario.

There were also pressures in the opposite direction, mostly in response to financial difficulties experienced by enterprises in import substitution

[21] The linkages among government revenue, the terms of trade, and the real exchange rate have been tested using panel data over the 1980–91 period. The econometric results, available from the authors, indicate that tax revenue in sub-Saharan Africa is negatively correlated with movements in the real value of the domestic currency: an appreciation of the real exchange rate would yield a fall in the ratio of tax revenue to GDP. It is also negatively correlated with the terms of trade. Here, the positive price effect of an increase in import prices dominates the negative income effect of the terms of trade deterioration. Similar results are obtained when total revenue is used as an independent variable.

activities, which resulted in lower average effective rates and a further narrowing of the tax base. Such tax relief included granting tax exemptions, allowing the build-up of tax arrears, and granting generous tax holidays. Fixed-rate countries have been particularly prone to these tendencies, with acute competitive pressures arising from the appreciation of the CFA franc and from the real currency depreciation in some neighboring countries. Thus, it has been estimated that roughly half of the imports into fixed-rate countries have been exempted from customs duties for various reasons, including generous provisions under investment codes. The proliferation of exemptions has reduced the average effective customs tariff, as measured by the ratio of receipts from import duties to total recorded imports, to less than half the statutory rate for the seven members of the West African monetary union.[22]

Another factor that has contributed to the narrowing of the import tax base is the growing importance of donor-financed imports, which are tax exempt. This has affected both country groups in a similar fashion. After the debt crisis, there has been a significant shift in project financing away from commercial sources, which paid full customs duties, to official bilateral and multilateral donors, which expect to be fully tax exempt for foreign assistance. Some of the changes in the import tax base can be traced to the average effective tariff. In the early 1980s, the average effective tariff in fixed-rate countries was 22 percent, substantially higher than in the variable-rate countries (15 percent). Despite efforts at broadening the taxation of imports (through the imposition of a minimum tariff), the effective tariff fell to 20 percent, mostly because of tariff reform, changes in the composition of imports, and the broadening of exemptions. In the variable-rate countries, many of which started off with a very low tariff base in the early 1980s (Tanzania, Zambia, and Uganda), the average effective import tariff rose from 15 percent in 1980 to 18 percent in 1990–91.

Impact of Inflation on Tax Revenue

In addition to the exchange rate, other components of the macroeconomic mix can also have a significant impact on tax revenue. In particular, when inflationary policies are pursued by the government—through expansionary fiscal policies, monetization of the budget deficit, and frequent devaluations—tax revenue can be affected negatively in various

[22] See Nashashibi, Lorie, and Doizé (1990).

ways. Collection lags, the effects of maintaining specific excise taxes, administered prices for basic commodities and utilities, and declines in real interest rates toward negative levels are among the factors that erode the real value of the tax base and contribute to widening the fiscal deficit (Tanzi (1977)). An empirical overview of the relationship between inflation and tax revenue in sub-Saharan Africa confirms that tax revenue has declined in countries where there have been sharp accelerations of inflation (such as Uganda, Zambia, and Zaire). However, for the remainder of the variable-rate countries, *the level of inflation*—the persistence of inflation in the 15–20 percent range, as opposed to 2 percent in the fixed-rate countries—did not seem to affect tax revenue significantly.

Average inflation in the variable-rate countries has certainly been higher than in the fixed-rate countries. During the first half of the decade, the inflation rate averaged 24 percent and accelerated to 33 percent during the 1986–91 period, as a result of higher monetary expansion and a substantial real depreciation of the exchange rate in a number of countries. Seigniorage revenue was almost double that of the fixed-rate countries during the first half of the decade (2.1 percent of GDP) and increased to 3.2 percent of GDP during the second half.[23] However, within these broad aggregates, one can distinguish two groups of countries. The first group of 11 countries made progress in reducing average inflation from 17 percent during the first half of the 1980s to 13 percent during the second half of the decade (Botswana, Burundi, The Gambia, Ghana, Kenya, Madagascar, Malawi, Mauritania, Mauritius, Rwanda, and Zimbabwe). Their tax revenue increased from 17.8 percent of GDP in 1985–86 to 21.1 percent of GDP in 1990–91. In the remaining six high-inflation countries of the variable-rate sample (Nigeria, Sierra Leone, Tanzania, Uganda, Zaire, and Zambia), inflation accelerated from 40 percent in 1980–85 to 75 percent during the 1986–91 period, with most of the deterioration owing to the sharp price increases in Uganda, Zaire, and Zambia. During the same period, tax revenue in these countries declined from 15.3 percent of GDP to 13.1 percent of GDP. The increase in the tax revenue of the first group of countries, despite persistent inflation at about 15 percent a year, may have to do with the structure of taxation in sub-Saharan Africa and efforts at containing inflationary pressures.

First, most of the consumption taxes levied on these imports and

[23] See World Bank (1994). The literature on optimal taxation has made a case for some degree of seigniorage revenue.

import substitutes are expressed in ad valorem terms, and the remaining specific excises have been gradually converted to an ad valorem basis through tax reforms (as has happened in Burundi, The Gambia, Malawi, Mali, and Mauritania). Moreover, the liberalization of exchange rate policies has ensured that the market exchange rate is the rate at which imports are valued at customs.[24] Thus, increases in the prices of traded goods tend to be translated into higher tax revenue with relatively low collection lags, typically within one or two months. Second, interest income, which is particularly sensitive to erosion under inflation and which is a substantial source of taxation in Latin America, does not as yet constitute a significant source of tax revenue in most sub-Saharan African countries. Moreover, the restructuring of interest rates and yields on government obligations in several variable-rate countries (The Gambia, Ghana, Kenya, Malawi, Mauritius, Nigeria, Tanzania, and Uganda) in the 1980s have tended to introduce positive real interest rates (Galbis (1993)).

Third, income taxes, which can be seriously eroded by inflation because of collection lags, account for the lowest share of tax revenues and have been reformed in a number of countries. Corporate income tax rates have been lowered and collection lags have been reduced by levying tax payments monthly or quarterly on expected income in the current year rather than on the income of the previous year (Kenya and Malawi). The individual income tax, which is mostly levied on wages in the formal sector, is withheld at source and is not eroded by inflation. On the other hand, lowering marginal tax rates toward the 35–40 percent level and raising the standard deduction with inflation have minimized the fiscal drag. These factors have enabled the variable-rate countries to maintain the ratio of direct taxes to GDP over the entire 1980–91 period (at 4.8 percent of GDP) despite the high inflation environment.

IV. Public Expenditure

In our full sample of 28 countries, public expenditures have remained remarkably steady at about 28 percent of GDP between 1980 and 1991. But as in the case of the overall budget deficit, the expenditure pattern of the two groups of countries is quite different.

[24] Under the dual exchange rate regimes that have been pursued in the past (Nigeria, Tanzania, and Zambia), the official overvalued exchange rate would serve as a basis for customs valuation of imports, resulting in the erosion of this tax base under inflation.

Expenditure Policy in Fixed-Rate Countries

Under the internal adjustment strategy, major policy requirements for maintaining a pegged exchange rate as a nominal anchor are that governments conduct fiscal policy in such a way that (i) the budget deficit is reduced to a level that makes demand for foreign exchange consistent with the pegged exchange rate and (ii) the relative prices between traded and nontraded goods can change in response to protracted external shocks so as to maintain competitiveness.

It is useful to recall that since 1985 the double impact of the deterioration of the terms of trade and the appreciation of the CFA franc resulted in a substantial decline in the competitiveness of the traded goods sector. In the face of this external shock, fiscal and monetary policies are called upon to lower real wages—and the prices of nontradables—so that the costs of producing traded goods decline sufficiently to restore their profitability. In most CFA countries, this would have entailed lowering *nominal* wages since inflation had already been reduced to very low levels. Since most tax revenue originates from imports and the traded goods sector, the drop in competitiveness has lowered tax revenue and contributed to a widening of the fiscal deficit. Here lies the essential contradiction of the internal adjustment strategy: the real appreciation of the exchange rate, by undermining competitiveness, reduces tax revenue, which widens the fiscal deficit. Yet this expansionary fiscal impulse goes against the contractionary fiscal policy needed for lowering real wages. Hence, expenditure policy is called upon not only to reduce the deficit and change relative prices but also to make up for the decline in tax revenue.

Despite the need to reduce the large fiscal deficits that emerged at the beginning of the 1980s, public expenditures in the fixed-rate countries have actually increased during the first half of the decade from 27.9 percent of GDP in 1980–81 to about 29.4 percent of GDP in 1985–86 (Table 6 and Figure 10). Since tax revenue was declining during the period, fiscal discipline would appear to have been relaxed in many fixed-rate countries, with costly consequences for the second half of the decade. During the 1986–91 period, with the further decline in tax revenue, the emergence of payment arrears, and the tightening of monetary policy, government expenditures were reduced to 25.3 percent of GDP by 1990–91. This major expenditure effort, equivalent to almost 4 percentage points of GDP, is particularly noteworthy considering that GDP was also declining and that there is usually little short-run policy flexibility in reducing current expenditures. Indeed, this decline was mostly attained by cutting capital expenditures. Subsequently, when

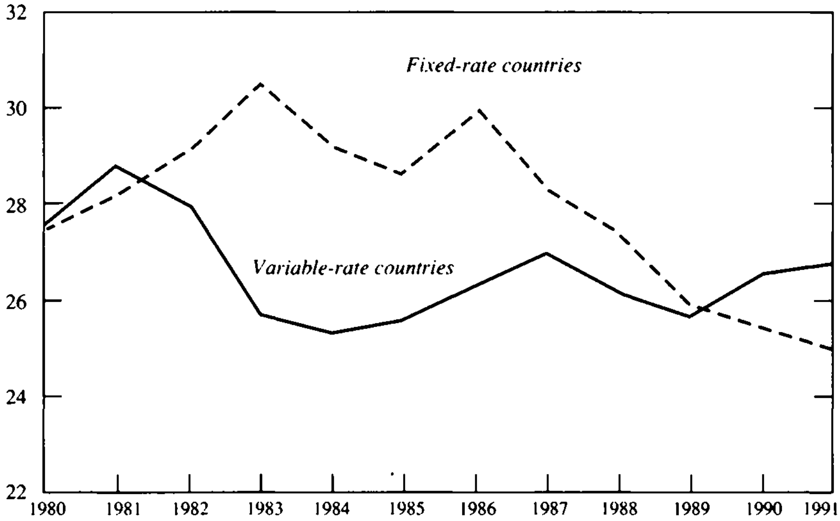
Table 6. *Expenditure Performance*
(In percentage of GDP)

	1980–81	1985–86	1990–91	Change from 1980–81 to 1990–91
Total expenditure				
Fixed-rate countries	27.9	29.4	25.3	−2.6
Fixed-rate countries, except Cameroon, Congo, and Gabon	27.9	27.2	24.1	−3.8
Variable-rate countries	28.2	26.0	26.7	−1.5
Variable-rate countries, except Botswana and Nigeria	27.6	26.1	25.8	−1.7
Current expenditure				
Fixed-rate countries	16.8	17.4	18.2	1.4
Fixed-rate countries, except Cameroon, Congo, and Gabon	16.9	16.9	16.3	−0.6
Variable-rate countries	19.1	18.9	19.3	0.2
Variable-rate countries, except Botswana and Nigeria	19.4	19.5	19.4	0.0
Wages				
Fixed-rate countries	7.8	7.6	8.7	0.9
Fixed-rate countries, except Cameroon, Congo, and Gabon	8.4	7.7	8.2	−0.2
Variable-rate countries	6.9	6.1	5.8	−1.1
Variable-rate countries, except Botswana and Nigeria	7.0	6.2	5.9	−1.1

Goods, services, and transfers				
Fixed-rate countries	7.1	6.4	5.3	−1.8
Fixed-rate countries, except Cameroon, Congo, and Gabon	6.8	6.1	4.8	−2.0
Variable-rate countries	10.4	8.8	9.2	−1.2
Variable-rate countries, except Botswana and Nigeria	10.6	9.0	9.4	−1.2
Interest				
Fixed-rate countries	1.9	3.3	4.1	2.2
Fixed-rate countries, except Cameroon, Congo, and Gabon	1.7	3.2	3.3	1.6
Variable-rate countries	1.8	4.1	4.3	2.4
Variable-rate countries, except Botswana and Nigeria	1.9	4.3	4.1	2.2
Capital expenditure				
Fixed-rate countries	10.4	11.2	6.3	−4.1
Fixed-rate countries, except Cameroon, Congo, and Gabon	10.2	9.6	7.0	−3.3
Variable-rate countries	8.5	6.7	7.4	−1.2
Variable-rate countries, except Botswana and Nigeria	7.4	6.2	6.7	−0.8

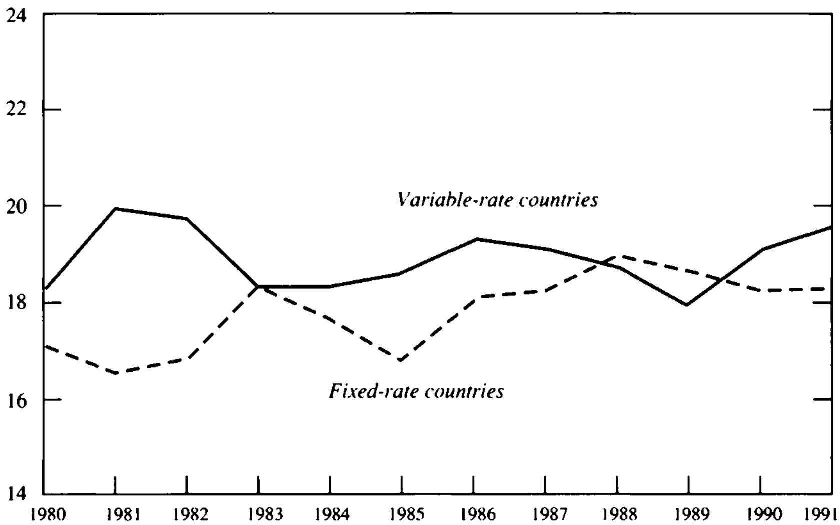
Source: National authorities and IMF staff estimates.

Figure 10.
Sub-Saharan Africa: Total Expenditure, 1980–91
(As percent of GDP)



Sources: National authorities and IMF staff estimates.

Figure 11.
Sub-Saharan Africa: Current Expenditure, 1980–91
(As percent of GDP)



Sources: National authorities and IMF staff estimates.

restoring competitiveness required a reduction in current expenditures (by reducing wages), these expenditures actually increased from 17.4 percent of GDP in 1985 to 18.2 percent of GDP in 1990–91 (Figure 11).

As a rough order of magnitude, balancing the budgets of the fixed-rate countries (including grants) by the end of the 1980s would have required a fiscal retrenchment of 6 percent of GDP.[25] At least half of this amount would have had to come from reducing wages to the average level in variable-rate countries, in an attempt to lower the prices of nontradables. Actually, the budgetary wage bill in fixed-rate countries (in relation to GDP) rose by 17 percent between 1980 and 1991 (Figure 12).[26]

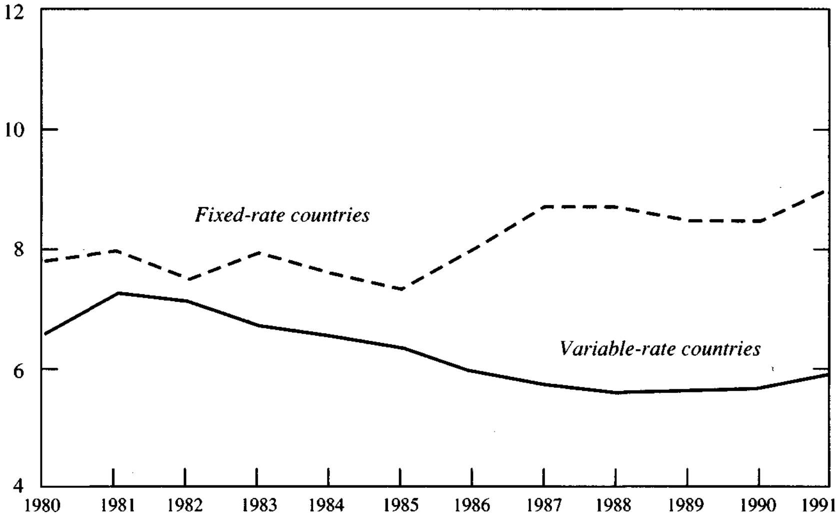
The government wage bill in the fixed-rate oil-exporting countries (Cameroon, Congo, and Gabon) continued to grow from 6.4 percent of GDP in 1980–81 to 10.3 percent by 1990–91 (Table 6), partly through increases in salaries and partly through an expansion in the size of the civil service, in order to staff the large infrastructure created in the early 1980s. Again, this reflects lax financial discipline within the two monetary unions, since this increase coincided with an accumulation of domestic and external arrears. In contrast, the non-oil countries made early efforts to reduce the wage bill, from 8.4 percent of GDP in 1980–81 to 7.6 percent in 1985–86. These efforts were maintained in the second half of the decade, but the fall in real GDP was larger than the nominal reductions in the wage bill, with the result that the share of wages in GDP rose from 7.6 percent in 1985–86 to 8.3 percent of GDP in 1990–91. This underscores the political difficulty faced by fixed-rate countries in reducing nominal wages and public employment under recessionary conditions.

The difficulties in exercising expenditure restraint in fixed-rate countries were exacerbated by a doubling of interest costs, from 2 percent of GDP in the early 1980s to 4 percent of GDP by the end of the decade. It should be noted, however, that this interest cost understates the full interest cost that would have been incurred by fixed-rate governments in the presence of functioning financial markets. Indeed, there are various forms of debt incurred by these governments—domestic payment arrears

[25] As the fiscal deficit is reduced, tax revenue would be expected to fall further in relation to GDP, so that the actual fiscal effort required may be greater than the initial deficit.

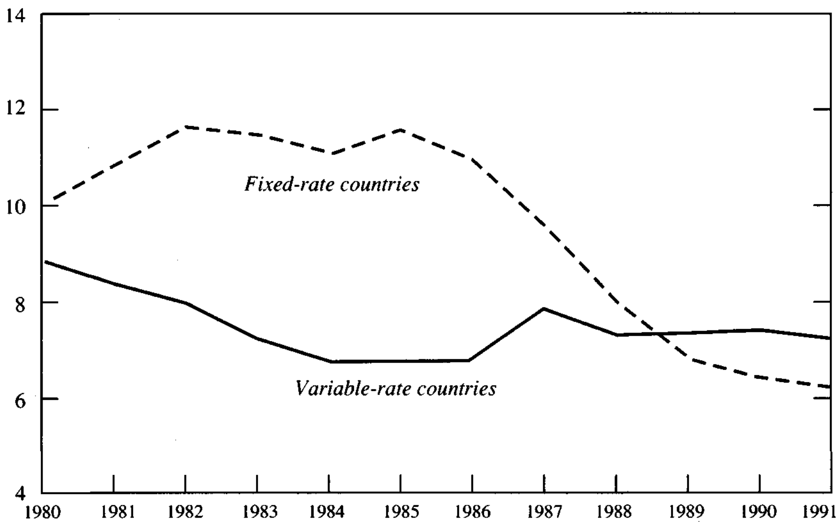
[26] When the government wage bill in fixed-rate countries is contrasted with some indicators of private sector income, such as producer prices for coffee and cocoa, which were reduced by half during the second half of the 1980s, or with trends in private consumption as estimated by national accounts during the same period, it appears that government adjustment efforts to improve the fiscal balance have lagged well behind those of the private sector (Secrétariat du comité monétaire de la zone franc (1992 and other issues)).

Figure 12.
Sub-Saharan Africa: Expenditure on Wages and Salaries, 1980–91
(As percent of GDP)



Sources: National authorities and IMF staff estimates.

Figure 13.
Sub-Saharan Africa: Capital Expenditure, 1980–91
(As percent of GDP)



Sources: National authorities and IMF staff estimates.

and arrears on social security contributions—on which no interest has been paid or budgeted.

The rise in both interest and wages in relation to GDP inevitably squeezed expenditures on goods and services and transfers. These were reduced from 7.1 percent of GDP at the beginning of the 1980s to 5.6 percent of GDP in 1990–91. This reduction has affected budgetary transfers to enterprises and consumers, although they have been relatively small in fixed-rate countries. But, more seriously, expenditures on schools, health facilities, and government administrations (including vehicles for tax administration) have been sharply cut in some countries and may have led to a deterioration of the infrastructure and the services provided by the government. This type of fiscal adjustment can be temporary only and would eventually require the restoration of expenditures on goods and services to a more sustainable level. It suggests that the underlying budget deficit in the fixed-rate countries during the 1988–91 period may be larger than the measured deficit.

The increase in current expenditures in the fixed-rate countries, concurrent with the sharp decline in revenue and the financing difficulties experienced in the 1985–91 period, forced substantial cuts in public capital expenditures. After increasing during the first half of the decade, they were reduced by almost half, from 11.5 percent of GDP in 1985–86 to 6.1 percent in 1990–91 (Figure 13). In some countries, this reduction proved beneficial, as it eliminated unproductive or marginal investments; in other countries (Congo, Côte d'Ivoire, and Senegal) gross capital expenditures were reduced to 2–3 percent of GDP, which may have resulted in a net reduction of the capital stock.

Expenditure Policy in the Variable-Rate Countries

In the variable-rate countries, total expenditures declined throughout the 1980s from about 28.3 percent of GDP in 1980–81 to 26.7 percent of GDP in 1990–91. As in the case of the fixed-rate countries, most of this adjustment can be attributed to a reduction in capital spending. However, in contrast to the fixed-rate countries, current expenditures remained virtually stable throughout the decade at 19.3 percent of GDP. If interest costs are excluded, however, current expenditures fell by 2.5 percent of GDP between 1980–81 and 1990–91.

In contrast to the fixed-rate countries, the wage bill in the variable-rate countries declined from 6.9 percent of GDP in 1980–81 to about 6 percent of GDP in the mid-1980s (and marginally thereafter), ending the decade at 5.8 percent of GDP, or about one-third lower than the wages-

to-GDP ratio in the fixed-rate countries. Much of this decline resulted from lags in nominal wage adjustments behind price increases. Policies of active exchange rate depreciation in the variable-rate countries have succeeded in reducing real wages.

A reduction in real wages was needed to accommodate the sharp increase in interest costs from 1.8 percent of GDP in 1980–81 to 4.3 percent of GDP in 1990–91. It is interesting to note that, by the end of the decade, despite very substantial real exchange rate depreciation, the interest burden in the variable-rate countries in relation to GDP was only marginally higher than interest expenditures in the fixed-rate countries (4.3 percent of GDP versus 4.1 percent of GDP), even though the two groups of countries started off the decade with similar interest charges. In particular, during the second half of the decade, when the depreciation of the real exchange rate was at its highest, the interest burden hardly changed. Yet it is a common criticism of an active exchange rate policy that servicing of the public external debt escalates, resulting in a deterioration of the budget. Three factors have contributed to this outcome. First, the widening of fiscal deficits in the fixed-rate countries over the decade raised interest costs, while the opposite occurred in the variable-rate countries. Second, the poor growth record of the fixed-rate countries has reduced GDP, while the better growth record of the variable-rate countries has increased it. Third, there was significant capitalization of interest through debt rescheduling and, particularly since 1987, through the lengthening of maturities, the partial cancellation of debt, and concessions on interest rates. This may have benefited the variable-rate countries to a greater degree than the fixed-rate countries. Because the variable-rate countries are generally poorer than the fixed-rate countries, they benefit more from favorable rescheduling terms, including debt cancellation. Moreover, the interest burden becomes endogenous to the adjustment process when reducing the fiscal deficit and undertaking structural changes facilitate debt rescheduling; adjustment may also provide greater concessionality in foreign financing.[27] However, a greater degree of debt rescheduling, coupled with an active exchange rate policy, would also result in a larger stock of debt. Indeed, the ratio of external debt to GDP in the variable-rate countries doubled over the decade to 107 percent of GDP in 1986–1991 (Table 5).

There was also a reduction in government expenditures on goods and services, which declined in the variable-rate countries from 10.5 percent

[27] The same factors may account for a doubling of grants to 3.6 percent of GDP in variable-rate countries over the 1980s, while in fixed-rate countries they declined from 3.2 percent of GDP in 1980–81 to 2.4 percent of GDP in 1990–91.

of GDP in 1980–81 to 9.3 percent by 1985–86, remaining stable at that level for the remainder of the decade. By 1990–91, expenditures on goods and services as a percent of GDP were 3 percentage points above their level in fixed-rate countries.

Capital expenditures were also reduced in the variable-rate countries, mostly during the first half of the decade (from 8.7 percent of GDP to 6.4 percent of GDP), and were raised to a lesser extent during the remainder of the decade. Hence, during the second half of the decade, when the decline in the terms of trade was its worst, expenditure reduction in the variable-rate countries was broadly distributed, while in the fixed-rate countries it was concentrated in lower spending on goods and services and on capital investment.

V. Summary and Conclusions

Fiscal adjustment in sub-Saharan Africa during the 1980s has been difficult. A protracted decline in the region's terms of trade and falling per capita income have clouded the political climate in most countries and surrounded any adjustment strategy with substantial risks. Seen from this perspective, the improvement in the fiscal performance of most sub-Saharan countries is particularly noteworthy. Nevertheless, opportunities for adjustments and consolidation of stabilization efforts have been missed.

During the first half of the decade (1980–85), the real exchange rate depreciated in the fixed-rate countries by about 15 percent in conjunction with a small improvement in the terms of trade. This provided an opportunity for fiscal retrenchment, and the fixed-rate countries were able to improve their revenue performance from 21 percent to 22 percent of GDP. However, because they were reaching the upper limits of revenue increases, the burden of adjustment fell squarely on expenditures. Yet, the oil-producing fixed-rate countries continued to increase capital expenditure in relation to GDP, offsetting revenue gains and widening their deficits. And while the other fixed-rate countries managed to reduce their fiscal deficit marginally (from 10.5 percent of GDP to 9.5 percent of GDP), this was clearly inadequate to restore domestic balance and consistency between monetary and fiscal policy. In the variable-rate countries, while government revenue in relation to GDP hardly changed (18.2 percent of GDP), there was a major effort on the expenditure side, which reduced the deficit from 11 percent of GDP to 7 percent of GDP, notwithstanding a doubling of interest costs.

During the second half of the decade, the terms of trade deteriorated sharply in the fixed-rate countries (by 30 percent) while the real effective

exchange rate appreciated by about 10 percent. The combined impact of these two external shocks, coupled with import liberalization and lower trade tariffs, weakened the traded goods sector by reducing the competitiveness of both exports and import substitutes. This undermined the tax base, as indicated by a fall in the import-GDP ratio from 27 percent to 21 percent. Falling incomes in the export and manufacturing sectors, cheaper imports, lower volume of recorded imports, and an expansion of the informal sector reduced budgetary revenue in fixed-rate countries by 4.3 percentage points of GDP, a major deterioration considering that it occurred over a span of only five years and that GDP was stagnant.

The policy response to the fall in tax revenue and the rising interest burden has been to increase tax rates, reduce public spending on goods and services, cut investment expenditures, and finance the fiscal gap through foreign borrowing, grants, and accumulation of domestic and external payment arrears. Increases in tax rates and resorting to domestic payment arrears have weakened the formal traded goods sector and strengthened the informal sector, further contributing to a narrowing of the tax base and to a deterioration of tax compliance. While adjustment should have occurred through lower civil service wages and employment, the wage bill actually increased in relation to GDP. Moreover, some of the expenditure cuts—particularly in operations, maintenance, and supplies—may have been counterproductive and unsustainable.

The decline in the terms of trade in the variable-rate countries during 1986–91 was more moderate than in the fixed-rate countries (20 percent versus 30 percent) and was fully matched by a reduction in the real effective exchange rate. This helped to increase the import ratio in variable-rate countries (from 26 percent of GDP to 29 percent of GDP), thereby protecting the tax base and raising revenue. However, the momentum of reducing expenditure stalled during the second half of the decade at about 26 percent of GDP. Even in the 11 countries undergoing sustained adjustment efforts, inflation remained imbedded at the two-digit level because of insufficient fiscal retrenchment and expansionary monetary policy. Nevertheless, by the end of the decade, the government wage bill in the variable-rate countries was one-third lower than the wage bill in the fixed-rate countries, and the primary deficit was further reduced to 1.4 percent of GDP (from 9 percent in 1980–81), which is the most telling indicator of fiscal adjustment in variable-rate countries.

Drawing on the fiscal experience of the two country groups analyzed in this study, a number of policy-oriented findings can be highlighted.

—A major factor in the deterioration of fiscal performance in the fixed-rate countries during the second half of the 1980s was that the real exchange rate increasingly diverged from its equilibrium path. Con-

versely, the variable-rate countries were able to improve their fiscal performance, despite their expansionary monetary policies, because their real exchange rate was converging toward its equilibrium path. Thus, while using a fixed exchange rate as a nominal anchor can have major advantages in limiting monetary expansion and ensuring price stability, it can contribute to the attainment of other macroeconomic objectives—particularly fiscal balance, competitiveness, and growth—only if it is adjusted in response to protracted external shocks.

—Because the tax base in sub-Saharan Africa is so heavily concentrated on the imports and import substitutes of the modern traded goods sector, protecting the competitiveness of that sector is essential for the preservation of tax revenue. Attempts at restoring the fiscal balance and competitiveness by wage deflation may be partly self-defeating and, given the magnitude of the required adjustment, unrealistic. It is self-defeating in the sense that an overvalued exchange rate, by undermining tax revenue, leads to a widening of the deficit and an expansionary fiscal impulse, when what is required under internal adjustment is a contractionary policy. An internal adjustment based on a fixed exchange rate strategy would require not only a major decline in nominal wages, a politically difficult undertaking under any circumstance, but one of such a magnitude as to offset the decline in government revenue triggered by the erosion of the tax base. Even if such a strategy were feasible, the cost in forgone output would be considerable. Under these circumstances, it would be very difficult for fiscal policy to serve as a proxy for exchange rate policy. Rather, exchange rate policy and fiscal policy should be viewed as complementary. An inappropriate exchange rate policy undermines the ability of the fiscal authorities to attain fiscal balance, let alone competitiveness. Thus, presenting the internal and external adjustment strategies as two alternatives is not realistic for policy purposes.

—The transmission mechanism between the reduction in the budgetary wage bill and an increase in competitiveness—say, for export crops—is not clear in the institutional set-up of sub-Saharan countries. Non–market-clearing wages can persist despite substantial unemployment because of the dual structure of the labor market: the formal sector, which is influenced by union and political power, and the informal one, in which the implicit wage in the agricultural subsistence sector can act as a floor. More work needs to be done on wage determination in African countries before wage deflation is the recommended policy for changing relative prices.

—The variable-rate countries as a group did not reduce their real exchange rate during the second half of the decade beyond the deterioration in their terms of trade. While individual countries may have

pursued a more active exchange rate policy, these countries did not resort—as a group—to "beggar thy neighbor" policies. Nevertheless, there is a need for regional surveillance of macroeconomic policies in Africa and the establishment of a behavioral framework within which outlier countries can be monitored.

—Fiscal discipline in the fixed-rate countries, as implied by the statutory arrangements of the western and central African monetary unions, was insufficient for the smooth operation of this regime. When faced with major fiscal gaps, the fixed-rate countries found it difficult to reduce current expenditures, particularly nominal wages. Instead, a variety of ways were found to finance these gaps, including the build-up of domestic and external arrears, the use of social security funds, and the circumvention of the statutory limits on government borrowing by drawing on credit extended to parafiscal agencies (such as stabilization funds) or resources available to public enterprises, particularly those processing mineral deposits (such as oil in Cameroon, Gabon, and the Congo, and phosphates in Togo).

—The rules followed by the two regional banks in West Africa in determining credit to the government (the use of the previous year's fiscal revenue as a measure of creditworthiness) may have exacerbated fluctuations in income. Considering that fiscal revenues are closely related to terms of trade and exchange rate movements and depend on a narrow range of exports, other credit rules may need to be considered. In particular, if government expenditure were appropriate, a credit rule based on a small real yearly increase in "core" expenditures may be more justifiable.

—Inflation remains imbedded in most variable-rate countries at the two-digit level, mostly because of frequent devaluations without adequate monetary and fiscal restraint. Such inflationary pressures have persisted, even in those countries that reduced their budget deficits and eliminated bank financing of the deficits. Since external grants have become a major source of deficit financing for most variable-rate countries, there is a need to look into the inflationary pressures that external grants generate and the fiscal stance that would be consistent with their elimination.

REFERENCES

Baban, Roy C., and Joshua E. Greene, "The Export Performance of Sub-Saharan Africa, 1970–1990: A Survey," IMF Working Paper 92/55 (Washington: International Monetary Fund, 1992).

Bhatia, Rattan J., *The West African Monetary Union: An Analytical Review,* IMF Occasional Paper 35 (Washington: International Monetary Fund, 1985).

Blejer, Mario, and Ke-young Chu, eds., *Measurement of Fiscal Impact: Method-*

ological Issues, Occasional Paper No. 59 (Washington: International Monetary Fund, 1988).

Boughton, James M., "The CFA Franc Zone: Currency Union and Monetary Standard," IMF Working Paper 91/133 (Washington: International Monetary Fund, 1991).

Chu, Ke-young, "External Shocks and Fiscal Adjustment in Developing Countries: Experiences During 1962–82," IMF Working Paper 87/48 (Washington: International Monetary Fund, 1987).

Devarajan, Shantayanan, and Jaime De Melo, "Membership in the CFA Zone: Odyssean Journey or Trojan Horse," World Bank WPS 482 (Washington: World Bank, 1990).

———, "Adjustment with a Fixed Exchange Rate in Cameroon, Côte d'Ivoire, and Senegal," *World Bank Economic Review,* Vol. 1 (1987), pp. 447–87.

———, and Dani Rodrik, "Do the Benefits of Fixed Exchange Rates Outweigh Their Costs?" World Bank WPS 3727 (Washington: World Bank, 1991).

Easterly, William, and Klaus Schmidt-Hebbel, "The Macroeconomics of Public Sector Deficits," World Bank WPS 775 (Washington: World Bank, 1991).

Edwards, Sebastian, *Real Exchange Rates, Devaluation, and Adjustment* (Cambridge, Mass.: MIT Press, 1989).

Elbadawi, Ibrahim, and Nader Majd, "Fixed Parity of the Exchange Rate and Economic Performance in the CFA Zone," World Bank WPS 830 (Washington: World Bank, 1992).

Galbis, Vicente, "High Real Interest Rates Under Financial Liberalization: Is There a Problem?" IMF Working Paper 93/7 (Washington: International Monetary Fund, 1993).

Guillamont, Patrick, and Sylviane Guillamont, *Stratégies de Développement Comparées: Zone Franc et Hors Zone Franc* (Paris: *Economica,* 1988).

———, and Patrick Plane, "Participating in African Monetary Unions: An Alternative Evaluation," *World Development,* Vol. 16 (May 1988), pp. 569–96.

Heller, Peter, and Jack Diamond, *International Comparisons of Government Expenditure Revisited: The Developing Countries, 1975–86,* Occasional Paper No. 69 (Washington: International Monetary Fund, 1990).

Honohan, P., "Monetary Cooperation in the CFA Zone," World Bank WPS 389 (Washington: World Bank, 1990).

Hewitt, Daniel, "Military Expenditure: International Comparison of Trends," IMF Working Paper 91/54 (Washington: International Monetary Fund, 1991).

Levy, Victor, and John L. Newman, "Wage Rigidity: Micro and Macro Evidence on Labor Market Adjustment in the Modern Sector," *The World Bank Economic Review,* Vol. 3 (January 1989), pp. 97–117.

Nashashibi, Karim, Henri Lorie, and Anne Doizé, "Fiscal Deficits in the UMOA Countries: Analysis and Possibilities for Reform" (unpublished; Washington: International Monetary Fund, 1990).

———, and others, *The Fiscal Dimension of Adjustment in Low-Income Countries,* Occasional Paper No. 95 (Washington: International Monetary Fund, April 1992).

———, "Externally Financed Budget Deficits in Developing Countries" (unpublished; Washington: International Monetary Fund, November 1989).

Neurrisse, André, *Le Franc C.F.A.* (Paris: Librairie Générale de Droit et de Jurisprudence, 1987).

Sahn, David E., "Public Expenditures in Sub-Saharan Africa During a Period of Economic Reforms," *World Development,* Vol. 20 (1992), pp. 673–93.

Schelizi, Z., V. Ghandi, and J. Ehdaie, "Patterns of Taxation in Sub-Saharan Africa: Trends in Tax Effort and Composition During the Period 1966–81," CPD Discussion Paper No. 1985–48 (Washington: World Bank, 1985).

Secrétariat du comité monétaire de la zone franc, *La Zone franc* (Paris: Secrétariat du comité de la zone franc, several annual issues).

Stymne, J., "Debt Growth and the Prospects for Debt Reduction: The Case of Sub-Saharan African Countries," IMF Working Paper 89/71 (Washington: International Monetary Fund, 1989).

Svedberg, Peter, "The Export Performance of Sub-Saharan Africa 1970–1985," Seminar Paper No. 409 (Stockholm: Institute for International Economic Studies, 1988).

Tanzi, Vito, "Inflation, Lags in Collection, and the Real Value of Tax Revenue," *Staff Papers,* International Monetary Fund, Vol. 24 (1977), pp. 154–67.

———, "Fiscal Policy Responses to Exogenous Shocks in Developing Countries," *American Economic Review,* Vol. 76 (1986), pp. 88–91.

———, "Quantitative Characteristics of the Tax System of Developing Countries," in *The Theory of Taxation for Developing Countries,* ed. by D. Newbery and N. Stern (Oxford: Oxford University Press, for the World Bank, 1987).

———, "Structural Factors and Tax Revenue in Developing Countries: A Decade of Evidence," *Open Economies: Structural Adjustment and Agriculture,* ed. by I. Goldin and A. Winters (Cambridge: Cambridge University Press, 1992).

———, *Taxation in Sub-Saharan Africa,* Occasional Paper No. 8 (Washington: International Monetary Fund, 1981).

———, "The Impact of Macroeconomic Policies on the Level of Taxation in Developing Countries," *Staff Papers,* International Monetary Fund, Vol. 36 (1989), pp. 633–56.

Vizy, Marc, *La Zone Franc* (Paris: C.H.E.A.M., 1989).

World Bank, "Adjustment in Africa: Reforms, Results, and the Road Ahead," A World Bank Policy Research Report (Oxford: Oxford University Press, 1994).

———, *Sub-Saharan Africa: From Crisis to Sustainable Growth* (Washington: World Bank, 1989).

———, *World Debt Tables 1991–92: External Debt of Developing Countries* (Washington: World Bank, 1992).

Young, Crawford, "Africa's Colonial Legacy," in *Strategies for African Development,* ed. by R.J. Berg and J.S. Whittaker (Berkeley: University of California Press, 1986).

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

The Stabilizing Effect of the ERM on Exchange Rates and Interest Rates

Some Nonparametric Tests

MICHAEL J. ARTIS and MARK P. TAYLOR*

This paper uses nonparametric procedures to test for a shift in the volatility of nominal and real exchange rates for members and nonmembers of the ERM. The results imply a reduction in volatility for ERM members, especially during the latter half of its operation. We also demonstrate that this enhanced stability was not bought at the expense of increased interest rate volatility. The issue of interest rate volatility during the British pound's participation in the ERM is also examined. [JEL F31, F42]

EVENTS SINCE THE so-called "Black Wednesday" of September 16, 1992, have caused some observers to doubt the future of the exchange rate mechanism (ERM) of the European Monetary System (EMS). The speculative spasm that seized the markets in 1992 precipitated the floating of both the lira and the pound sterling (the former "inside" and the latter "outside" the ERM) and caused the first devaluation of the Spanish peseta. Since that time several other currencies—the Portuguese escudo, the Irish punt, the Danish kroner, and the French franc—have come under intermittent, and at times intense, speculative pressure.

* Michael J. Artis is Professor of Economics at the University of Manchester and Mark P. Taylor is a Senior Economist in the IMF's Research Department. This paper was completed while Michael Artis was a Visiting Scholar in the Research Department. The authors are grateful to a number of IMF colleagues and two anonymous referees for comments on a previous version of the paper.

Whether and how the ERM might be reconstituted to insulate it from such pressures is still a controversial issue. Although worries concerning substitutability among member currencies and the long-term viability of the ERM had been voiced earlier (Artis and Taylor (1989)), few argue that the ERM has not succeeded at stabilizing volatility in member exchange rates. This paper provides empirical support for that presumption and examines some additional propositions where the presumption is less clear. Thus, we examine whether what is true of intra-ERM exchange rates is also true of exchange rates between ERM members and nonmembers, and we look at whether the stabilizing effect of the ERM holds for real exchange rates. In addition, because stable exchange rates might be bought at the cost of unstable interest rates, we explicitly examine whether this is true.

Further, the EMS, far from being a chronologically homogeneous regime, has evolved over time. In particular, Giavazzi and Spaventa (1990) observe that before the disturbances the system was exceptionally stable. We examine this proposition and confirm that the stabilizing effect of the ERM became more marked over this period of time. Given the recent intense pressures within the ERM, however, we also investigate the volatility in interest rates during the United Kingdom's participation in the mechanism. The core of this paper is the application of a nonparametric method for testing whether the ERM had a stabilizing effect; this is in contrast to the large number of studies that employ parametric methods. There are good reasons, as we explain later, for our nonparametric approach.

I. Provisions of the EMS

The European Monetary System was instituted in March 1979, largely as a reaction to the volatility of exchange rates among European countries during the years following the breakdown of the Bretton Woods system. At the heart of the EMS is the exchange rate mechanism. Countries participating in the ERM agree to maintain their bilateral exchange rates within ±2.25 percent of a central rate. In an exception, Italy negotiated a temporarily wider band of ±6 percent at the outset of the system and moved to the narrow band in January 1990, an example subsequently followed by the United Kingdom, Spain, and Portugal when these countries entered the ERM. (Spain and Portugal are still working within this broader band while the United Kingdom now lets the pound float.) Spain entered the ERM in June 1989; the United Kingdom in October 1990; and Portugal in April 1992.

Maintenance of the bilateral rates within the bands is a symmetrical obligation: the strong-currency country is equally obliged to prevent its currency from piercing the ceiling as the weak-currency country is to prevent it from falling through the floor. Resources to support these obligations through market intervention are made available through credit lines, notably through the Very Short Term Financing Facility. The Basle-Nyborg agreements of 1987 strengthened the ERM by expanding these credit lines, lengthening the repayment period associated with them and establishing the "presumption" that credit would be available for intervention within the band (so-called "intramarginal intervention"), whereas previous provisions allowed only for intervention at the limits of the band ("marginal intervention").[1] The Basle-Nyborg agreement also called for a strengthening of cooperative monetary policy.

The central rates in each bilateral band can be changed in a realignment. Table 1 shows the timing and extent of such realignments before the speculative surge of September 1992.[2] As can be seen, the use of realignments declined rapidly after the early 1980s.

Although provisions concerning bilateral rates form the core of the ERM, the system is formally organized around a composite currency, the European currency unit (ECU); central rates for participating currencies are expressed in terms of ECU. While it is purely a formal construction, the ECU allowed for an interesting technical innovation in the EMS—the divergence indicator and threshold positions. According to these provisions, when the ECU value of a currency reaches 75 percent of its band limit relative to other currencies, a presumption is created that the country concerned should take corrective action involving monetary and fiscal policy or a realignment. This technical provision—also called the divergence threshold—provides an early warning of wide bilateral movements and, more important, isolates the errant currency.

There seems to have been little doubt in the minds of those who constructed these provisions that the errant currency would be the

[1] Although this presumption was established, the financing central bank could still, in principle, object. Other special conditions were also attached to intramarginal interventions under the Basle-Nyborg agreements, relating to credit limits and currency of repayment.

[2] The lira was formally devalued (by 7 percent) on September 12, 1992, but commenced a float soon afterwards on September 17, following the example of sterling which was floated from September 16. The peseta was devalued on September 17, by 5 percent and again, by a further 6 percent, on November 22, at which time the escudo was devalued by a similar amount. The punt was devalued by 10 percent on January 30, 1993; on May 13 the peseta was devalued by a further 8 percent with a parallel devaluation of the escudo of 6.5 percent.

Table 1. *Changes in EMS Central Rates Prior to September 1992*

	Dates of realignments											
	9/24 1979	11/30 1979	3/22 1981	10/5 1981	2/22 1982	6/14 1982	3/21 1983	7/21 1985	4/7 1986	8/4 1986	1/12 1987	1/8 1990
Percentage change in parity												
Belgian franc	0.0	0.0	0.0	0.0	−8.5	0.0	+1.5	+2.0	+1.0	0.0	+2.0	0.0
Danish kroner	−2.9	−4.8	0.0	0.0	−3.0	0.0	+2.5	+2.0	+1.0	0.0	0.0	0.0
German mark	+2.0	0.0	0.0	+5.5	0.0	+4.25	+5.5	+2.0	+3.0	0.0	+3.0	0.0
French franc	0.0	0.0	0.0	−3.0	0.0	−5.75	−2.5	+2.0	−3.0	0.0	0.0	0.0
Irish punt	0.0	0.0	0.0	0.0	0.0	0.0	−3.5	+2.0	0.0	−8.0	0.0	0.0
Italian lira	0.0	0.0	−6.0	−3.0	0.0	−2.75	−2.5	−6.0	0.0	0.0	0.0	−3.7
Dutch guilder	0.0	0.0	0.0	+5.5	0.0	+4.25	+3.5	+2.0	+3.0	0.0	+3.0	0.0

deutsche mark.[3] It is one of the curiosities of the ERM that the mark has *not* often been at the higher end of its permitted range and that, for most of its operation, the ERM's anti-inflationary stance has been so strong that the mark was not singled out (Padoa-Schioppa (1983)), and the mark naturally evolved as the anchor of the system.

It is important to note that the introduction of the ERM did not require the abolition of exchange controls nor were controls over capital movements initially removed, notably in France and Italy (although these countries did abolish remaining restrictions on capital movements in January and May 1990 respectively, ahead of the deadline of July 1 imposed by a European Community directive). Artis and Taylor (1988) provide evidence—the offshore-onshore interest rate differentials for the mark and lira after March 1979—that these controls were used substantially. The controls may have fostered stability by giving the authorities the whip hand in negotiating realignments and by avoiding the immediate convergence of monetary policy, which the freedom from capital controls, coupled with the obligation to defend central bilateral parities, would have implied.

The immediate objective, then, of the ERM has been the stabilization of bilateral nominal exchange rates among its members. Although the ERM in its early phase of operation was used to restrain exchange rate "overshooting"—preserving the competitiveness of participant countries through frequent realignment—the system also evolved into a counter-inflationary framework. Without convergent inflation rates, real exchange rate stability does not necessarily follow from stable nominal rates. Yet, it can be argued that, like a customs union, the EMS must have an "inner rationale" of maintaining broadly stable conditions of competitiveness. Otherwise, the purpose of reducing exchange rate protection will be questioned. It is just as important, therefore, to explore real exchange rate stabilization as it is to test for nominal rate stabilization.

A further issue involves the stability of EMS exchange rates vis-à-vis outside currencies. To the extent that the ERM encourages greater coherence among the partner currencies, the more inevitable it becomes that the currencies vis-à-vis third currencies will become more homogeneous. To take an example that is not entirely fanciful, if the dollar-mark rate is exceptionally volatile, some of the mark's volatility against the dollar will be imparted to the franc. (To the extent to which this does *not* happen, tensions within the ERM will be caused by shifts of sentiment

[3] Ludlow (1982) gives a detailed and informative account of the negotiations leading to the institution of the EMS. Van Ypersele (1985) provides a more detailed account of the institutional features of the System.

about the dollar, which would exert pressure on the mark-franc exchange rate.) Thus, we also explore whether the reduced volatility among intra-EMS parities has been offset by increased volatility in extra-EMS parities for some European currencies.

II. Volatility in the EMS

Before the events of September 1992, there had been 12 realignments in the EMS. This fact, together with the fact that wide variations are allowed by the margins and that less than full convergence of inflation has occurred among ERM members (Masson and Taylor (1992)), calls into question whether the EMS actually does induce greater stability in either the nominal or the real exchange rate.

The difference, stressed by John Williamson (1985), between the concepts of exchange rate *volatility* and *misalignment* is important here. Volatility is a "high frequency" concept referring to movements in the exchange rate over comparatively short periods of time. Misalignment, on the other hand, refers to the capacity of an exchange rate to depart from its fundamental equilibrium value (however defined) over a protracted period of time. In a world of risk-neutral producers and consumers, it can be shown that higher exchange volatility actually enhances overall welfare (see De Grauwe (1992), pp. 64–67), but this conclusion becomes more debatable when allowance is made for risk aversion and incomplete forward markets. A related question concerns the effect of ERM membership on interest rates. If the union is successful in generating a convergence of interest rates toward those of the low-inflation anchor, there may be positive welfare gains since higher interest rates may generate important principal-agent problems in domestic capital markets (Stiglitz and Weiss (1981)). Also, some authors (Baldwin (1989), European Commission (1990)) have argued—using an endogenous growth model framework—that lower interest rates may generate permanently higher growth for GDP.[4]

With respect to the effects of exchange rate volatility on trade flows, the evidence is mixed. A study of Akhtar and Hilton (1984) found a negative correlation between exchange rate volatility and U.S.-German trade flows, but a comparable study by the International Monetary Fund (1983) failed to confirm this finding for other trade flows, time periods, and volatility measures. Cushman (1986), however, does find volatility

[4] In fact, these authors apply these arguments to the prospect of a single European currency, but similar arguments apply in the present case.

effects on trade when "third country" effects are controlled (for example, dollar-mark volatility may affect U.S.-U.K. trade).

Despite these caveats, a number of studies have concentrated on evidence that the EMS has reduced exchange rate volatility, most notably those by Ungerer, Evans, and Nyberg (1983), Ungerer and others (1986, 1990), the European Commission (1982), Padoa-Schioppa (1983), Rogoff (1985), and Artis and Taylor (1988). There are many possible approaches to this question—the choice of exchange rates (bilateral, effective, nominal, or real), data frequency (daily, weekly, monthly, or quarterly), the standard against which stability is to be judged (the level or change in exchange rates, conditional or unconditional), and the precise statistical measure chosen (standard deviation). Then, there is the question of the counterfactual—supplied in these studies and others like them by the behavior of a control group of non-EMS currencies. Without exception, the EMS in these studies has been judged as having improved the stability of intra-EMS bilateral exchange rates, although the improvement is less marked for effective rates.

III. Some Nonparametric Volatility Tests

Many of the studies that have tested for a post-March 1979 downward shift in exchange rate volatility have relied on purely descriptive statistics. As such, they can only suggest results, and it is difficult to assess the performance of the EMS from this evidence. The more straightforward approach to the problem, namely estimating a specific parameterization of the volatility and testing for a structural shift, is fraught with pitfalls. This is because economists are uncertain about the statistical distribution of exchange rate changes.

It is a stylized fact that percentage exchange rate changes tend to follow leptokurtic (fat-tailed, highly peaked) distributions. Westerfield (1977), for example, finds that the stable paretian distribution with characteristic exponent less than two fits the change in the logarithm of spot exchange rates better than the normal distribution. In a similar vein, Rogalski and Vinso (1977) suggest Student's t-distribution as a good approxmiation. It may well be that the distribution of exchange rate changes is normal but that the variance shifts through time—perhaps according to the amount of "news." This would give the appearance of a stable leptokurtic distribution. Some evidence for such behavior is provided by Boothe and Glassman (1987), who find that mixtures of normal distributions provide some of the best fits.

We wish to stress the importance of the distributional properties of

exchange rate changes in any volatility study. Studies that rely on simple variance measures implicitly invoke a normality assumption, the legitimacy of which is being challenged by a growing number of studies (see Boothe and Glassman (1987) for additional references). For example, it is conceivable that exchange rate changes at a certain frequency have a Cauchy distribution, for which no finite moments of any order exist.

To circumvent some of these problems, we employ nonparametric tests for volatility shifts, which do not require estimation of the distributional parameters. Instead, exchange rate changes are ranked by size, and inferences are drawn with respect to the shape of the ranking. Intuitively, if a significant number of lower-ranked percentage changes are recorded in the latter half of the sample period, a reduction in volatility would be indicated. Note, however, that although the test procedure is nonparametric in the sense that no volatility measures are actually *estimated*, we still must choose an appropriate distribution for exogenous disturbances. This issue is discussed later in this section.

In the nonparametric technique we employ, let Δe_t be the change in the logarithm of the exchange rate at time t. Then, the hypothesis is

$$\Delta e_t = \mu + \sigma_t \epsilon_t, \tag{1}$$

$$\sigma_t = \exp(\alpha + \beta z_t), \tag{2}$$

where μ, α, and β are unknown constant scalars, ϵ_t is independently and identically distributed with distribution function F and density function f, and z_t is a binary variable reflecting the hypothesized shift in volatility at time $N + 1$. Thus,

$$z_t = \begin{cases} 1, & t \leq N \\ 0, & t > N \end{cases}$$

Given equation (1), the null hypothesis of no shift in volatility is

$$H_o : \beta = 0 \tag{3}$$

Hajek and Sidak (1967) develop a number of nonparametric tests for dealing with this kind of framework, which under appropriate regularity conditions are locally most powerful (Hajek and Sidak (1967), pp. 70–71). The test statistics take the form

$$\zeta = \sum_{t=1}^{T} (z_t - \bar{z})\, \alpha(u_t), \tag{4}$$

where $\bar{z}$ is the arithmetic mean of the z_t sequence of T observations and u_t is defined as follows. Let $r()$ be the rank of Δe_t—that is, $r(\Delta e_i)$ is the

$r(\Delta e_i)^{\text{th}}$ smallest absolute change in the total sequence considered—then $u_t = r(\Delta e_t)/(T + 1)$.

Clearly, u_t must lie in the closed interval $[1/(T + 1), T/(T + 1)]$, for no ties in rank. The function $\alpha(\cdot)$ in equation (4) is a score function defined by Hajek and Sidak that depends upon the assumed density of ϵ_t. Hajek and Sidak define a class of functions that can be used in place of the score function in large samples, since $\alpha(\cdot)$ may be difficult to evaluate in practice. If F is the assumed distribution function of ϵ_t,

$$F(x) = \int_{-\infty}^{x} f(u)\, du, \tag{5}$$

and if $F^{-1}(u)$ is the inverse of F,

$$F^{-1}(u) = \text{Infimum}\ \{x \mid F(x) \geq u\}, \tag{6}$$

then the asymptotic score function, $\psi(\cdot)$, is defined (Hajek and Sidak (1967), p. 19)

$$\psi : (0, 1) \rightarrow \mathcal{R}, \tag{7}$$

$$\psi(u) = -F^{-1}(u) \left[\frac{f'\{F^{-1}(u)\}}{f\{F^{-1}(u)\}} \right] - 1.0. \tag{8}$$

Under equation (1), the statistic

$$\eta = \sum_{t=1}^{T} (z_t - \bar{z})\, \psi(u_t) \tag{9}$$

($\alpha(\cdot)$ in equation (4) is replaced by $\psi(\cdot)$) will be asymptotically normally distributed.

Under the null hypothesis (3), η will have a zero mean and variance ρ^2 given by (Hajek and Sidak (1967), pp. 159–160)

$$\rho^2 = \left\{ \sum_{t=1}^{T} (z_t - \bar{z})^2 \right\} \int_0^1 \left\{ \psi(u) - \bar{\psi} \right\}^2 du, \tag{10}$$

where

$$\bar{\psi} = \int_0^1 \psi(u)\, du.$$

Thus, for a given choice of f, the statistic (η/ρ) will be asymptotically standard normal under the null hypothesis of no shift in volatility. Significantly negative values of η reflect a negative value for β in equation (2)—an increase in volatility after the shift point—while significantly positive values of η imply a reduction in volatility.

As noted at the outset of this section, although the test procedure is nonparametric in the sense that no volatility measures are actually estimated, we cannot avoid choosing an appropriate distribution for ϵ_t in implementing the procedure. To minimize the risk of choosing an inappropriate distribution, we selected four well-known ones in the belief that the true distribution will be close to one of them. If qualitatively similar nonparametric results are obtained for a range of assumed distributions, then the results may be said to be robust to this uncertainty. The densities correspond to the normal, logistic, double exponential, and Cauchy distributions. The density and asymptotic score functions (as defined in equation (8)) for these distributions are given in the appendix. All of the chosen distributions are symmetric, and both the double exponential and Cauchy distributions have fat tails.[5]

IV. Data

The data we use are monthly data on bilateral U.S. dollar exchange rates and on nominal effective exchange rates, both taken from the *International Financial Statistics* data tape for the period January 1973 through October 1990, and on Eurocurrency interest rates (three-month maturity) for the period January 1975 through February 1993. The end point of the exchange rate sample coincides with Britain's entry into the ERM: this choice of sample allows us to use sterling as a representative non-ERM currency. Bilateral rates against the German mark and the pound sterling were also constructed by assuming a triangular arbitrage condition. Real exchange rates were constructed by deflating by the wholesale price relatives (data also from the IFS tape).[6] The currencies used included three ERM currencies—German mark, French franc, and Italian lira—and three non-ERM currencies—U.S. dollar, pound sterling, and Japanese yen. We also obtained monthly data on three-month Eurodeposit interest rates to test the hypothesis that exchange rate fixity may impart interest rate volatility. All results are for shifts in the volatility of monthly changes.

[5] Another relevant distribution would have been Student's *t*. However, the score function (8) for this distribution would have been very difficult to compute. A possibility not considered is that there was a change in distribution of ERM exchange rate changes post-March 1979 (e.g., shifted from normal to Cauchy). Tests for this kind of behavior could conceivably be based on likelihood ratios, although one might suspect that the discriminatory power of such procedures would be low.

[6] Wholesale prices were used as a proxy for tradable goods prices.

V. Overall ERM Effect

In the first set of tests, we look for a post-March 1979 shift in exchange rate volatility. In some sense, we expect such tests to be affected by the behavior of the dollar during the 1980s, which was highly volatile. Thus, if intra-ERM exchange rates displayed reduced volatility during a period when most exchange rates against the dollar were highly volatile, this would be strong evidence of an overall ERM effect.

Nominal Exchange Rates

As expected, the nonparametric tests reveal a significant reduction in the volatility of ERM currencies against the mark after March 1979 (Table 2), while the volatility of the mark against the pound, yen, and dollar appears unchanged.[7] Table 2 also shows that although the dollar-lira exchange rate became more volatile after March 1979, this is not the case for the dollar-franc and dollar-mark exchange rates. (The test statistics, however, are uniformly negative, suggesting a tendency toward increased volatility.) The volatilities of the dollar-pound and dollar-yen exchange rates have risen significantly since March 1979.

Finally, Table 2 reveals an unequivocal reduction in the volatility of nominal effective ERM exchange rates after March 1979. Exactly the converse is true of the U.S. dollar nominal effective rate. The results for the effective pound and yen rates, although significant in only one case, are nevertheless uniformly negative and large in absolute size, indicating a tendency toward increased volatility.

Real Exchange Rates

Table 3 shows a marked and significant reduction in the volatility of the mark-lira rate after March 1979, and a similar—albeit statistically insignificant—reduction in the mark-franc rate (as indicated by the large, positive values of the test statistics). The real exchange rates of the mark against the non-ERM countries do not exhibit a significant shift (Table 3).

[7] As derived above, the test statistic will be asymptotically standard normal (that is, with a mean of zero and variance of unity) under the null hypothesis of no shift in volatility. Significantly negative values of the test statistic (less than about -2) reflect a negative value for β in (2)—i.e., an increase in volatility after the shift point—while significantly positive values (bigger than about $+2$) imply a reduction in volatility.

Table 2. *Test Statistics for a Shift in Nominal Exchange Rate Volatility* (1973:1–1979:3 vs. 1979:4–1990:10)

Exchange rate	Normal	Logistic	Double exponential	Cauchy
Mark nominal exchange rate				
Mark–franc	6.405 (0.000)	5.396 (0.000)	5.272 (0.000)	5.399 (0.000)
Mark–lira	7.563 (0.000)	6.118 (0.000)	5.942 (0.000)	5.093 (0.000)
Mark–pound	0.222 (0.824)	0.135 (0.893)	0.103 (0.918)	−0.159 (0.874)
Mark–yen	−0.498 (0.619)	−0.444 (0.657)	−0.451 (0.652)	−0.546 (0.585)
Mark–dollar	−0.502 (0.616)	−0.685 (0.493)	−0.644 (0.520)	−1.310 (0.190)
Dollar nominal exchange rate				
Dollar–franc	−0.935 (0.350)	−1.098 (0.272)	−1.323 (0.186)	−2.869 (0.004)
Dollar–lira	−1.835 (0.067)	−1.929 (0.054)	−2.058 (0.040)	−3.770 (0.000)
Dollar–mark	−0.604 (0.546)	−0.755 (0.450)	−0.720 (0.471)	−1.361 (0.173)
Dollar–pound	−2.625 (0.009)	−2.171 (0.030)	−2.189 (0.029)	−2.830 (0.017)
Dollar–yen	−1.956 (0.050)	−2.088 (0.037)	−2.179 (0.029)	−4.286 (0.000)
Nominal effective exchange rate				
Franc	2.720 (0.007)	2.099 (0.036)	1.941 (0.052)	1.148 (0.251)
Lira	3.143 (0.002)	2.225 (0.026)	2.144 (0.032)	0.637 (0.524)
Mark	3.077 (0.022)	2.356 (0.018)	2.294 (0.022)	1.419 (0.156)
Pound	−1.694 (0.090)	−1.635 (0.102)	−1.717 (0.086)	−2.634 (0.008)
Yen	−1.531 (0.126)	−1.120 (0.263)	−1.101 (0.271)	−0.458 (0.647)
Dollar	−3.006 (0.003)	−2.670 (0.008)	−2.738 (0.006)	−3.652 (0.000)

Note: Figures in parentheses denote marginal, two-sided significance levels. All test statistics are distributed as standard normal under the null hypothesis of no shift in volatility. Significantly positive test statistics indicate a reduction in volatility after the break point; significantly negative statistics indicate the converse.

Table 3. *Test Statistics for a Shift in Real Exchange Rate Volatility* (1973:1–1979:3 vs. 1979:4–1990:10)

Exchange rate	Normal	Logistic	Double exponential	Cauchy
Mark real exchange rate				
Mark–franc	1.652 (0.099)	1.469 (0.142)	1.492 (0.136)	1.993 (0.046)
Mark–lira	7.390 (0.000)	6.159 (0.000)	6.088 (0.000)	6.170 (0.000)
Mark–pound	0.502 (0.615)	0.313 (0.754)	0.287 (0.774)	−0.154 (0.877)
Mark–yen	0.006 (0.996)	−0.023 (0.981)	−0.017 (0.986)	0.030 (0.976)
Mark–dollar	0.314 (0.754)	−0.041 (0.967)	−0.005 (0.996)	−0.914 (0.361)
Real effective exchange rate				
Franc	3.543 (0.000)	2.916 (0.004)	2.860 (0.004)	2.563 (0.010)
Lira	4.780 (0.000)	3.872 (0.000)	3.814 (0.000)	3.573 (0.000)
Mark	1.232 (0.218)	1.019 (0.308)	0.985 (0.325)	1.001 (0.317)
Pound	−0.808 (0.419)	−0.773 (0.440)	−0.830 (0.407)	−1.200 (0.230)
Yen	−1.390 (0.164)	−1.017 (0.309)	−0.948 (0.343)	−0.365 (0.715)
Dollar	−4.273 (0.000)	−3.703 (0.000)	−3.735 (0.000)	−4.641 (0.000)

Note: Figures in parentheses denote marginal, two-sided significance levels. All test statistics are distributed as standard normal under the null hypothesis of no shift in volatility. Significantly positive test statistics indicate a reduction in volatility after the break point; significantly negative statistics indicate the converse.

Table 3 also shows the results of the tests applied to real effective exchange rates. Strong and significant reductions in volatility are indicated for the franc and the lira, and positive statistics were also recorded for the mark. For the pound and the yen, the test statistics are negative but insignificant, while the real effective rate of the dollar shows a strongly signficant rise in volatility after March 1979.

An Overall ERM Effect?

Overall, the results reported in this section indicate that the ERM has reduced exchange rate volatility—both real and monimal—since March 1979. This is particularly impressive in light of evidence that the volatility of non-ERM exchange rates—particularly the dollar—has *risen* over the same period.

VI. A New EMS?

A number of commentators (including Giavazzi and Spaventa (1990)) have noted a shift in the nature of the EMS toward less frequent realignments and more concerted action toward internal adjustment. It is not entirely clear when this shift began, and thus we tested for a shift using two different subperiods of ERM operation. We first tested for a shift in exchange rate volatility after the realignment of March 1983 (Tables 4 and 5). The results indicate a downward shift in the volatility of ERM exchange rates—real and nominal, bilateral and effective. For the dollar and the pound, however, there is little shift in volatility, although yen exchange rates appear to have become more stable over this period. The second subperiod begins after the realignment of April 1986. The results of testing for a shift in volatility after this realignment (Tables 6 and 7) are broadly comparable to those reported for the first subsample—although yen volatility appears lower.

We therefore conclude that the hypothesis that a "new" and harder EMS arose after 1982, in which greater emphasis was given to harmonizing the macroeconomic objectives of EMS members (Giavazzi and Spaventa (1990)).

VII. Volatility Transfer

It is sometimes argued that advanced macroeconomic systems naturally generate a "lump of uncertainty" that can be pushed out of the economy at one point only to reappear somewhere else (see Bachelor

Table 4. *Test Statistics for a Shift in Nominal Exchange Rate Volatility* (1979:4–1983:3 vs. 1983:4–1990:10)

Exchange rate	Normal	Logistic	Double exponential	Cauchy
Mark nominal exchange rate				
Mark–franc	2.200 (0.028)	1.694 (0.090)	1.688 (0.091)	1.139 (0.255)
Mark–lira	2.326 (0.020)	1.934 (0.053)	1.969 (0.049)	2.087 (0.037)
Mark–pound	1.671 (0.095)	1.339 (0.818)	1.337 (0.818)	1.203 (0.229)
Mark–yen	3.224 (0.001)	2.636 (0.008)	2.648 (0.008)	2.626 (0.009)
Mark–dollar	−0.564 (0.573)	−0.500 (0.617)	−0.481 (0.631)	−0.528 (0.598)
Dollar nominal exchange rate				
Dollar–franc	0.755 (0.450)	0.572 (0.567)	0.623 (0.533)	0.706 (0.480)
Dollar–lira	0.165 (0.869)	0.027 (0.978)	0.107 (0.915)	−0.005 (0.996)
Dollar–mark	−0.228 (0.819)	−0.292 (0.771)	−0.255 (0.799)	−0.467 (0.640)
Dollar–pound	0.037 (0.970)	−0.026 (0.979)	−0.038 (0.970)	−0.439 (0.661)
Dollar–yen	1.899 (0.058)	1.582 (0.114)	1.594 (0.111)	1.512 (0.131)
Nominal effective exchange rate				
Franc	0.951 (0.341)	0.748 (0.455)	0.797 (0.425)	0.760 (0.448)
Lira	1.340 (0.180)	1.074 (0.283)	1.025 (0.306)	0.859 (0.390)
Mark	1.740 (0.082)	1.196 (0.232)	1.152 (0.249)	0.201 (0.841)
Pound	0.696 (0.486)	0.473 (0.636)	0.407 (0.684)	0.027 (0.978)
Yen	1.460 (0.144)	1.241 (0.215)	1.269 (0.204)	1.552 (0.121)
Dollar	0.294 (0.769)	0.158 (0.875)	0.152 (0.879)	−0.258 (0.796)

Note: Figures in parentheses denote marginal, two-sided significance levels. All test statistics are distributed as standard normal under the null hypothesis of no shift in volatility. Significantly positive test statistics indicate a reduction in volatility after the break point; significantly negative statistics indicate the converse.

Table 5. *Test Statistics for a Shift in Real Exchange Rate Volatility*
(1979:4–1983:3 vs. 1983:4–1990:10)

Exchange rate	Normal	Logistic	Double exponential	Cauchy
Mark real exchange rate				
Mark–franc	1.731 (0.083)	1.337 (0.181)	1.283 (0.200)	0.658 (0.511)
Mark–lira	3.976 (0.000)	3.453 (0.001)	3.388 (0.001)	3.923 (0.000)
Mark–escudo	3.751 (0.000)	2.854 (0.004)	2.699 (0.007)	1.414 (0.157)
Mark–pound	1.809 (0.070)	1.426 (0.154)	1.407 (0.160)	1.209 (0.227)
Mark–yen	3.397 (0.001)	2.790 (0.005)	2.820 (0.005)	2.703 (0.007)
Mark–dollar	−0.417 (0.679)	−0.404 (0.687)	−0.394 (0.693)	−0.536 (0.592)
Real effective exchange rate				
Franc	2.636 (0.008)	2.177 (0.029)	2.169 (0.030)	2.269 (0.023)
Lira	3.985 (0.000)	3.433 (0.001)	3.313 (0.001)	3.489 (0.001)
Mark	3.185 (0.001)	2.611 (0.009)	2.569 (0.010)	2.562 (0.010)
Pound	1.143 (0.253)	0.849 (0.396)	0.823 (0.410)	0.432 (0.666)
Yen	2.350 (0.019)	1.947 (0.052)	1.977 (0.048)	2.155 (0.031)
Dollar	0.362 (0.717)	0.261 (0.794)	0.273 (0.784)	0.028 (0.978)

Note: Figures in parentheses denote marginal, two-sided significance levels. All test statistics are distributed as standard normal under the null hypothesis of no shift in volatility. Significantly positive test statistics indicate a reduction in volatility after the break point; significantly negative statistics indicate the converse.

Table 6. *Test Statistics for a Shift in Nominal Exchange Rate Volatility* (1979:4–1986:5 vs. 1986:6–1990:10)

Exchange rate	Normal	Logistic	Double exponential	Cauchy
Mark nominal exchange rate				
Mark–franc	0.248 (0.804)	0.105 (0.917)	0.083 (0.934)	−0.362 (0.717)
Mark–lira	1.575 (0.115)	1.201 (0.230)	1.312 (0.190)	1.067 (0.286)
Mark–pound	1.474 (0.140)	1.179 (0.238)	1.064 (0.288)	0.691 (0.489)
Mark–yen	1.431 (0.153)	1.034 (0.301)	1.035 (0.301)	0.471 (0.638)
Mark–dollar	−0.175 (0.861)	−0.306 (0.759)	−0.295 (0.768)	−0.698 (0.485)
Dollar nominal exchange rate				
Dollar–franc	0.912 (0.362)	0.647 (0.517)	0.695 (0.487)	0.466 (0.641)
Dollar–lira	−0.156 (0.876)	−0.266 (0.790)	−0.252 (0.801)	−0.559 (0.576)
Dollar–mark	−0.177 (0.860)	−0.277 (0.782)	−0.276 (0.782)	−0.587 (0.557)
Dollar–pound	−0.119 (0.905)	−0.119 (0.905)	−0.096 (0.924)	−0.209 (0.834)
Dollar–yen	−0.456 (0.649)	−0.467 (0.641)	−0.428 (0.669)	−0.927 (0.354)
Nominal effective exchange rate				
Franc	2.200 (0.028)	1.829 (0.067)	1.868 (0.062)	2.104 (0.035)
Lira	0.974 (0.330)	0.798 (0.425)	0.767 (0.443)	0.721 (0.471)
Mark	2.388 (0.017)	2.025 (0.043)	2.021 (0.043)	2.349 (0.019)
Pound	0.818 (0.413)	0.569 (0.569)	0.463 (0.643)	−0.076 (0.939)
Yen	−0.811 (0.417)	−0.789 (0.430)	−0.782 (0.434)	−1.282 (0.200)
Dollar	1.041 (0.298)	0.777 (0.437)	0.775 (0.438)	0.433 (0.665)

Note: Figures in parentheses denote marginal, two-sided significance levels. All test statistics are distributed as standard normal under the null hypothesis of no shift in volatility. Significantly positive test statistics indicate a reduction in volatility after the break point; significantly negative statistics indicate the converse.

Table 7. *Test Statistics for a Shift in Real Exchange Rate Volatility* (1979:4–1986:5 vs. 1986:6–1990:10)

Exchange rate	Normal	Logistic	Double exponential	Cauchy
Mark real exchange rate				
Mark–franc	1.290 (0.197)	1.070 (0.285)	1.102 (0.270)	1.093 (0.274)
Mark–lira	2.096 (0.036)	1.728 (0.084)	1.679 (0.093)	1.612 (0.107)
Mark–pound	1.442 (0.149)	1.111 (0.267)	1.032 (0.302)	0.596 (0.551)
Mark–yen	1.751 (0.080)	1.359 (0.174)	1.437 (0.151)	1.144 (0.253)
Mark–dollar	−0.400 (0.689)	−0.532 (0.595)	−0.521 (0.602)	−1.177 (0.239)
Real effective exchange rate				
Franc	2.706 (0.007)	2.231 (0.026)	2.193 (0.028)	2.176 (0.030)
Lira	4.720 (0.000)	3.975 (0.000)	3.892 (0.000)	3.857 (0.000)
Mark	2.415 (0.016)	1.996 (0.046)	1.900 (0.057)	1.767 (0.077)
Pound	1.472 (0.141)	1.157 (0.247)	1.121 (0.262)	0.889 (0.374)
Yen	−0.016 (0.987)	−0.182 (0.855)	−0.236 (0.814)	−0.968 (0.333)
Dollar	0.826 (0.409)	0.563 (0.574)	0.588 (0.556)	0.104 (0.918)

Note: Figures in parentheses denote marginal, two-sided significance levels. All test statistics are distributed as standard normal under the null hypothesis of no shift in volatility. Significantly positive test statistics indicate a reduction in volatility after the break point; significantly negative statistics indicate the converse.

(1983, 1985)).[8] In particular, this suggests that reducing exchange rate volatility inevitably causes interest rate volatility to rise. Such a conclusion might follow from inverting a standard exchange rate equation and noting that the interest rate is the only other major "jump variable" in the system. Such a phenomenon might be termed a "volatility transfer." Insofar as the burden of increased interest rate volatility falls on the general public more squarely than that of exchange rate volatility, which presumably falls mainly on the tradable goods sector, then the welfare argument must hinge on which sector would find it easier to hedge the induced risk. Given the already well-developed forward foreign exchange markets, it is probable that such an argument would come down against membership in the ERM.

However, it is not at all clear that ERM membership is equivalent to "inverting the exchange rate equation." Insofar as membership enhances the credibility of policy, there may be a significant reduction in speculative attacks on the exchange rate and hence a *reduction* in the volatility of short-term interest rates (if the authorities use interest rates as a short-term measure for "leaning into the wind"). Such credibility arguments rest crucially on the assumption that the costs to the authorities of revaluation outweigh the costs of internal adjustment—and, in particular, the costs of disinflation (see Giavazzi and Giovannini (1989)).

To shed some light on these arguments, we carried out the nonparametric tests for monthly changes in Eurocurrency short-term interest rates; the results are reported in Table 8.

Table 8 reveals that the overall effect of the ERM has not been to increase interest rate volatility—if anything, the test statistics show *reduced* interest rate volatility for ERM members, which is strongly significant for lira interest rates. By contrast, dollar interest rates have seen a significant rise in volatility during the operation of the EMS.

During operation of the EMS, there seem to have been further reductions in interest rate volatility (less marked for the lira), although a specific ERM effect cannot be separated from a global effect, since significant reductions in volatility are also indicated for dollar, yen, and sterling interest rates (Tables 8.B and 8.C).

Overall, the results of this section indicate that the stability of nominal and real ERM exchange rates was not bought at the expense of increased interest rate volatility.

[8] Many of the arguments relating to systemic macroeconomic risk can be traced to Poole (1970).

Table 8. *Test Statistics for a Shift in Eurocurency Interest Rate Volatility*

Exchange rate	Normal	Logistic	Double exponential	Cauchy
Period: 1975:1–1979:3 vs. 1979:4–1990:10				
Franc	1.53 (0.126)	1.42 (0.155)	1.37 (0.171)	1.15 (0.250)
Lira	2.04 (0.041)	2.19 (0.028)	2.13 (0.033)	2.41 (0.016)
Mark	1.18 (0.238)	0.97 (0.332)	0.84 (0.401)	0.72 (0.472)
Pound	1.14 (0.254)	1.67 (0.095)	1.43 (0.153)	1.37 (0.171)
Yen	1.13 (0.258)	0.93 (0.352)	1.13 (0.258)	1.18 (0.238)
Dollar	−2.18 (0.029)	−2.13 (0.033)	−2.26 (0.024)	−2.19 (0.028)
Period: 1979:4–1983:3 vs. 1983:4–1990:10				
Franc	4.892 (0.000)	4.231 (0.000)	4.191 (0.000)	4.779 (0.000)
Lira	1.994 (0.046)	1.585 (0.113)	1.592 (0.111)	1.349 (0.177)
Mark	6.051 (0.000)	5.001 (0.000)	4.945 (0.000)	4.826 (0.000)
Pound	3.857 (0.000)	3.266 (0.001)	3.326 (0.001)	3.892 (0.000)
Yen	5.213 (0.000)	4.293 (0.000)	4.232 (0.000)	4.068 (0.000)
Dollar	7.642 (0.000)	6.385 (0.000)	6.204 (0.000)	5.994 (0.000)

Period: 1979:4–1986:5 vs. 1986:6–1990:10				
Franc	4.542 (0.000)	3.924 (0.000)	3.915 (0.000)	4.575 (0.000)
Lira	−0.062 (0.950)	−0.153 (0.879)	−0.220 (0.826)	−0.768 (0.442)
Mark	3.639 (0.000)	3.084 (0.002)	3.150 (0.001)	3.651 (0.000)
Pound	3.009 (0.003)	2.545 (0.011)	2.585 (0.010)	3.054 (0.002)
Yen	3.553 (0.000)	2.907 (0.004)	2.829 (0.005)	2.606 (0.009)
Dollar	5.072 (0.999)	4.301 (0.000)	4.198 (0.000)	4.347 (0.000)
Period: 1987:2–1990:10 vs. 1990:11–1992:9				
Franc	1.314 (0.189)	0.913 (0.361)	0.714 (0.475)	0.398 (0.691)
Lira	−0.331 (0.741)	−0.183 (0.855)	−0.043 (0.966)	0.577 (0.564)
Mark	0.699 (0.485)	0.588 (0.556)	0.674 (0.500)	0.763 (0.445)
Pound	1.223 (0.221)	0.821 (0.411)	0.638 (0.523)	−0.532 (0.595)
Yen	−0.307 (0.759)	−0.235 (0.814)	−0.225 (0.822)	−0.191 (0.848)
Dollar	−0.277 (0.821)	−0.299 (0.765)	−0.215 (0.830)	−0.652 (0.515)

Note: Figures in parentheses denote marginal, two-sided significance levels. All test statistics are distributed as standard normal under the null hypothesis of no shift in volatility. Significantly positive test statistics indicate a reduction in volatility after the break point; significantly negative statistics indicate the converse.

VIII. British Participation in the ERM

Given the comparatively brief duration of the U.K.'s membership in the ERM (October 1990–September 1992) and the growing tensions inside the system during this period, it is interesting to examine whether our nonparametric test procedures can detect these tensions. While this period—less than two years—is arguably too short to yield reliable conclusions, it does seem reasonable that the reduced credibility of the system should be reflected in increased interest rate volatility among member countries. Thus, in Table 8 we report results of the nonparametric procedures applied to interest rate data for the period from the January 1987 realignment to the exit of the pound from the ERM in September 1992, with a hypothesized shift point in volatility after the pound's entry in October 1990. The results reveal no significant shift in interest rate volatility after October 1990 either for the pound or for any other currency examined.

These findings may be explained by returning to the distinction between misalignment and volatility. It seems that some of the tensions within the system reflected cumulative currency misalignments, particularly in the case of the high-inflation currencies; in the case of the United Kingdom, some commentators have argued that the pound joined at too high a rate vis-à-vis the mark (see Wren-Lewis and others (1991)). The pound's rapid devaluation since leaving the ERM gives this hypothesis some empirical support. Thus, if a primary cause of the tensions within the system was currency *misalignment*—a low-frequency concept—we should not necessarily expect to see this reflected in the *volatility* of interest rates—a high-frequency concept.

More generally, we have argued elsewhere (Artis and Taylor (1989)) that the ERM did not render member currencies perfect substitutes in international portfolios, as required of a fully credible exchange rate union (Canzoneri (1982)) and did not exhibit a convincing capacity for correcting cumulative misalignments over time. It seems plausible that market interest in these longer-run issues was heightened by the Danish rejection of the Maastricht Treaty in 1992 and the less than overwhelming ("petit oui") support for the agreement in the French referendum a few months later.

IX. Conclusions

In this study we investigated the volatility of the exchange rates of the ERM countries up to October 1990, before the entry of the United Kingdom into the ERM, and the volatility of interest rates during sub-

samples of the period extending through the pound's participation in the ERM. Because there are doubts about the true distribution of exchange rate and interest rate changes, a nonparametric statistical method was used. The volatility of ERM exchange rates and interest rates was compared with that of a control group of non-ERM currencies before and after the inception of the ERM. Their behavior through time was also examined.

The essential findings are very clear. During the operation of the ERM, the volatility of intra-ERM exchange rates (specifically other ERM rates against the mark) fell while the volatility of non-ERM currencies remained the same or increased. This effect was big enough to replicate itself for ERM countries' overall effective exchange rates also. Similar conclusions, albeit not quite so striking, were obtained for real bilateral and real effective exchange rates. The general impression that the ERM evolved toward greater stability is also confirmed by this data set. The same technique was applied to the evolution of volatility in "offshore," or "Eurocurrency," interest rates: it also appears that volatility has been somewhat reduced for the ERM countries compared to the control group. This is inconsistent with the "volatility transfer" hypothesis, according to which reduced stability in exchange rates would imply added volatility in interest rates. There is no significant shift in interest rate volatility during the pound's participation in the ERM.

Given the recent instability in the ERM, the conclusion that the mechanism has exerted an unequivocally stabilizing influence on its member currencies may seem surprising. We would again refer to the distinction made throughout this paper between the short-run ("high frequency") concept of volatility and the longer-run ("low frequency") concept of misalignment. In earlier work (Artis and Taylor (1989)), we showed that the ERM was not entirely successful either in correcting long-term misalignments of real exchange rates among member currencies or in rendering member currencies perfect substitutes in international portfolios, as would be expected in a fully credible exchange rate union (Canzoneri (1982)): "Both these findings are worrying since it is easy to imagine the stock of credibility which the EMS has earned being dissipated as sophisticated and forward-looking international capital markets begin to focus on the longer-run stability properties. . . ." (Artis and Taylor (1989), p. 305). Given the added instabilities that arise naturally in the transition to monetary union (Masson and Taylor (1992)), the recent abolition of exchange controls, which had previously been extensively used by France and Italy (Artis and Taylor (1988)), and the uncertainty occasioned by the sequence of national referenda on the Maastricht Treaty during 1992, the "events of '92" are neither surprising nor inconsistent with the short-run stabilizing influence of the ERM documented in this paper.

APPENDIX

Density and Asymptotic Score Function for the Nonparametric tests

Distribution	Density function, $f(x)$	Asymptotic score function, $\psi(u)$
Normal	$(2\pi)^{-1/2} \exp(-1/2\, x^2)$	$\{\Phi^{-1}(u)\}^2 - 1$
Logistic	$e^{-x}(1 + e^{-x})^{-2}$	$(2u - 1)\ln\{u/(1 - u)\} - 1$
Double exponential	$1/2 \exp(-\lvert x \rvert)^2$	$-\ln(1 - \lvert 2u - 1 \rvert) - 1$
Cauchy	$\pi^{-1}(1 + x^2)^{-1}$	$2 \tan^2\{\pi(u - 1/2)\}[1 + \tan^2\{\pi(u - 1/2)\}]^{-1} - 1$

REFERENCES

Akhtar, M.A., and R.S. Hilton, "Exchange Rate Uncertainty and International Trade: Some Conceptual Issues and New Estimates for Germany and the United States," Research Paper No. 8403, Federal Reserve Bank of New York (New York: Federal Reserve Bank of New York, May 1984).

Artis, Michael J., and Mark P. Taylor, "Exchange Rates Capital Controls, and the European Monetary System: Assessing the Track Record," in *The European Monetary System*, ed. by Giavazzi, Micossi, and Miller (Cambridge: Cambridge University Press, 1988).

———, "Some Issues Concerning the Long-Run Credibility of the European Monetary System," in *Exchange Rates and Open Economy Macroeconomics,* ed. by R. MacDonald and M.P. Taylor (Oxford: Blackwell, 1989).

Bachelor, Roy A., "European Monetary System, 13th Report from the Treasury and Civil Service Committee and Evidence" (London: House of Commons Select Committee, 1983).

Baldwin, Robert, "On the Growth Effects of 1992," NBER Working Paper No. 3119 (Cambridge, Mass.: National Bureau for Economic Research, 1989).

Boothe, P., and D. Glassman, "Statistical Distribution of Exchange Rates: Some Empirical Evidences," *Journal of International Economics*, Vol. 22 (1987), pp. 297–319.

Canzoneri, Matthew B., "Exchange Intervention Policy in a Multiple Country World," *Journal of International Economics*, Vol. 13 (1982), pp. 267–89.

Cushman, David O., "Has Exchange Risk Depressed International Trade? The Impact of Third Country Exchange Risk," *Journal of International Money and Finance*, Vol. 5 (1986), pp. 361–80.

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

European Commission, "Documents Relating to the European Monetary System," *European Economy*, No. 31 (Brussels: European Commission, 1982).

———, "One Market, One Money," *European Economy*, No. 44 (Brussels: European Commission, 1990).

Giavazzi, Francesco, and Alberto Giovannini, *Limiting Exchange Rate Flexibility: The European Monetary System* (Cambridge, Mass.: MIT Press, 1989).

Giavazzi, Francesco, and Luigi Spaventa, "The 'New' EMS," in *The European Monetary System in the 1990s*, ed. by Paul De Grauwe and L. Papademos (London: Longman, 1990).

Hajek, J., and Z. Sidak, *Theory of Rank Tests* (New York and Prague: Academic Press, 1967).

International Monetary Fund, *Exchange Rate Volatility and World Trade*, Occasional Paper No. 28 (Washington: International Monetary Fund, 1983).

Ludlow, P., *The Making of the European Monetary System* (London: Butterworths, 1982).

Masson, Paul R., and Mark P. Taylor, "Issues in the Operation of Monetary

Unions and Common Currency Areas," in *Policy Issues in the Evolving International Monetary System*, by Morris Goldstein and others, Occasional Paper No. 96 (Washington: International Monetary Fund, 1992).

Padoa-Schioppa, Tomaso, "European Monetary System, 5th Report and Evidence" (London: House of Lords Select Committee, 1983).

Poole, William, "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Macro Model," *Quarterly Journal of Economics*, Vol. 85 (1970).

Rogalski, R.J., and J.D. Vinso, "Price Level Variations as Predictors of Flexible Exchange Rates," *Journal of International Business Studies*, Vol. 2 (1977), pp. 71–81.

Rogoff, Kenneth, "Can Exchange Rate Predictability Be Achieved Without Monetary Convergence," *European Economic Review*, Vol. 28 (1985), pp. 93–115.

Stiglitz, Joseph, and A. Weiss, "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, Vol. 71 (1981).

Ungerer, H., O. Evans, and P. Nyberg, *The European Monetary System: The Experience 1979–82*, Occasional Paper No. 19 (Washington: International Monetary Fund, May 1983).

Ungerer, Hurst, and others, *The European Monetary System: Recent Developments*, Occasional Paper No. 48 (Washington: International Monetary Fund, 1986).

———, *The European Monetary System: Developments and Perspectives*, Occasional Paper No. 73 (Washington: International Monetary Fund, 1990).

Van Ypersele, J., *The European Monetary System* (Cambridge: Woodhead, 1985).

Westerfield, J.M., "Empirical Properties of Foreign Exchange Rates Under Fixed and Floating Rate Regimes," *Journal of International Economics*, Vol. 7 (1977), pp. 181–200.

Williamson, John, "The Exchange Rate System" (revised ed.), *Policy Analyses in International Economics*, Vol. 5 (Washington: Institute for International Economics, 1985).

Wren-Lewis, Simon, and others, "Evaluating the United Kingdom's Choice of Entry Rate into the ERM," *The Manchester School*, Vol. 49 (Supplement, 1991), pp. 1–22.

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

Shorter Paper and Comments

The Use of Foreign Exchange Swaps by Central Banks

CATHARINA J. HOOYMAN*

The paper discusses the use of foreign exchange swaps by central banks. Such use has aimed at affecting domestic liquidity, managing foreign exchange reserves, and stimulating domestic financial markets. The discussion is illustrated using selected countries. The paper cautions about the use of foreign exchange swaps to defend a particular exchange rate at a time when foreign exchange reserves are under pressure. It notes, finally, that use of foreign exchange swaps by central banks has been losing importance. [JEL E52, E58, F33, G15]

IN A FOREIGN EXCHANGE SWAP, two parties exchange specific amounts of two different currencies and repay the amount of the exchange at a future date, according to a predetermined rule reflecting both interest payments and amortization of the principal. Different types of foreign exchange swaps became fashionable in the 1980s and received much attention in the literature, most of it concerning the fast-growing interbank swap market. Less publicized, however, has been the fact that for several decades central banks have also used foreign exchange swaps to affect domestic liquidity, manage their foreign exchange reserves, and stimulate domestic financial markets. This paper describes the various ways in which central banks use foreign exchange swaps to achieve these goals.

* Catharina J. Hooyman is affiliated with the University of Amsterdam and Tinbergen Institute, the Netherlands. This paper was written during a summer internship in 1992 in the Monetary and Exchange Affairs Department of the International Monetary Fund. The author wishes to thank Tomás J. T. Baliño, Tom Nordman, Anton Op de Beke, Steve Gilmore, Sohail Qureshi, numerous desk officers, and, especially, Daniel Hardy for valuable comments and support.

The survey does not claim to be exhaustive, but rather illustrates all broad categories of central bank swaps (more elaborate evidence is given in Hooyman 1993). Information about central banks' use of foreign exchange swaps is scarce since it usually does not appear on balance sheets, and central banks often keep their swap operations confidential in order to avoid adverse signaling effects. Furthermore, if data are available, their quality is often uncertain.

I. Foreign Exchange Swaps: Some Background Information

This section provides some general information about foreign exchange swaps and foreign exchange markets.

Definition and Pricing

In the 1980s, a common foreign exchange swap involved the exchange of interest payments over time, a transaction usually called a "currency swap." For a much longer time, however, a simpler swap transaction has been used, in which only the principal amounts are exchanged on the initial and maturity dates at predetermined exchange rates. This latter kind of swap is the most relevant for this paper, since central banks use it more often than currency swaps.

According to the covered interest parity condition, the forward premium, or discount, reflects the corresponding differential between interest rates on the international markets, as risk-free arbitrage should ensure that no unexploited profit opportunities exist. This forward premium, sometimes called a swap premium or swap rate, is considered the price of the instrument. In countries with well-developed forward exchange markets, this price is market determined, and quotes can be obtained from commercial banks.

Currency swaps (like interest rate swaps and cross-currency interest rate swaps) are used by a wide variety of participants—banks, corporations, insurance companies, international agencies (the World Bank was a driving force in the development of the market), and sovereign states. There are four broad reasons to use swaps: (i) to exploit differences in credit rating and differential access to markets, thereby obtaining low-cost financing or high-yield assets; (ii) to hedge interest rate or currency exposure; (iii) to manage short-term assets and liabilities; and (iv) to speculate. Central banks have been known to use currency swaps for hedging and asset-liability management, but very rarely; hence, those operations will not be elaborated upon.

Central Bank Swaps in the Balance Sheet

When a central bank carries out a foreign exchange swap, it has significant economic effects. Suppose the Federal Reserve system buys foreign exchange with domestic currency and simultaneously agrees to sell the same amount of foreign currency at a certain date in the future at the forward exchange rate. As the Fed's foreign assets increase, some items on the liabilities side must increase, depending on the counterparty. If the latter is another country's central bank, this institution's account at the Fed is credited, meaning that the Fed issues dollars and the other central bank issues its own currency. As long as both central banks do not spend the foreign currency, however, there is no effect on either currency in circulation or banks' reserves, so both money supplies remain constant. If the counterparty is the banking system, banks' reserves are credited with the domestic currency equivalent of the foreign exchange purchase, and banks' foreign assets decline. Thus, reserve money increases, which normally causes an expansion of the money supply.

Another possible approach is to treat such swap operations as collateralized loans. In analytic terms, expansionary foreign exchange swaps with deposit money banks are similar to direct loans by the central banks. Depending on a complex of legal, historical, and institutional considerations, negotiable financial instruments may continue to be registered as assets in the balance sheet of the deposit money banks; their reserves item would increase and be balanced on the liability side by "central bank discounts received" (IMF (1984), p. 53). In this case, the central bank's balance sheet shows an increase in domestic assets (claims on deposit money banks), not foreign assets. When the swap is unwound, both domestic assets and banks' reserves return to their old levels, with an interest payment going into the central bank's profit and loss account. This is the same amount as the interest earned on the foreign asset, adjusted for the difference between the spot and forward exchange rates.

Another consequence of the expansionary swap is the creation of a forward foreign liability for the central bank, matched by a forward domestic asset. This matched pair of contingent accounts can be recorded on or off the balance sheet, depending on the local practice. The four-entry system inflates and complicates the balance sheet, but it has the benefit of showing clearly which part of the foreign reserves is temporary, and therefore which exchange risk the central bank would be running if it lost its cover.

The swap in this example may not be intended as a money market tool; nonetheless, it still has the (temporary) expansionary effect on reserve money discussed above. Thus, if the swap was undertaken to gain foreign

exchange or to provide banks with a forward cover, the consequent monetary expansion would have to be sterilized. In short, the goals for using foreign exchange are often conflicting.

A last topic to be discussed is the risk for the central bank in foreign exchange swaps. When foreign exchange swaps are seen as collateralized loans, normally there will be little risk: the central bank need not worry about default risk, since it has the collateral. Also, it is not exposed to exchange rate risk, as long as it has the foreign assets to cover the forward foreign liability (there is, of course, the risk of opportunity cost). There is exchange risk as soon as either the asset or the liability disappears—that is, if the counterparty defaults before the swap matures or if the central bank runs out of foreign exchange reserves. The latter situation has occurred in many countries and will be discussed in a later section. Finally, there is a settlement risk involved with swaps, as is always the case in any foreign exchange operation; the so-called Herstatt risk. However, the risk is very small.

Foreign Exchange Markets

In foreign exchange markets, spot transactions are settled on the second business day following trade. This is because the funds are ultimately transferred by having the central bank of the country issuing the currency transfer liabilities from the sending party's account to that of the receiving party, and the two central banks may be in different time zones. The time difference also causes the (tiny) Herstatt risk, which occurs when one party is not able to receive the other party's currency delivery lag. The delivery lag is made up by the time difference plus the difference between each country's local time for final settlement. A central bank doing foreign exchange swaps is thus exposed to default risk by its counterparty, since after fulfilling its own obligation it has to wait several hours for payment.

In addition to dealing in the spot foreign exchange market, central banks engaged in foreign exchange swaps also enter the forward market. To be suitable for the use of swaps as a money market tool, this market should be deep, and quotes of the forward exchange rate should be readily available. The first criterion ensures that large transactions are not disruptive. The second requirement means that central banks should not be "making the market," the price should be truly market determined. If the market were thin, and the rate were determined by the

central bank, swap transactions could have disruptive effects on exchange rate expectations.[1]

II. Open Market Policy

In the following section, theory and facts are presented about the use of foreign exchange swaps as a money market tool in industrial and developing countries; such use differs noticeably between the two groups of countries.

Industrial Countries

In the past two decades, an increasing number of central banks in industrialized countries have included foreign exchange swaps among the instruments with which they fine-tune domestic liquidity, even though actual use of such swaps has remained limited in most countries.

As a result of the enormous increase in the volume, integration, and liberalization of international financial markets, and the emergence of new markets and instruments, a trend has developed for central banks to move from direct to indirect monetary policy. Instead of fixing of interest rates by administrative means or imposing controls, they increasingly use market-based instruments, such as open market operations. The bulk of these operations is in the form of central bank credit, but numerous other instruments are used.

Central banks employ foreign exchange swaps because (i) they prefer to have a wide range of intervention techniques at their discretion (possibly because they may wish to vary the predictability of their policy actions); (ii) in many countries the domestic short-term secondary market is not deep enough to permit market intervention, whereas the market in foreign exchange is generally active, making it possible to trade large volumes in any one deal; (iii) unlike outright foreign exchange operations, swaps have no direct effect on the spot (or forward) exchange rate;[2] and (iv) swaps are flexible in that the technical procedures are informal and the swaps themselves are inconspicuous and easily reversible.

[1] According to the *un*covered interest parity, the forward rate is equal to the expected future spot rate. Thus, if the central bank, which is supposed to defend the currency, would set the domestic currency at a forward discount, this would have a strong announcement effect on the spot foreign exchange market.

[2] Swaps can be seen as analytically similar to temporary operations in domestic securities: the exchange rate will normally be influenced only to the extent of the swaps' impact on interest rates. Of course, this may amount to the same thing, as interest rate policy and exchange rate policy are often inextricably connected.

Possible drawbacks are that (i) foreign exchange swaps might influence the exchange rate because of a strong announcement effect; (ii) foreign exchange transactions take two days to become effective, which makes foreign exchange swaps less appropriate when swift action is required; (iii) in foreign exchange transactions there is no simultaneous exchange of one currency for the other, which gives rise to settlement risk;[3] (iv) often, there are only a limited number of large banks to act as counterparties—banks have to get the necessary dollars in the international market, and if the sum is large relative to their capital, a risk premium will be charged; (v) if active short-term securities markets exist, central banks often prefer to conduct their operations there; and (vi) if monetary policy targets interest rates instead of a monetary aggregate, allocation of central bank credit might be a more suitable instrument, since the immediate effect of swaps is on the high-powered money supply.

Since swap transactions are temporary by nature, they are suitable for short-term technical adjustment: either to influence the general liquidity of the market, so as to neutralize the effect of fortuitous or seasonal factors, or to bring about temporary market imbalances that can push interest rates in a desired direction. However, swaps can easily be rolled over for a longer-term effect, and the maturity of swap operations can be extended, the present range being from 24 hours to 24 months. Central bank foreign exchange swap operations may be conducted anonymously in the market at the usual maturities, but more flexible contracts may be concluded bilaterally with banks. They generally involve U.S. dollars.

Switzerland is the only country where foreign exchange swaps are the main instrument for the management of bank reserves, mainly because of the lack of short-term government securities (the Swiss government does not usually run a budget deficit). The total amount of swaps outstanding averages around 40 percent of the monetary base.

Apart from Switzerland, the two countries that rely most on central bank foreign exchange swaps are the Netherlands and Germany. In both, short-term securities markets are extremely thin, but the central banks use foreign exchange swaps infrequently because they rely on other open-market instruments. In Germany, foreign exchange swaps have been used by the Bundesbank since 1958. For the first decade, it used contractive swaps both to influence the domestic money market and to stimulate short-term foreign investment by offering attractive swap rates. From the late 1960s, swaps were motivated mainly by attempts to calm

[3] Here one must abstract from the possibility that the central bank ends up with a net short position in foreign currency, since in such situations central banks would not be using swaps as a money market tool.

the international monetary situation and strengthen confidence in the dollar parity. Foreign exchange swap transactions have served the purpose of "fine-tuning" the money market only since 1979. Besides swaps, the Bundesbank carries out so-called foreign exchange transactions under repurchase agreements. These are essentially the same as swaps, but the ownership of the foreign asset does not change. Quantitatively, foreign exchange swaps and repurchase agreements have sometimes been of considerable importance for fine-tuning but do not make up a substantial part of the monetary base (Deutsche Bundesbank (1989), pp. 77–79).

The importance of the foreign exchange swap in the Netherlands was never great, and operations have ceased during the past few years (but are not officially abandoned). Other countries that have used currency swaps (albeit rarely) are the United Kingdom, Norway, Finland, Belgium, Ireland, and Austria (Bingham (1985), Kneeshaw and Van den Bergh (1989), and annual reports of various central banks).

Developing Countries

Now consider the usefulness of foreign exchange swaps as a money market tool for less developed countries. The preceding discussion suggests that swaps could be recommended for countries that (i) wish to conduct monetary policy in a market-oriented way; (ii) target a monetary aggregate; (iii) lack a well-developed market for short-term securities (especially countries where there is no securitized government debt); and (iv) have deep spot and forward foreign exchange markets. The latter criterion is probably the most binding. Developing countries that do use foreign exchange swaps include Kuwait, Saudi Arabia, and Malaysia, all of which meet the forward market criterion,[4] as well as Oman, United Arab Emirates, Bahrain, and Turkey, which do not.

The central bank of Saudi Arabia (SAMA) provides liquidity to banks through foreign exchange swaps using spot sales of U.S. dollars to itself with a repurchase agreement based on the market-determined forward exchange rate. The swap facility is not a very important instrument in terms of size, but it has been helpful occasionally—for instance, when swaps were used during the recent regional crisis to provide emergency liquidity to the market.

The Central Bank of Oman (CBO) started its swap facility in March 1980. All foreign exchange swaps (always expansionary) are initiated by

[4] South Africa also qualifies, and it does use swaps, but not as a money market tool.

the commercial banks, each with an individual ceiling. Initially, U.S. dollars were swapped at par (the Omani rial is pegged to the U.S. dollar, and the exchange rate has been very stable). This is, of course, related to the fact that there is no developed forward market in Omani rials, and the financial system is fairly regulated. The drawbacks of this situation became clear in 1986, when the domestic interest rate exceeded international rates. This gave the banks the possibility of risk-free windfall profits, leading to a peak level of RO 27 million outstanding in August 1986. In July 1986, the facility was modified to cure this flaw, and the oustanding swap amount subsequently declined sharply. In 1989, there was a squeeze on banks' Omani rial liquidity, so the use of the swap facility peaked again, at RO 35 million in April 1989 (averaging 7.19 percent of reserve money for 1989). Apart from these peaks, the share of reserves acquired through foreign exchange swaps in the monetary base has been less than 1 percent. Another disadvantage of the lack of market-orientedness in Oman's swap system is that the CBO cannot do reverse swaps to withdraw liquidity because they would cause destabilizing expectations of exchange rate movements. The system is not very flexible as a result.

In Turkey, monetary control has been exercised largely through the reserve requirement ratio and occasionally through limits on central bank credit. Beginning in 1987, open market operations have been expanded but are still limited by the size of the central bank's portfolio and the thinness of the securities market. The Turkish lira–U.S. dollar swap facility has been in operation for more than a decade, and the swaps are carried out as an exchange of mutual deposits, which gives rise to the four-entry system mentioned earlier. The foreign exchange deposit causes the central bank's foreign assets to rise and creates a (forward) foreign liability. The Turkish lira deposit increases banks' reserves and the central bank's domestic assets. During the term of the swap, the central bank's foreign asset is valued at the historical exchange rate, but its foreign exchange liability is revalued with the current exchange rate. As the Turkish lira has been depreciating constantly since the 1970s, the net of these four items is always negative. This net figure is called "foreign exchange swaps" in the Turkish banking system accounts. The interest on the Turkish lira deposit, which should in theory compensate for the capital loss, goes into the profit and loss account.

This treatment implies that no forward exchange rate is agreed upon; the central bank relies on the uncovered interest parity to hold. The amount of swaps outstanding peaked in Turkey in 1988, probably because the anticipated liberalization of interest rates in October 1988 was expected to widen interest differentials substantially, thereby reducing the

profitability of swap operations to commercial banks. The outstanding amount has decreased notably since then and has been virtually stable since mid-1990, when the central bank stopped actively using swaps.

III. Management and Acquisition of Foreign Exchange Reserves

Swaps are also used as a tool for the management and acquisition of foreign exchange reserves, with both the banking system and other central banks acting as counterparty. Such use sometimes coincides with the use of swaps to stimulate the domestic financial system.

Asset-Liability Management

Central banks nowadays are under more pressure to actively manage their assets than in the past; besides defending the exchange rate, they have to search for better returns as well. But, since intervention in the foreign exchange market requires instant access to reserves, liquidity is crucial. In determining the currency composition of their reserves, some central banks take account of the currency composition of their country's import basket, with higher weights for currencies with liquid bond markets and for currencies that a country uses for intervention. This provides a rationale to use currency swaps to temporarily adjust the currency distribution when it has been distorted as a result of intervention. The central bank of Norway uses currency swaps and forwards to maintain the liquidity of its assets while leaving the currency distribution unchanged.

Currency swaps also provide cross-currency hedging (and interest rate hedging if cross-currency interest rate swaps are used). This is done when the foreign assets are denominated in different currencies than the foreign liabilities. In less developed countries, foreign asset-liability management is not so much about maximizing returns while maintaining liquidity as it is a matter of minimizing the inevitable financial risk. One way to do that is to improve the matching and currency composition of foreign asset and liability structures. A developing country that has engaged in such activities is Trinidad and Tobago. The country has considerable external debt, denominated in Japanese yen, which was partially taken over by the central bank. The widely fluctuating foreign reserves (mainly earnings from oil exports) are largely in U.S. dollars, since the TT dollar is pegged to the U.S. dollar. The central bank has used currency swaps to hedge against changes in the U.S. dollar–yen exchange rate, as well as swaps from floating-rate into fixed-rate liabilities.

Another rationale for using currency swaps in asset-liability manage-

ment is to influence published official foreign exchange asset positions—that is, hide the real fluctuations in them. The Banque de France has reportedly done this, and many others may also be doing so quietly, since it is unlikely to be public knowledge.

Acquisition of Foreign Exchange Through Swaps Among Central Banks

In addition to commercial banks, other central banks can also act as a counterparty in foreign exchange swap transactions. If central banks deem it necessary to intervene heavily in the foreign exchange market, they may acquire the needed reserves by drawing on a swap line with another central bank. It is for this purpose that the Federal Reserve's swap network exists: a system of reciprocal short-term credit arrangements between the Fed and 14 other central banks (those of most western European countries, plus Canada, Mexico, and Japan) and the Bank for International Settlements. The network enables the Fed to acquire currencies needed to counter disorderly market conditions through market operations, and it enables the swap partners to acquire the dollars they need to conduct their own operations. The Fed's swap lines amount to $30.1 billion.

Swap drawings to finance official exchange intervention do not affect the money supply under these operating procedures. For example, if the Fed draws on the swap line with the Bundesbank, the Bundesbank, which does not immediately need the dollars, invests them in U.S. securities so that these dollars find their way back into U.S. bank reserves (Kubarych (1978), p. 19). These Federal Reserve swaps were prominent in the late 1970s but their importance has diminished since then (although the swap lines have increased). Swaps between central banks do not arise for foreign exchange intervention purposes exclusively; they can also be connected with trade. For example, Guatemala and Costa Rica swapped their currencies in the early 1980s to deal with regional trade deficits.

Acquisition of Foreign Exchange Reserves in Situations of Scarcity

Many developing countries suffer from substantial external indebtedness and wide fluctuations in their export earnings, as well as steady demand for imported basic goods. These countries must hold foreign exchange reserves in order to prevent excessive short-term fluctuations in the exchange rate, and liquidity is of prime importance. When faced

with an acute shortage of liquid foreign exchange reserves, central banks in a number of countries have resorted to swaps to increase their gross foreign exchange holdings. Examples are Chile, Argentina, Uruguay, Ecuador, the Philippines, Indonesia, and the Republic of Korea.

Argentina experienced a balance of payments crisis starting in 1982, whereupon it announced a number of far-reaching measures to deal with the situation. They included foreign exchange swaps with domestic banks and residents in possession of foreign exchange deposits (such as importers). Also, domestic banks and corporations with foreign debt were encouraged to renegotiate the debts; the central bank of Argentina (BCRA) rewarded the renegotiation with what amounted to an exchange rate guarantee for the remaining life of the debt. Both measures exposed the BCRA to exchange rate risk; without cover, it was dependent on the uncovered interest parity to hold. The BCRA could not set its swap rate in this way, because it had to set a preferential rate to attract the swaps and because the announcement effect of a "realistic" swap rate would cause undesirable capital movements. Consequently, the BCRA suffered huge losses, with negative external operating results averaging about 0.5 percent of GDP since 1984. Other countries where central bank foreign exchange swaps have led to large losses from the depreciation of the domestic currency are the Philippines and South Africa.

Korea's financial sector has been quite regulated. Before 1986, Korea had a persistent balance of payments deficit, so foreign exchange reserves were scarce. The central bank of Korea engaged in swaps with foreign commercial banks, which were not allowed to establish a network of local branches. This caused the domestic currency funding to be very limited, and to acquire working capital the banks borrowed from their head offices and swapped the proceeds into Korean won. The swap rate was chosen to provide an incentive for capital inflows, and individual bank swap limits were regularly increased. The larger part of these swaps was sterilized. As of 1986, the balance of payments went into surplus, which eliminated the need for swaps. The local branches of foreign banks were granted the option of having access to the discount window of the Bank of Korea and were allowed to issue certificates of deposit. The swap limits have not since been increased.

To sum up, the use of foreign exchange swaps in a situation of scarce foreign reserves has the advantage of providing a short-term capital inflow. However, they also have an expansionary monetary effect that has to be sterilized; if there is no (active) forward market, the central bank has to set a swap rate that may cause adverse signaling effects. Moreover, if the country is in severe crisis, using swaps to temporarily boost gross foreign exchange reserves will only delay the necessary adjustment.

Above all, if the central bank cannot keep its position covered (as in times of speculation against the currency), it is liable to suffer exchange losses. In the case of Argentina, these negative effects clearly outweighed the benefit. There are examples of other countries (such as Korea), however, that did not experience particularly adverse effects.[5]

The use of foreign exchange swaps to gain reserves is difficult to distinguish from swaps used to stimulate financial markets or to subsidize certain activities when the central bank acts as a market-maker in forward foreign exchange markets, provides forward cover and exchange rate guarantees, or subsidizes domestic financial institutions. Such activities have been common in developing countries. Regardless of the purpose, these operations are very similar to those discussed above; they amount to the central bank taking an open position, thereby assuming an exchange rate risk that often results in large losses. These kinds of activities are therefore not recommended, since central banks can do more to stimulate financial markets by conducting a credible exchange rate policy. Subsidies should be on the government budget, and the financial sector does not gain in the long run if its central bank incurs substantial losses.

Central bank losses negatively affect the economy in a number of ways. Large losses erode the central bank's capital, which may jeopardize its independence. Moreover, the losses represent an injection of liquidity, which might make it necessary for the central bank to issue interest-bearing liabilities. This policy embodies a risk that future losses may grow exponentially. Persistent losses may also create pressure to implement an expansionary policy as a way to reduce future losses (Vaez-Zadeh (1991)).

IV. Summary and Conclusions

Central banks of industrial countries use foreign exchange swaps to "fine-tune" the money market, to acquire foreign reserves for intervention purposes, and to manage their assets and liabilities. In some

[5] Another way to obtain liquid foreign resources is through gold swaps; loans in foreign currency backed by deposits of gold. The classic operation consists of selling gold at the current price and repurchasing the same gold at a future date. Such an operation makes it possible to obtain temporary financing while paying an interest rate below the current market rate (the risk premium is reduced because the gold serves as collateral). Uruguay and Ecuador have been active in gold swaps, swapping gold for Swiss francs. As long as the central banks keep their positions covered (hold on to the foreign exchange to make sure they have a matching asset by the time the swap matures), there is no exchange risk involved. Unfortunately, when reserves are scarce this is often neglected, so that the bank will suffer a loss if the domestic currency depreciates during the swap interval.

developing countries, central banks also use foreign exchange swaps to control bank liquidity. In other cases, they use swaps to defend scarce official reserves and to stimulate or subsidize the domestic financial system.

As a money market tool, foreign exchange swaps are flexible and can be used in the absence of a well-developed short-term financial market. The operations should be conducted in accordance with market mechanisms; therefore, a deep forward foreign exchange market is necessary for the instrument to function smoothly.

As a means of defending foreign reserves in times of balance of payments problems, the appropriateness of foreign exchange swaps is considerably more questionable. In countries where these problems are acute, swaps can have adverse effects, because they leave the central bank with an open currency position that almost always creates losses. For the same reason, using central bank swaps to stimulate the development of financial markets is not recommended. Central bank losses could seriously damage monetary policy, and thus monetary stability. In countries with moderate and temporary balance of payments difficulties, swap schemes can be useful.

On a global scale, it appears that central bank foreign exchange swaps are losing favor; most swap facilities have been restricted or dormant in recent years. This is probably connected with the ongoing development of financial markets around the world. As short-term securities markets develop, swaps become less attractive as a money market tool, and the poor experiences of countries that have used swaps during difficult times have further discouraged their use.

REFERENCES

Bingham, T. R. G., *Banking and Monetary Policy* (Paris: OECD, 1985).

Deutsche Bundesbank, *The Deutsche Bundesbank: Its Monetary Policy Instruments and Functions,* Special Series No. 7, 3rd edition (Bonn: Deutsche Bundesbank, 1989).

Hooyman, C. J., *The Use of Foreign Exchange Swaps by Central Banks: A Survey,* IMF Working Paper 93/64 (Washington: International Monetary Fund, 1993).

International Monetary Fund, *A Guide to Money and Banking Statistics in International Financial Statistics* (Washington: International Monetary Fund, 1984).

Kneeshaw, J. T., and P. Van den Bergh, *Changes in Central Bank Money Market Operating Procedures in the 1980s* (Basle: Bank for International Settlements, 1989).

Kubarych, R. M., "Monetary Effects of Federal Reserve Swaps," *Federal Reserve Bank of New York Quarterly Review* (Winter 1977–78), pp. 19–21.

Vaez-Zadeh, Reza, "Implications and Remedies of Central Bank Losses," in *The Evolving Role of Central Banks*, ed. by P. Downes and R. Vaez-Zadeh (Washington: International Monetary Fund, 1991).

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

Unification of Foreign Exchange Markets

A Comment on Agénor and Flood

CHING-CHONG LAI and WEN-YA CHANG*

Recently in an interesting paper in *IMF Staff Papers*, Pierre-Richard Agénor and Robert Flood (1992) set up a theoretical model of dual exchange rates and analyzed how the economy would work given the shock of an announcement to unify dual exchange markets. Provided that the unified exchange rate under the flexible regime is greater than the official exchange rate under the dual regime, their major findings are (i) at the announcement, the domestic price of foreign currency in the parallel exchange market depreciates immediately, (ii) the parallel exchange rate keeps rising and net foreign assets keep falling during the pre-unification period, and (iii) the unified exchange rate of the post-unification regime is independent of the timing of unified implementation.

This paper is written for two purposes. First, we intend to show that the unified exchange rate of the post-reform regime is inversely related to the date of reform. Second, we attempt to extend the graphical presentation of Agénor and Flood (1992) and to trace the possible dynamic behavior of relevant variables during the unified process.[1]

*Ching-chong Lai is a Research Fellow at Sun Yat-Sen Institute for Social Sciences and Philosophy, Academia Sinica, in Taipei. Wen-ya Chang is an Associate Professor in the Economics Department and Graduate Institute of Economics of Fu-Jen Catholic University.

[1] Agénor, Bhandari, and Flood (1992) provide an excellent survey of the exchange regime reform.

I. The Model

The Agénor and Flood (1992) model can be described by the following log-linear equations:

$$m - p = -\alpha i; \qquad \alpha > 0 \tag{1}$$

$$i = i^* + \dot{s} - \gamma(s - e); \qquad \gamma > 0 \tag{2}$$

$$m = \theta R + (1 - \theta)D; \qquad 0 < \theta < 1 \tag{3}$$

$$p = \nu s + (1 - \nu)e; \qquad 0 < \nu < 1 \tag{4}$$

$$\dot{R} = -\Phi(s - e); \qquad \Phi > 0 \tag{5}$$

$$\dot{D} = 0. \tag{6}$$

With the exception of the domestic interest rate i and the foreign interest rate i^*, all variables are expressed in logarithms. The variables are defined as follows: m = nominal money supply; p = domestic price level; s = parallel exchange rate; e = official exchange rate; R = foreign reserves; and D = domestic credit.

Equation (1) is the standard equilibrium condition for the money market. A logarithmic approximation of the interest rate parity under dual exchange rates is given in equation (2).[2] Equations (3) and (4) are the definition of money supply and of domestic price level, respectively. Equation (5) describes how foreign reserves change over time, while the divergence between the official and the parallel exchange rates captures the arbitrage activity between two foreign exchange markets.[3] Finally, equation (6) depicts that domestic credit remains constant at all times.

The experiment of the Agénor and Flood (1992) study is that the foreign authorities announce the regime will switch from a dual to a unified flexible regime at a specific date T in the future. Because the regime switch involves dual exchange rates and flexible exchange rates, in what follows we examine the dynamic features of both regimes. For expository convenience, throughout this paper 0^- and 0^+ denote the instants before and after announced unification, T^- and T^+ denote the instants before and after implemented unification.

Dual Exchange Regime

Under dual exchange rates, e is given at the level $\bar{e}$. Letting $D = \bar{D}$ and $i^* = 0$, from equations (1)–(6), we have

[2] For a detailed derivation, see Flood and Marion (1983).
[3] Gros (1988) provides a good microfoundation for linking arbitrage behavior with exchange rate differentials.

$$\begin{bmatrix} \dot{s} \\ \dot{R} \end{bmatrix} = \begin{bmatrix} (\alpha\gamma + \nu)/\alpha & -\theta/\alpha \\ -\Phi & 0 \end{bmatrix} \begin{bmatrix} s \\ R \end{bmatrix} + \begin{bmatrix} -(1-\theta)\overline{D}/\alpha + (1 - \nu - \alpha\gamma)\bar{e}/\alpha \\ \Phi\bar{e} \end{bmatrix}. \quad (7)$$

Let ρ_1 and ρ_2 be the two roots of the dynamic system, we then have

$$\rho_1 \rho_2 = -\Phi\theta/\alpha. \quad (8)$$

Following Agénor and Flood (1992), we assume that $\rho_1 < 0 < \rho_2$. The general solutions for s and R during the pre-unification period are

$$s = s^* + A\exp(\rho_1 t) + B\exp(\rho_2 t); \qquad 0^+ \leq t \leq T^- \quad (9)$$

$$R = R^* + \frac{[(\alpha\gamma + \nu) - \alpha\rho_1]}{\theta} A\exp(\rho_1 t) + \frac{[(\alpha\gamma + \nu) - \alpha\rho_2]}{\theta} B\exp(\rho_2 t); \qquad 0^+ \leq t \leq T^- \quad (10)$$

where s^* and R^* are the steady-state values of s and R, and A and B are undetermined coefficients.

The dynamic behavior of the system can be illustrated by means of a phase diagram like Figure 1, which is similar to Figure 1 in Agénor and Flood (1992). It is clear from equation (7) that the slopes of loci $\dot{s} = 0$ and $\dot{R} = 0$ are

$$\partial s/\partial R|_{\dot{s}=0} = \theta/(\alpha\gamma + \nu) > 0, \quad (11)$$

$$\partial s/\partial R|_{\dot{R}=0} = 0. \quad (12)$$

Equipped with the information of arrow directions and with equation (8), we can see that the model has a saddlepoint solution. In the phase space plane, the stable branch SS, which is the unique trajectory that leads to the stationary equilibrium, is associated with $B = 0$ in equations (9) and (10), while the unstable branch UU is associated with $A = 0$. All other unstable trajectories indicated in the figure correspond to the values $A \neq 0$ and $B \neq 0$. It is clear from equations (9) and (10) that the slope of SS is

$$\partial s/\partial R|_{SS} = \theta/[(\alpha\gamma + \nu) - \alpha\rho_1] > 0. \quad (13)$$

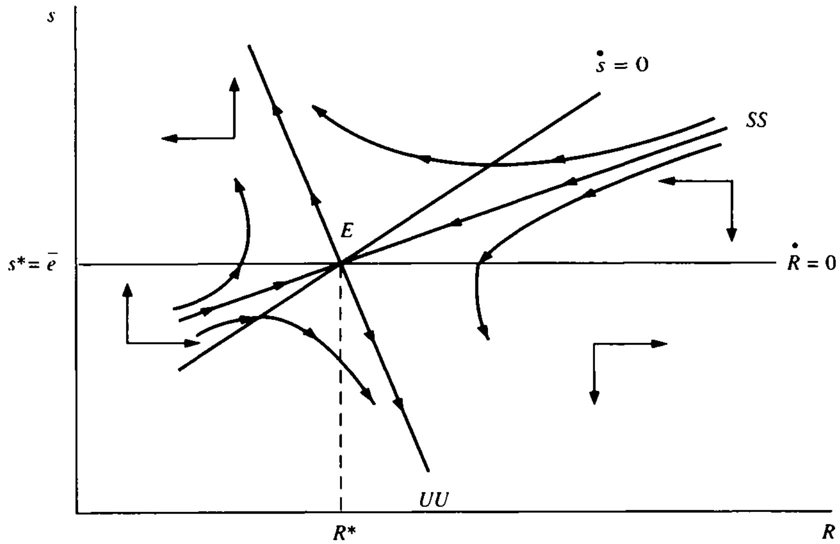
Flexible Exchange Regime

As in Agénor and Flood (1992), we proceed to consider the post-unification (flexible) regime. Under flexible exchange rates, given $\dot{R} = 0$ and $\epsilon = e = s$, from equations (1)–(6), we derive

$$\alpha\dot{\epsilon} - \epsilon = -m, \quad (14)$$

where $m = \theta R_{T^+} + (1-\theta)\overline{D}$.

Figure 1. *Dynamic Behavior of Exchange Rate System Before Unification*



The general solution for equation (14) is

$$\epsilon = Ce^{t/\alpha} + \theta R_{T^+} + (1 - \theta)\overline{D}; \qquad t \geq T^+. \tag{15}$$

As the unique root of flexible exchange rates is positive, the convergent condition thus requires that the speculative bubble should be ruled out. As a consequence, substituting $C = 0$ into equation (15) gives

$$\epsilon = \theta R_{T^+} + (1 - \theta)\overline{D}; \qquad t \geq T^+. \tag{16}$$

Two points should be noted from equation (16). First, from T^+ onward, the economy is operating under flexible exchange rates, and hence the stock of foreign reserves is frozen at the level R_{T^+}. Second, the unified exchange rate ϵ is solely determined by the stock of foreign reserves, which in turn is crucially related to the timing of regime switch T.[4] It should be noted that equations (9)–(16) are a restatement of the analysis in Agénor and Flood (1992).

[4] Since the flexible regime is adopted after the dual regime, the stock of foreign reserves under the dual regime, denoted by R_{T^+}, will serve as an initial condition for the dynamics under the flexible regime beyond time T^+.

The AA schedule presented in the diagrams of the next section traces the combinations of ϵ and R that satisfy equation (16). It is clear from equation (16) that the slope of AA is

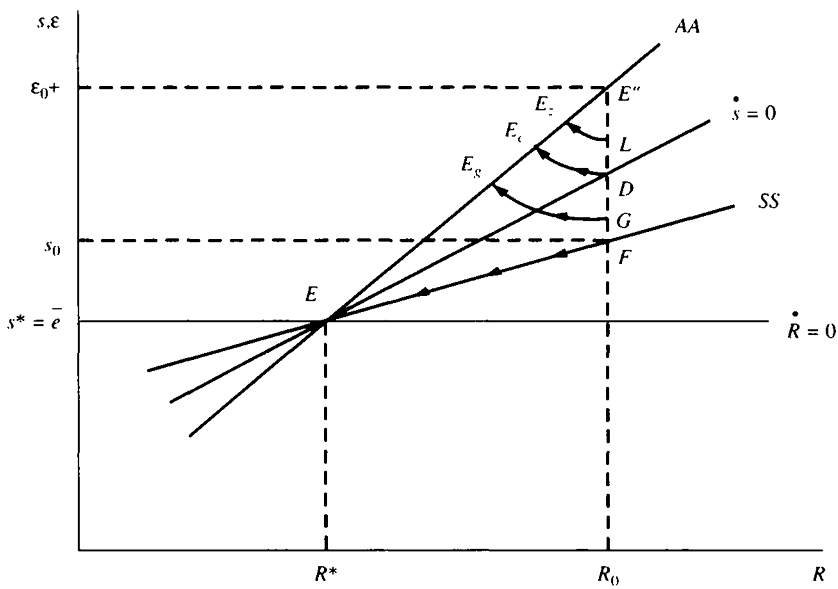
$$\partial\epsilon/\partial R|_{AA} = \theta > 0. \tag{17}$$

II. The Unification Process

We are now in a position to deal with the evolution of the economy during the unification process.[5] As indicated in equations (13) and (17), the AA schedule may be either flatter or steeper than the SS schedule. We thus have to consider two possible situations.

Figure 2 illustrates the Agénor and Flood (1992) situation where the slope of AA is greater than that of SS. Following Agénor and Flood (1992), assume that initially the system is at the point F on the saddlepath SS; the initial parallel exchange rate and foreign reserves are s_0 and R_0, respectively.

Figure 2. *Agénor and Flood's Unification Process*



[5] The detailed mathematical derivations behind the graphical results are available upon request.

Corresponding to the timing of the regime switch, the economy will display a different unification process. First, if $T = 0$, at the instant 0^+, the parallel exchange rate will discretely jump from s_0 to ϵ_{0^+}, while the foreign reserves remain intact at its initial level R_0. In consequence, the economy will immediately move upward from the point F to E'' on the AA schedule. From 0^+ onward, the economy will remain intact at E''.[6] Second, if $T \rightarrow \infty$ (the announcement effect is negligible on the public), the parallel exchange rate will not jump, and the economy will move along the saddlepath trajectory SS to point E.

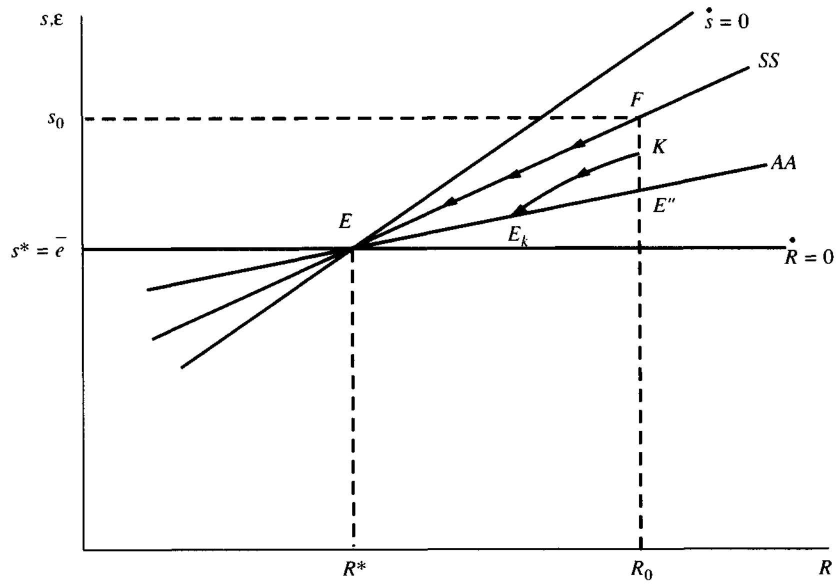
As claimed by Agénor and Flood (1992), corresponding to a specific value of T (namely T_c), the economy will happen to jump to a point such as D on the $\dot{s} = 0$ schedule at the moment of announced unification.[7] During the dates between 0^+ and T_c^-, the economy will move along the unstable branch DE_c. At the instant T_c^+, the exchange-rate unification comes into force, the system has reached the point E_c on the AA line. During the post-unification period, the economy will stay at E_c. As is evident from Figure 2, the parallel exchange rate continues to rise and foreign reserves continue to fall throughout the pre-unification period. If T (namely T_g) is greater than T_c, the system will follow an unstable branch GE_g during the time interval 0^+ and T_g^-. From T_g^+ onwards, the economy will remain intact at the point E_g. The dynamic path indicates that the parallel exchange rate will first fall then rise during the pre-switch period. However, if T (namely T_l) is less than T_c, the system will follow an unstable branch LE_l during the time interval 0^+ and T_l^-. From T_l^+ onwards, the economy will remain intact at the point E_l. The dynamic path indicates that the parallel exchange rate will keep rising during the pre-switch period.

Figure 3 illustrates the situation where the AA schedule is flatter than the SS schedule. Three distinct patterns of adjustment should be recognized. One possibility is associated with $T = 0$. If this is the case, the economy will jump downward to E'' on the AA schedule and stay at that point at all times. The second possibility is associated with $T \rightarrow \infty$. If this is the case, the economy will remain at its initial position F at the instant of unified announcement, and then move along SS toward the point E. The last possibility is associated with $0 < T < \infty$ (namely T_k). In this case, the economy will jump downward to the point K on impact and then move along an unstable trajectory like KE_k between 0^+ and T_k^-. At time

[6] It is clear from equation (16) that, if $T = 0$, ϵ_{0^+} is equal to initial money stock $m_0(= \theta R_0 + [1 - \theta]\bar{D})$. Thus, whether AA is steeper or flatter than SS is crucially related to the size of initial money stock.

[7] See footnote 21 of Agénor and Flood (1992).

Figure 3. *An Alternative Unification Process*



T_k^+, the exchange rate unification takes place, and thereafter the economy stays at point E_k. Under such a situation, the economy is characterized by both a falling parallel exchange rate and falling foreign reserves during the period prior to the unification.[8,9]

III. Concluding Remarks

The Agénor and Flood (1992) paper is an important addition to the literature on the dynamic adjustment of anticipated unification. In this comment, we have pointed out that some points are not fully explored in the Agénor and Flood (1992) work. Based on our analysis, given that the unified exchange rate under the flexible regime is greater than the

[8] To save space, we omit the case where *AA* is steeper than *SS* but flatter than $\dot{s} = 0$.

[9] Agénor and Flood (1992, p. 933) also refer to another situation where the unified exchange rate under the flexible regime is lower than the official exchange rate under the dual regime. Under such a situation, "a reform announcement would lead to an immediate appreciation of the parallel rate and a gradual increase in the stock of net foreign reserves."

official exchange rate under the dual regime, their conclusions should be generalized as follows:[10]

(i) The parallel exchange rate may either rise or fall at the instant of unified announcement;

(ii) The economy may display a variety of unified processes. In particular, both the parallel exchange rate and foreign reserves may fall throughout the pre-unification period; and

(iii) The unified exchange rate of the flexible regime is negatively related to the timing of unified implementation.

REFERENCES

Agénor, Pierre-Richard, Jagdeep S. Bhandari, and Robert P. Flood, "Speculative Attacks and Models of Balance of Payments Crises," *Staff Papers*, International Monetary Fund, Vol. 39 (June 1992), pp. 357–94.

Agénor, Pierre-Richard, and Robert P. Flood, "Unification of Foreign Exchange Markets," *Staff Papers*, International Monetary Fund, Vol. 39 (December 1992), pp. 923–47.

Flood, Robert P., and Nancy P. Marion, "Exchange-Rate Regimes in Transition: Italy 1974," *Journal of International Money and Finance*, Vol. 2 (December 1983), pp. 279–94.

Gros, Daniel, "Dual Exchange Rates in the Presence of Incomplete Market Separation," *Staff Papers*, International Monetary Fund, Vol. 35 (September 1988), pp. 437–60.

[10] Provided that (i) the coefficient θ is small, (ii) the *AA* curve is steeper than the *SS* curve, and (iii) the transition period is short, the following two conclusions reported in the Agénor and Flood (1992) analysis hold: the parallel exchange rate rises at the moment of unified announcement, and both the parallel exchange rate and foreign reserves fall throughout the pre-unification period. However, their third conclusion, which indicates that the unified exchange rate of the flexible regime is independent of the timing of unified implementation, does not hold.

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

Unification of Foreign Exchange Markets

A Rejoinder to Lai and Chang

PIERRE-RICHARD AGÉNOR and ROBERT P. FLOOD*

Professors Lai and Chang have provided valuable extensions to our analysis of exchange market unification. The use of the "terminal curve" *AA* in the graphical presentation of the unification process shows clearly the inverse relationship between the reform date and the unified exchange rate. Their contribution shows that the dynamic predictions of our model are potentially richer than we originally indicated.

However, having shown that almost anything can happen—depending on the position of the terminal curve *AA* and the reform date *T*—makes a comparison of the predictions of the model with the empirical evidence on exchange market unification all the more difficult. An examination of the most probable outcomes would therefore prove useful. The first issue is to determine the likely position of the terminal curve *AA*. Using the results established in our paper, it can be shown that the condition under which the curve *AA* is steeper than the pre-reform saddlepath is given by

$$2(1 - \alpha\Phi\theta) < \alpha\gamma + \nu.$$

If θ, the share of foreign reserves in the domestic money stock, is small—a likely situation for most developing countries contemplating exchange rate unification—this condition is likely to hold for reasonably high values of α (the interest elasticity of money demand), γ, and ν. Thus, in practice, the parallel market exchange rate is likely to jump upward upon announcement of the reform date, as emphasized in our paper.

The second issue that requires further investigation in the context of the unification process relates to the length of the transition period. In

* Pierre-Richard Agénor is an Economist in the Research Department. Robert P. Flood is a Senior Economist in the Research Department. He holds graduate degrees from the University of Rochester.

the context of the extended model presented in the second part of our paper, there exists a logical way of determining endogenously the optimal date of reform. Essentially, the idea is to minimize a loss function that highlights the trade-off involved in continuing the "old" regime—which entails a variety of costs, such as price distortions and diversion of export remittances to the parallel market—and implementing an immediate reform, because of the real costs that a short transition period or an "overnight" float may entail. As shown in the second part of our paper, the immediate increase in domestic wages and prices that takes place when the unified rate is expected to be more depreciated than the initial parallel exchange rate leads to an appreciation of the real exchange rate during the transition period, and a fall in output. A formal solution to this problem would further extend our understanding of exchange rate reforms.

IMF *Staff Papers*
Vol. 41, No. 1 (March 1994)

IMF Working Papers

Staff Papers *draws on IMF Working Papers, which are research studies by members of the Fund's staff. A list of Working Papers issued in 1993:4 follows.*

"The Role of Financial Institutions in the Transition to a Market Economy," by Hans J. Blommestein and Michael G. Spencer [93/75]
"Testing the Credibility of Belgium's Exchange Rate Policy," by Ioannis Halikias [93/76]
"Asset Price Inflation in the 1980s: A Flow of Funds Perspective," by Monica Hargraves, Garry J. Schinasi, and Steven R. Weisbrod [93/77]
"A Framework for Assessing Fiscal Sustainability and External Viability, with an Application to India," by Karen Parker and Steffen Kastner [93/78]
"Efficiency Wages and Labor Mobility in an Open Economy," by Pierre-Richard Agénor and Julio A. Santaella [93/79]
"Government Finance in a Model of Currency Substitution," by Anne Sibert and Lihong Liu [93/80]
"Supply-Side Economics in an Integrated World Economy," by Enrique G. Mendoza and Linda L. Tesar [93/81]
"Devaluation and Competitiveness in a Small Open Economy: Ireland 1987–1993," by Leonardo Bartolini [93/82]
"Expected Devaluation and Economic Fundamentals," by Alun H. Thomas [93/83]
"International R&D Spillovers," by David T. Coe and Elhanan Helpman [93/84]
"Japanese Banks and the Asset Price 'Bubble'," by Steven M. Fries [93/85]
"The Impact of Worldwide Military Spending Cuts on Developing Countries," by Tamim Bayoumi, Daniel Hewitt, and Steven Symansky [93/86]
"Nonmonetary Instruments and Their Use During the Transition from a Centrally Planned to a Market Economy," by Paul Hilbers [93/87]
"Financial Liberalization and the Information Content of Money in Indonesia, Korea, and the Philippines," by A. Javier Hamann [93/88]
"Price Controls and Electoral Cycles," by Pierre-Richard Agenor and Carlos M. Asilis [93/89]
"On Credible Disinflation," by Jorge E. Roldos [93/90]
"Labor Market Aspects of Industrial Restructuring in Canada," by Eswar Prasad [93/91]
"Recent U.S. Investment Incentives," by Christopher M. Towe [93/92]
"U.S. Health Care Reform," by Ellen M. Nedde [93/93]
"Endogeneity in Structural Unemployment Equations: The Case of Canada," by Caroline Van Rijckeghem [93/94]
"The Integration of World Capital Markets," by Morris Goldstein and Michael Mussa [93/95]

IMF Papers on Policy Analysis and Assessment

Papers on Policy Analysis and Assessment, a new series of research papers, are intended to make staff work in the area of policy design available to a wide audience. A list of all PPAAs issued in 1993:4 follows. Papers may also be considered for inclusion in the journal.

"The World Market for Natural Gas: Macroeconomic and Financial Implications," by Thomas P. Enger [93/15]
"Aspects of the Design of Financial Programs with the Adoption of Indirect Monetary Controls," by R. Barry Johnston [93/16]
"Selective Government Interventions and Economic Growth: A Survey of the Asian Experience and Its Applicability to New Zealand," by Jonathan D. Ostry [93/17]
"The Structural Crisis in the Swedish Economy: Role of Labor Markets," by Ramana Ramaswamy [93/18]
"The Taxation of High Income Earners," by Parthasarathi Shome [93/19]
"International Capital Transactions: Should They Be Restricted?" by Norman S. Fieleke [93/20]

An annual subscription to IMF Working Papers is available for US$200.00. It includes 12 monthly shipments and priority mail delivery. Requests should be made to:

International Monetary Fund
Publication Services
Washington, D.C. 20431, U.S.A.

Telephone: (202) 623-7430
Telefax: (202) 623-7201

Inquiries about individual IMF Papers on Policy Analysis and Assessment should also be directed to the address above.

In statistical matter throughout this issue,

dots (. . .) indicate that 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, 1992–94 or January–October) to indicate a total of the years or months inclusive of the beginning and ending years or months;

a stroke (/) is used between years (for example, 1993/94) to indicate a fiscal year or a crop year;

a colon (:) is used between a year and the number indicating a quarter within that year (for example, 1994:1);

components of tables may not add to totals shown because of rounding.