INTERNATIONAL MONETARY FUND

STAFF PAPERS

CONTENTS

VOL. 35 No. 3 SEPTEMBER 1988

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 center 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. The authors have received considerable assistance from their colleagues on the staff of the Fund, and this general statement of indebtedness should be accepted in place of a detailed list of acknowledgments.

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$18.00 a volume or the approximate equivalent in the currencies of most countries. Four numbers constitute a volume. Single copies may be purchased at $5.00. Special rate to university libraries, faculty members, and students: $9.00 a volume. Subscriptions and orders should be sent to:

Publication Services
International Monetary Fund
Washington, D.C. 20431, U.S.A.

Telephone number: (202) 623-7430
Cable address: Interfund

INTERNATIONAL MONETARY FUND

STAFF PAPERS

Vol. 35 No. 3 SEPTEMBER 1988

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

CONTENTS

Vol. 35 No. 3 SEPTEMBER 1988

This page intentionally left blank

An International Debt Facility?

W. MAX CORDEN*

The proposal to set up an international debt facility to buy the debt of developing countries at a discount and then mark down its contractual value is analyzed. The paper considers the central question of how the debtor countries, creditor banks, and owners of the facility would be affected; in particular, what redistribution of gains and losses there would be among them. The "market price effect" and the "ceiling effect" are distinguished. A crucial consideration is whether debt retained by banks is subordinated to debt bought by the facility.

A COMMON PROPOSAL designed to deal with the developing countries' debt problem is that there be set up an "international debt facility" that would buy debt at a discount and then mark down its contractual value, hence providing debt relief. This facility could be envisaged either as a major scheme that would, over a period, deal with most or all outstanding commercial debt owed or guaranteed by governments, or as a more modest arrangement dealing with only small portions of debt, possibly only that which is owed by the governments of particular countries.

Many such proposals have been advanced. They vary in their details,[1] and there are many difficulties, some major. Nevertheless, the frequency with which such proposals are made makes it worth examining them carefully in their many permutations.

* Mr. Corden is Senior Advisor in the Research Department, on leave from the Australian National University, where he is Professor of Economics. He is a graduate of the University of Melbourne and the London School of Economics and Political Science. He has taught at Oxford University and at various Australian and U.S. universities. He thanks Ken Rogoff and colleagues in the Fund for helpful comments.

[1] As far as I am aware, the first proposals of this general kind were advanced by Felix Rohatyn in *Business Week* (February 28, 1983) and by Peter Kenen in *The New York Times* (March 6, 1983). The proposal has been made by Sachs and Huizinga (1987) and by Percy Mistry, formerly of the World Bank (in *The*

I. Main Issues

There are three principal parties to the proposed transaction—the debtor governments, the creditor banks, and the "owners" of the facility. The central question is how the costs and benefits would accrue to various parties. Is there an element of "foreign aid" or of a "bank bail-out"? Alternatively, would the banks give up something? Could all three parties gain or, at least, could some gain without the others losing significantly, if at all? In other words, is there some systemic benefit?

The proximate redistributive effects—and possibly also the systemic effects—will depend crucially on three prices: the price at which the debt is bought, the price or value to which it is marked down, and the price or perceived value to which remaining debt that is retained in the private sector moves as a result of the whole operation. The full economic effects will depend, of course, on how the various parties react to or deal with the proximate gains or losses.

In considering the details of such a scheme there are many choices to be made.

- The debt might be bought by the facility at the minimum price required for the banks to part with it voluntarily; it might be bought at current market prices; it might be bought at the market prices that existed at some earlier "cut-off" date; or it might be bought at some other set of prices representing discounts on the contractual value. Conceivably it might even be bought at its contractual value.
- Purchased debt might be marked down to the cost at which the facility bought it, or to a higher or lower value than cost.
- The debt that is not sold by banks might maintain its present contractual status; it might be subordinated to the debt that the debtor countries will now incur to the facility; or it might be marked down by the debtors to an extent that would force the banks to sell all their debt to the facility.

A crucial question is how the facility would be financed. Here, also, there are differences among the various proposals, and the possibilities will be discussed shortly.

Banker, September 1987). In 1988, proposals of this general nature have been made by Dr. Sengupta, an Executive Director of the Fund, and by James Robinson, Chairman of American Express. There is an analysis of this kind of proposal in Feldstein and others (1987). The Omnibus Trade and Competitiveness Act of 1988, passed by both houses of the U.S. Congress, included a provision for the Secretary of the Treasury to "study the feasibility and advisability" of establishing an "International Debt Management Authority."

II. A Simple Scheme

Let us suppose that the scheme applies to any one debtor country. The facility goes into the market and offers to buy given amounts of debt. In the detailed example spelled out below, it will be assumed that the facility offers to buy half the stock of debt. Of course debt is not homogeneous, so that various decisions would have to be made on which debt to buy. It is quite likely that the facility would have to pay more than the initial market price, but one can assume that the facility would buy debt at a discount from the contractual value. The important question of what would determine the price, and in which direction it would move, will be taken up later.

The facility pays for the purchased debt with new bonds guaranteed by its owners. The banks are thus able to dispose of debt with the original contractual value that is subject to default risk in exchange for debt of a lower contractual value that is subject to much lower, possibly zero, default risk. One's first thought is that those that sell could not be worse off as a result. After all, selling is voluntary; there is no compulsion in this scheme. This conclusion is not necessarily true, and will have to be looked at again in Section IV.

The facility would then mark down the contractual value of the debtor country's debt that it has acquired. It marks the debt down to the cost price to it—that is, the contractual value of the new bonds it has issued to the banks. No funds are thus required from the owners of the facility—that is, the governments that have underwritten the facility. But, of course, the facility's new assets are somewhat risky and, because of the guarantees on the bonds it has issued, this risk has been taken over by the facility's owners. Given this risk, there will be a potential need for funds from the owners, who may actually wish to finance contingency reserves specifically to allow for the risk. A major, and possibly overwhelming, obstacle to the establishment of a facility is the reluctance of governments to assume such risks. A question to be discussed below is whether this risk can be reduced or eliminated.

The debtor country apparently benefits because the contractual value of its debt has been reduced. But the gain to it will not necessarily be as great as it might seem at first. One possible view is that the market's perception of default risk, which led to the initial discount, was justified in the sense that this represented the true probability of default. In other words, there was a good chance in any case that the country would not repay the full contractual value of its debt. Reducing the contractual value as a result of the operation of the facility would not necessarily reduce actual payments (or the probability of actual payments expected

to be made) to the same extent. Indeed, one might ask whether there is likely to be any gain to the debtor country at all.

III. How Debtor and Banks May Gain

This matter of the possible gain to the debtor, and also to the banks, can be analyzed more precisely if the concept of the debtor's "capacity to pay" is introduced. This is defined as the ability to make resource transfers abroad to cover interest payments and repayment of principal. It depends on the country's output over a period of time, on its minimum consumption level, and on its ability to transform output into tradables (exports and import-competing products) required to generate the transfer. It also depends on the terms of trade.

It will be supposed at this stage that capacity to pay can vary as the result of various exogenous, uncertain events—such as terms of trade developments—but does not vary because of changes in the policies of the debtor country itself, which are simply taken as given. This is the assumption of "exogeneity." It will also be assumed for the moment that expectations about capacity to pay are the same among market participants, debtor countries, and the decision makers of the facility. This is the assumption of "uniformity of expectations." These two simplifying assumptions are important for the analysis of gains and losses from the establishment of a facility and therefore will be reconsidered in Section VII.

There are two steps in the analysis. First I show why the banks might gain at the expense of the facility, and then I show why the debtor might gain. The second effect depends crucially on uncertainty.

Gain to the Banks

To begin with, there is the "market price effect."[2] It can be shown that the banks will gain at the expense of the facility, provided that the debt that they retain is not subordinated to the marked-down debt the facility now holds. The reason is that the market price of the debt will rise (the discount will fall). The argument is quite simple when there is complete certainty about the capacity to pay (or repay).

Let us suppose that the contractual debt is US$1,000 and the capacity to pay is $600. Assume at this stage that the latter is fixed. Hence the debtor country will neither gain nor lose; whatever happens, it pays $600. This is thought of as a single sum paid in a single future period, the

[2]This discussion builds on Dooley (1988).

sum consisting of principal and interest combined. Given the initial contractual debt, default or debt relief is then inevitable. The "default ratio" would be 40 percent.

The facility then offers to buy a proportion of the $1,000 in contractual debt from the banks at a discounted price. Here it will be assumed that it buys half the debt ($500) and that the price it pays is 80 cents to the dollar. Hence the facility pays $400 and marks the debt it has acquired down to its cost price. The contractual value of the total debt owed by the debtor country is thus reduced to $900. With the same capacity to pay as before, the default ratio becomes 33.3 percent. The facility will finally get 66.7 percent repayment of the debt it holds, thus making a loss of $133. The banks will get $333 for the debt they have retained (with a contractual value of $500), and when this is combined with the $400 they received from the facility they end up with $733, which is an improvement of $133 on what they would have received if the facility had not bought and marked down some of the debt. The discount on debt held by the banks has fallen from 40 percent to 33.3 percent.

In this example the facility's purchase price is 80 cents but the market price has risen only from 60 cents to 67 cents. The purchase price could therefore be reduced, leading to a bigger decline in the contractual value and, hence, to a further rise in the market price. The equilibrium price (where purchase and market prices are equal) would actually be 70½ cents when the facility buys half the debt. If it bought a greater proportion, the price would be higher. These results can be derived as follows:

$$C_2 = C_1(1 - q) + C_1 qp \tag{1}$$

$$p = R/C_2, \tag{2}$$

where

C_1 = initial contractual value
C_2 = contractual value after debt relief
q = proportion of debt bought by facility
R = capacity to pay
p = purchase price (equal to market price after purchase) as proportion of initial contractual price.

From equations (1) and (2),

$$R/C_1 = (1-q)p + qp^2. \tag{3}$$

From equation (3),

$$p = \frac{-(1-q) + \sqrt{(1-q)^2 + 4qR/C_1}}{2q}.$$

There has been a pure transfer from the facility to the banks. All this will be reflected in the market price rising (discount falling) when the facility enters the market. It has to pay a higher price than the initial price to induce the banks to sell any debt to it. The banks will foresee that debt not sold would rise in value when some marking down takes place, and hence they will only sell at a sufficiently higher price. The price would not necessarily rise to its equilibrium value immediately, and could also overshoot, since banks and others in the market would not be able to predict this equilibrium in advance. The account given here, with its impression of precision, merely indicates likely tendencies.

The essential point can be restated as follows. When the contractual value of the total debt is reduced while total capacity to pay stays constant, each dollar's worth of contractual debt must be worth more in the market than before, provided that all dollars of contractual debt would be treated equally if there were some default.

Gain to the Debtor

Uncertainty about capacity to pay and the "ceiling effect" can now be introduced.[3] The mean expected repayment might be $600, but it could also be greater, up to a ceiling of complete repayment of $1,000, and it could be less, with a floor of zero. There is thus both upside and downside risk, and this will be taken into account in the market price. If the contractual value of the debt is reduced, say, to $900, the ceiling will be lowered to $900. If the terms of trade, for example, turn out to be quite favorable, so that capacity to pay is actually $950, the actual payment will be $50 less than if the contractual value had stayed at its initial level. Thus the debtor gains at the expense of creditors from a markdown of the contractual value because the downside risk remains as before, whereas the offsetting upside risk (or gain) becomes less.[4]

[3] In several papers Paul Krugman has discussed the uncertainty aspects; see especially Krugman (1985).

[4] The example that has been used is quite simplified, although sufficient to make the main points. As noted earlier, repayment is thought of as a single sum ($600, when there is certainty) paid in a single future period, the sum consisting of principal and interest combined. The analysis could be elaborated to allow for a stream of interest and amortization payments over time, in which case the sum should be thought of as the present value. There is then scope for changes in the time profile of payments. In that case a distinction between interest and principal would have to be made. Debt relief may have an immediate effect in reducing interest payments, even though, if capacity to pay in total is really fixed, interest or amortization payments will increase later. Changes in the time profile of either interest or amortization payments that do not alter the present value leave the analysis presented above unchanged. The market discount is caused not only by the probability of default or forced debt relief as usually understood but

IV. Subordination: Can a Gain to the Banks Be Avoided?

An interesting question is whether a gain to the banks at the expense of the facility—that is, a "bank bail-out"—can be avoided. The key here is subordination of the debt retained by the banks relative to the debt now owned by the facility.

When one talks about a gain to the banks, one means a gain relative to the initial situation when there was already a discount in the market. Earlier, of course, the banks incurred a loss once the probability of some default or forced debt relief was perceived by them or the market. Presumably, as long as the banks get less than $1,000 they will have incurred some loss as normally defined, even though the margin above the London interbank offered rate (LIBOR) they charged originally must have taken into account the possibility of some default or of heavy pressure to provide some relief.

Suppose that, again, the facility buys half the debt and marks it down to cost. It buys it at 80 cents per dollar and so pays $400, total contractual debt being thus marked down to $900 as before. One now proceeds in two stages. First, let us assume that there is no doubt at all that the capacity to pay will be at least $400.

If it could be firmly established that, whatever is the capacity-to-pay outcome above $400, the debt held by the facility would always be paid first—that is, would be "senior" debt—then the facility would not make a loss, and its owners would run no risk. But the banks would lose potentially, and the debtors gain, because the "ceiling payment" has been reduced. Previously the maximum payment the banks could have received was $1,000, whereas now it is $500 for the debt they have retained plus the $400 they have already received from the facility. If capacity to pay turned out to be $950, previously the banks would have received $950, but now they can only receive $900. In contrast, the minimum they can receive remains $400.

Subordinating debt retained by the banks to facility-held debt thus ensures that the facility neither loses nor gains, taking on no risk, while the debtor countries gain potentially at the expense of the banks because of the ceiling effect. The expected loss to the banks would be reflected in a decline in the market price.[5]

---

also by the probability of forced rescheduling, pressures to participate in new money packages, and so on. These are all ways of changing the time profile and reducing the present value of repayments.

[5] An issue that is clearly important for the various proposals is whether it would be legally possible for existing debt that is retained by the banks to be subordinated to that acquired by the facility. Of course there would be no difficulty in such subordination if it were done with the agreement of the banks.

If it were desired for some of the loss to the banks to be shared with the facility, the latter could mark down the debt by more than the discounted purchase price, hence making a clear loss now. Alternatively, only part of its debt might be given seniority. Here there is scope for many variations in the details of such a scheme, and these may have significant effects on the gains or losses for the banks and the facility. The key point is that the banks and the facility combined must make a potential loss—that is, forgoing the benefits of a very favorable capacity-to-pay outcome. The risk of an unfavorable outcome has not been eliminated, but the possibility of a very favorable return (above $900) has.

The second stage of the analysis is to assume that capacity to pay could be less than $400. In that case there would be some possibility that even debt given senior status would not be fully repaid. Hence risk for the facility would not be completely eliminated. The conclusion that subdination of debt retained by the banks to the debt owned by the facility would eliminate all risk for the facility hinges completely on the assumption that some minimum total payment—equal to the value of the debt that the facility has bought ($400 in the example)—is utterly assured. But the larger the proportion of the initial total debt that the facility takes over and marks down, the less likely it is that all risk for the facility would be eliminated by giving the debt it holds senior status. If the facility had bought up all the debt, no one but the facility could assume the risk.

There can never be utter certainty about prospective capacity to pay, so that some risk for the facility is inherent in the scheme. This implicit risk is particularly relevant for proposals that would have the facility take over a large part of the foreign debt of a country. The inevitable risk helps to explain the reluctance of governments of creditor countries to support proposals for such a facility.

V. Reduction of Uncertainty

Another possibility, worth exploring carefully because it is implicit in some proposals, can now be considered. The suggestion is that the Fund or World Bank may be able to increase or ensure certainty of payment at the new, marked-down value of the debt. The assumption maintained so far—that the actual repayment outcome depends only on exogenously determined capacity to pay, subject to the "ceiling" imposed by the contractual value—is relaxed. Repayment can also depend on policies.

Suppose that initially the mean expected capacity to pay was $600, with a probability of creditors getting more or less. If the total debt were marked down to $600, there would then be a $600 ceiling to the repay-

ment. In addition, suppose that the Fund or World Bank were able to ensure that $600 also became the minimum repayment. This assurance might have been obtained with the aid of conditionality. Given this arrangement, there is no longer a necessary loss to the banks and the facility combined from the imposition of a reduced ceiling because that ceiling is associated with the imposition of a raised floor. Upside and downside risk have both been eliminated. Certainty has been obtained—or at least uncertainty has been reduced.

Certainty represents a net gain for the banks and the facility combined, given that they are risk averse. With subordination, the whole of this gain from certainty would go to the banks in the first case just discussed, in which a minimum repayment sufficient to cover the debt held by the facility was in any case ensured. But in the more general case, in which the facility has carried some of the risk previously (whether because no minimum was ensured or because there was no subordination), the gain would go partially to the facility.

It is often implied in debt relief proposals that the marked-down value of debt would have no more risk attached to it (or very little risk) because it would be close to the expected capacity to pay. It is doubtful whether this assumption is realistic. The implication is that willingness to repay—and the resolve to make the necessary adjustments—is not exogenous but rather can be made more "certain" in return for debt relief. Perhaps a commitment that would successfully reduce perceived default risk could be obtained from the debtor country in some way or other. Debtor governments could make certain policy commitments. No doubt the Fund's conditionality procedures can play a role here. Conditionality can conceivably reduce uncertainty and default risk, although it can surely not eliminate them.

A reduction in uncertainty of repayment without necessarily any net change in the mean expected repayment is clearly a gain to the banks and the facility. But it is not necessarily a gain for the system as a whole. If uncertainty in the capacity to pay—for example, uncertainty in terms of trade movements—could be reduced, that would be a net gain. But if uncertainty in capacity to pay continues while repayment becomes more certain owing to conditionality, there has simply been a transfer in the burden of uncertainty toward the debtor country. For example, if the terms of trade turned out to be particularly adverse, the country would have to bear the whole burden instead of sharing it with the banks or the facility through some degree of default or debt relief.

This approach assumes that the mean expected capacity or willingness to pay stays unchanged but that the floor is raised and the ceiling lowered. There are also two other possibilities.

The first is rather similar to the one just discussed. Conditionality may

raise the mean without raising the floor: it may raise expected capacity to pay through bringing about an improvement in policies. Although the banks lose through the ceiling effect, they may then nevertheless gain.

The other, quite contrary possibility is that the probability of repayment is actually reduced when the facility takes over debt. As just noted, it is usually argued that, through associating conditionality with the establishment of a facility, the certainty of repayment can be improved. But a contrary view is that a facility that is subject to political pressure and that has no strong penalties available to it may be a less effective debt collector than private banks, which can threaten the withdrawal of trade credit as a potential penalty.

VI. Interests of the Debtor Countries and Moral Hazard

There are several ways in which the debtor country might gain from the arrangement. Some have already been referred to, but they will now be brought together.

Reducing the Default Ratio

A gain that seems obvious at first sight but turns out to be primarily cosmetic is the reduction in the default ratio. The default ratio, D, equals $1 - R/C$, where R is the actual debt repayment made—that is, the resource transfer—and C is the contractual value of the debt. D is reduced when the contractual value is marked down, even though the actual payment (which has been assumed to be exogenous so far) does not change.

Does it really matter if this default ratio falls, possibly to zero, when the resource transfer remains unchanged? One might say that the effect is purely cosmetic. If an emperor has few clothes, is it really necessary to proclaim the fact? Against this it can be argued convincingly that debt relief voluntarily provided by the creditors is always better than default.

There would clearly be a preference on the part of the debtor country for debt relief over default if penalties were associated with default. Even in the absence of current penalties, reputation—and hence future creditworthiness—may be influenced by whether there has been formal default rather than debt relief. Note that it has been assumed here that default depends purely on exogenous capacity to pay; hence penalties related to the default ratio would seem less likely or reasonable. Because capacity to pay completely determines actual repayments, there would be no point in the creditors imposing penalties.

Lowering the Ceiling

The debtor country gains owing to the "ceiling effect." As has been pointed out, if capacity to pay turns out to be particularly favorable—above the new contractual value—the gain would go to the debtor rather than to the creditors because the ceiling for potential repayments has been lowered. This benefit to the debtor might disappear (and could even turn into a loss) if conditionality manages also to raise the floor for the repayment, shifting more of the risk toward the debtor country.[6]

Marking Debt Down Below Cost

The facility might mark the debt down by more than the cost price to it—possibly by much more—and the contractual value might then fall below the capacity to pay, even when the capacity to pay turns out to be quite low. At the limit, the debt might be marked down to zero. This would represent a straightforward transfer from the owners of the facility to the debtor countries—a case of foreign assistance. It is equivalent to the owner countries donating funds to the debtor country to buy back its debt.

Asymmetric Expectations

A fourth kind of gain has not been referred to so far but is implicit in much advocacy of debt relief and could be important. The markets, specifically the banks, may be pessimistic and believe that there is some probability of default. Hence there is a market discount on the debt. But the government of the debtor country may have no intention of defaulting. There are "asymmetric expectations." Capacity to pay, after all, is not something clear-cut. The government foresees difficulties and adjustment problems and seeks debt relief but—possibly for fear of penalties—does not intend ever to default, even though it has not succeeded in convincing the market of this. The issue of asymmetric expectations

[6] There may be a touch of perversity in the ceiling effect brought about by debt relief. Whenever capacity to pay improves exogenously—owing, for example, to a terms of trade improvement—some of the gains inevitably go to the debtor even before debt relief (that is, when the ceiling is high). Similarly, some of the losses from a deterioration would be borne by the debtor, and not wholly by the creditors. In that case, lowering the ceiling as a result of debt relief increases the gains for the debtor when events, such as the terms of trade, turn out well but does not help when events turn out badly.

will be discussed further below. Here it can be noted that, if the government of the debtor country has no intention at all of defaulting, the whole of the fall in the contractual value of the debt brought about by debt relief through the operation of the facility or in some other way would represent a clear-cut gain to the debtor country in reduced prospective resource transfers.[7]

The creditors, however, having different expectations, do not perceive this reduction in the contractual value—or all of it—as a loss to them. They may expect to lose through the ceiling effect but also see some virtue in an explicit recognition of what they believe to be realities—that the country has limited capacity or willingness to pay, that the emperor has fewer clothes than the initial contract specified.

Moral Hazard

For three of the four reasons given here (other than the third, marking debt down below cost), the debtor country would want the price at which the facility buys debt from the banks to be as low as possible. The lower the price, the greater is the decline in the contractual value; hence, the lower the default ratio, the lower is the ceiling applying when events turn out favorably, and the lower are actual repayments if default is never intended.

If this purchase price is equal to or closely related to the market price, the debtor country therefore has an incentive to get the market price down. This can be done by making "default noises"—just a hint here, a threat there—and the banks will be glad to sell at a low price, in the extreme case at any price above zero. This is the familiar "moral hazard" problem.

A possible solution from the point of view of creditors seems to be for the facility's purchase price not to be determined by the market price, or at least by the market price ruling once the likelihood of such a facility being established has become serious. Market prices at some earlier "cut-off" date might be taken. If the banks are to sell voluntarily, the purchase price will have to be no lower than the current market price. But it could be higher.

[7]There is an important qualification to this argument. If the debtor country's government takes the long view, it will realize that debt relief through the facility or otherwise—even though entirely voluntary on the part of the creditors and in no way associated with actual default—could still have an adverse effect on its country's future creditworthiness. After all, when investors look back they will see that a $1,000 loan finally turned out to be worth less, for whatever reason. The government will never have the opportunity to show that it would have paid the full initial contractual value.

The problem is to fix a price that does not give a gain, or an undue gain, to the banks; otherwise there would be a bail-out. But what is a gain? Given the expectations created by their anticipation of the debtor country's capacity to pay, combined with the default noises made by the debtor government or others in that country, a sale of the debt to the facility at a very low price may still seem to be a gain to the banks. This is true even though the price is likely to represent a loss relative to the expectations at the time the loans were originally made. Presumably the facility should aim to avoid either gain or loss to the banks relative to the situation at some "pre-discussion-of-facility" date—that is, an appropriately early cut-off date.

Most proposals for a facility do not pursue in detail the critical question of how the price at which debt is to be purchased is to be determined, given that there is a moral hazard problem and that there must presumably be a separate price (discount) for each country. It is at this point that the greatest practical problems arise. Sometimes elaborate calculations, which are essentially estimates of capacity to pay, are proposed, but the political difficulties such estimates would involve cannot be ignored. Given the thinness of existing markets, actual market prices, whether current or at some earlier cut-off date, may not be adequate guides.

VII. Two Assumptions Reconsidered

At the beginning of this paper two crucial assumptions were made: that the debtor's capacity to pay was exogenously determined—for example, by the terms of trade—and that expectations about the capacity to pay were the same among all the relevant parties. Given these assumptions, a fairly straightforward analysis followed that showed that a facility would yield a gain to the banks because of the market price effect and a gain to the debtor because of the ceiling effect. These gains would be at the expense of the facility, which would be taking over a risk. It was further assumed that debt owed to the facility would not be given seniority over debt retained by the banks. If the latter were subordinated, a gain to the banks and loss to the facility might be avoided.

Subsequently, the two initial assumptions have been removed in particular ways. In Section V, the possibility was explored that the facility (or the World Bank or Fund acting on its behalf) could actually affect the debtor's policies so that capacity to pay would be improved to ensure certainty of repayment of the marked-down value of the debt. In other words, capacity to pay might no longer be exogenous. In Section VI, one

case of asymmetric expectations was noted. The debtor government may have no intention of defaulting, but the market may not be convinced of this. In addition, moral hazard was introduced. Prospective repayment may depend not only on *capacity* to pay but also on *willingness* to pay (for given capacity), and threats of reduced willingness would affect the market price.

These complications to the initial approach are really special cases, but there are further cases that analysts of these issues sometimes have in mind. A more systematic approach is therefore needed.

First of all, the concept of expected capacity to pay determined by exogenous factors could be redefined as "expected total repayment" determined both by expected capacity to pay and by expected willingness to pay.[8] Both would be influenced, or even determined, by policies. When the original concept of capacity to pay is broadened in this way, it becomes more plausible. If the redefined concept is to apply to the initial analysis in this paper, it has to be assumed that expected policies are exogenous in the sense of not being expected to change as a result of the establishment of the facility or its activities.

The next step is to allow for endogenous policies affecting capacity and willingness to pay. The endogeneity of policies is central to many debt-strategy proposals. As noted in Section V, the basic idea is that the benefit to the debtor from debt relief provided through the intermediation of the facility would be reciprocated by improvements in the debtor's policies, and that some kind of assurance about these policies can be obtained, perhaps with the help of conditionality. In this way more certainty of repayment can be ensured.

With regard to endogenous willingness to pay, two points are usually made. The first, as noted above, is the moral hazard problem: threats of reduced willingness to pay can get the market price down. A second idea not mentioned so far is that, when the contractual value of the debt is partially forgiven so that it is brought down to a more realistic level, the debtor government may have a greater willingness to repay the remainder. If the contractual debt was $1,000 and capacity to pay was $600, some default would be inevitable. It has then been argued that a large default is as bad—and incurs similar penalties—as a more modest default, so that willingness to pay in that case might fall to zero. But if the contractual debt were marked down to $600, there would be a good chance that default could be avoided, and willingness to pay might become 100 percent.

[8] All this should be thought of in present-value terms. See footnote 4 in Section III.

As regards asymmetric expectations about the capacity and willingness to pay, there are several possibilities worth noting. First, as already mentioned, the debtor may not intend, and hence not expect, to default while the market contrarily believes that there is a positive probability of default, thus explaining the market discount. In that case the debtor government will believe it would gain from any debt relief, whereas the creditors—selling their debt voluntarily on the market (and assuming no subordination)—will not expect to lose. If the facility paid the banks a price above the initial market price—still with a discount—the banks may believe that they would actually gain, even though, if the debtor government's expectations were correct, the banks would actually have lost by selling.

A scheme could conceivably be worked out whereby the facility pays, for example, $400 for debt with a contractual value of $500 and marks that debt down to only $450, with the margin of $50 adequately compensating the facility for the risk it incurs so that it neither gains nor loses. In this case the creditors believe that they gain through the market price effect, the debtor government believes that it gains because there is some reduction in the contractual price, and the facility neither gains nor loses.

This scheme leads into the second possibility, whereby the facility actually makes a profit or at least is expected to do so by its owners or managers. The market may have an unduly pessimistic view of the debtor's prospects, and hence there may be a large market discount. But the facility only marks the debt down by a little, so that the contractual value of the marked-down debt it holds stays well above the cost price and there is a high degree of certainty that there will be low or zero default. All this depends on confidence in the ability to get the debtor's policies improved sufficiently to disprove the market's pessimistic expectations.

Finally, it has been argued in the main analysis here that the market price effect represents a benefit for the banks, at least relative to the situation after the debt crisis and the discount developed. But there may be some holders of debt who do not sell to the facility because—contrary to the expectations of marginal holders—they do not believe that the probability of default is high at all. They may value the debt they hold at the contractual value, not near the market price. They may have made a more optimistic assessment of capacity or willingness to pay. If they feel assured that there will be full repayment in any case, it would make no difference to them if the total contractual claims are reduced through the operation of the facility. But if they really believed that the debt is worth more than its market price, the question then arises why they did not buy up the debt held by others and so bring the price up to their

optimistic expectations. The argument assumes that the market is, in some sense, imperfect.

VIII. Would New Investment Increase as a Result of the Facility?

There are three parts to the answer to this question. If the debt of the facility is given senior rights the answer is not clear; it is possible that new investment would actually be discouraged.

First, the analysis has shown that for various reasons there may be an actual reduction in resource transfer from the debtor country as a result of the facility—that is, the debtor country may actually gain something. Indeed, in the view of some this is the primary objective of the exercise. An expectation of such a gain would lead also to an expectation of lower taxation than otherwise—including taxation of profits and capital—and this may well encourage new investment.

Second, if the debt held by the facility does not acquire senior rights, so that the discount in the market falls as described earlier, there should indeed be a tendency for investment inflows to resume or to increase. The facility will have assumed some of the burden of potential default on the existing debt, and new investors will have a lesser burden to bear than before.

Finally, the matter is not so simple if the existing debt is subordinated to the facility's debt. The question then is whether new debt incurred in the market would also be subordinated, or whether it would acquire seniority over the facility's debt. The reasonable assumption is that the facility would enjoy complete seniority. As noted above, in the absence of increased certainty, subordination would actually reduce the market price (raise the discount) owing to the "ceiling effect"; hence new investment would be further discouraged. If all old debt had been sold to the facility, in effect new debt would then be subordinated to old debt completely.

IX. Is There Really Need for an International Facility?

A central question remains. One might grant the desirability of a reduction in the contractual value of the debt but still wonder why an intermediary in the form of a facility along the lines proposed would be needed.

Although banks can sell the developing countries' debts in the market at a discount, bank managements may not feel free to grant outright relief in the form of reduction of the contractual value, possibly because of legal obstacles. In practice, however, relief in the form of long-term debt rescheduling and various debt transformations can be and has been granted—although such arrangements are different from reducing contractual value. One could also argue that there is no incentive for any private holder to grant relief because of the ceiling effect. There is always the possibility that the full contractual value will be repaid, so why forgo this possibility? Incentive for relief may, however, be created by the threat of more severe default.

One can think of three arguments in favor of the establishment of a facility from the points of view of the banks and the debtor countries involved.

A Channel for Resource Transfer

The most obvious argument from the point of view of both parties is that the facility could act as a channel for the transfer of current resources (that is, aid) from the countries that underwrite it, or for the possible transfer of future resources if some default risk is perceived. This, of course, is also an argument *against* a facility from the point of view of its potential owners if they are not interested in providing aid either in general or specifically to the debtor countries. This point will be taken up again below.

If foreign aid to the debtor countries is indeed intended, one alternative could be for the parties to negotiate debt relief contracts bilaterally and then for some or all of the industrial countries to guarantee the marked-down debts in part or in full. This arrangement would give particular industrial countries an opportunity to help those debtors that are of special interest to them. The familiar difficulty here is that the banks are not a single "party," as the problems of organizing concerted lending have shown.

For the debtor countries the other alternative is to receive direct bilateral aid. The aid could be used by the debtor country to buy back some of its own debt. Again, there would be an opportunity for industrial countries to discriminate in favor of particular debtor countries. But the fundamental question is highlighted in that case: whether funds received in aid are best spent in buying back debt. They could perhaps be better used to finance extra investment.

A More Orderly Process

It could be argued that, if world economic conditions turned adverse, the alternative to the operation of such a facility would be a decentralized process of debt restructuring with relief. In that situation numerous bilateral arrangements—with the banks represented by committees that have difficulty in getting support from sufficient banks-—could get rather disorderly. The facility could be an intermediary that would bring more orderliness to the process. An element of automaticity and consistency across countries and kinds of debt in the choice of purchase prices, the extent of relief, and so on, could smooth the restructuring and debt relief process. The facility might thus avoid default crises that could lead to political difficulties and disruption of trade flows.

Greater Realism and Certainty

A key feature of such proposals is that very uncertain obligations, with contractual values well above what is expected to be paid on a probability basis, would be replaced by more realistically valued debt that (in the view of its proponents) would be more certain to be repaid and, ideally, would be free of serious default risk.

It might be argued that the increase in certainty (if it could be obtained) is in general desirable even though it does, to an extent at least, shift the burden of exogenous uncertainty (for example, in the terms of trade) back toward the debtor countries. This is possibly a gain because some uncertainty is believed to be endogenous—a result of the lack of firm political will by debtor governments rather than capacity-to-pay uncertainty. Then there is a role for conditionality and, hence, for the Fund or the World Bank. This does not necessarily mean that the two institutions, or their owners, should, through the facility, take on the remaining risks.

One negative point is also important. It refers not to the actual operation of a facility but to the effects of expectations that it might come into operation. It concerns a moral hazard problem. If the banks and the debtor countries believe that there is some chance that an institution such as the facility might be established to take over some of the risks, they will have less incentive to arrive at debt relief agreements directly or without disruption. The threat of disruption, particularly of trade flows, could be an inducement leading the international community to establish such a facility. But if such an institution were never seen as

being even a possibility, the parties directly involved would have an incentive to arrive at agreements. They would try to avoid prolonged uncertainty and disruption because it is damaging to them all.

X. Is Any Compulsion Involved?

To what extent would such a scheme be voluntary? I first consider the debtor country and then the creditors.

On the Debtor

On the one hand, if conditionality were not involved, a debtor country would have nothing to lose in the short run from debt relief through the medium of the facility. But in the long run it might lose some creditworthiness, since future creditors may well think that what has happened once can happen again. Therefore a debtor government, confident that it will be able to repay the full contractual value of its debt and wanting to take a long view, may benefit from staying out of the scheme. This is true even though there may be a market discount on its debt that suggests that, so far, the debtor country has not been able to convey its confidence to the market.

On the other hand, if conditionality were part of the scheme, then each debtor country could decide whether it preferred to accept the burdens of conditionality and get debt relief through the facility, or whether it preferred to stay out. There would not need to be any compulsion.

On the Creditor

Each bank could be free to sell or to keep as much as it liked of the debt it holds at present. Sales of debt need not be compulsory. As described here, the facility would operate in the market, even though this is not a feature of all proposals. But this freedom of the banks to sell or not to sell could be somewhat illusory. The willingness to sell will be influenced by the debtors' actions. A decision by the debtor government to subordinate debt that is not sold would lower the market price—as would threats of, or actual, default. Furthermore, changes in bank regulations in creditor countries that reduced the attractiveness to banks of holding on to debt could also increase the willingness to sell to the facility.

XI. Concluding Remarks

From the viewpoint, first, of a debtor country, the availability of an international debt facility cannot be harmful to it because it cannot be compelled to participate and, if use of the facility is associated with conditionality, as is usually proposed, it may choose not to. But, for the reasons given in Section VI, a debtor country is quite likely to gain from participation.

From the viewpoint of creditors, the banks would gain if sales of debt to the facility were truly voluntary, if there were no subordination of debt that is not sold to the facility, and if the moral hazard problems discussed in Section VI were overcome. Otherwise the banks could lose. It has been noted that the moral hazard problem might be overcome by determining purchase prices of debt on some objective basis or on the basis of market prices at an early "cut-off" point. But this can present some of the most difficult practical problems involved with the establishment of a debt facility.

Finally, and most crucial, there are the interests of the potential owners or underwriters of the facility to consider. If the facility would purchase a significant part of the commercial debt of all the developing countries that currently have problems—as is suggested in many of the proposals—a large transfer of risk internationally from private banks to the underwriting governments or multilateral institutions would take place. The extent of the transfers would depend on the detailed arrangements that have been discussed here. Of course a facility could operate on a small scale, but then it would only make a small impact on the world debt situation.

The potential owners may see some benefit in increasing certainty (which might be brought about by the debt relief process combined with conditionality) and in avoiding a disorderly process of debt restructuring, default, and so on (as discussed in Section IX). Furthermore, the owners may wish to provide aid to particular debtors or assistance to particular banks, although a generalized facility is not the best way of doing this. But it is inevitable that, by underwriting the obligations that the facility issues, the owner governments would assume some risk even when the debt not sold to the facility is subordinated to that held by the facility, and even more so without subordination.

REFERENCES

Dooley, Michael P., "Buy-Backs and Market Valuation of External Debt," *Staff Papers,* International Monetary Fund (Washington), Vol. 35 (June 1988), pp. 215–29.

Feldstein, Martin, and others, "Restoring Growth in the Debt-Laden Third World: A Task Force Report to the Trilateral Commission," Triangle Papers, Report 33 (New York: The Trilateral Commission, April 1987).

Krugman, Paul R., "International Debt Strategies in an Uncertain World," in *International Debt and the Developing Countries*, ed. by Gordon Smith and John Cuddington (Washington: World Bank, 1985), pp. 79–100.

Sachs, Jeffrey, and Harry Huizinga, "U.S. Commercial Banks and the Developing-Country Crisis," *Brookings Papers on Economic Activity: 2* (1987), The Brookings Institution (Washington), pp. 555–601.

Capital Flight

A Response to Differences in Financial Risks

MICHAEL P. DOOLEY*

Empirical tests incorporating measures of capital flight from developing countries that are substantially different from those used in existing studies suggest that capital flight can be explained by differences in risk perceived by residents and nonresidents in holding claims on residents of the countries studied. To the extent that capital flight reflects differences among holders in expected yields on claims on residents of capital-flight countries, it may not be related to conventional determinants of net capital movements such as yield differentials between countries.

FOR THE PURPOSES of this paper, flight capital is defined as the stock of claims on nonresidents that do not generate investment income receipts in the creditor country's balance of payments data. This definition permits an empirical distinction between capital outflows motivated by normal portfolio decisions and those based on the desire to place assets beyond the control of domestic authorities.

Empirical tests incorporating this measure suggest that capital flight can be explained by differences between what residents and nonresidents perceive to be the risk-adjusted returns from holding claims on residents of the countries studied. The tests also suggest that attempts to arbitrage these different risk-adjusted returns may lead to increases in stocks of gross external debt. To the extent that capital flight reflects differences between the rates of return that residents and nonresidents can expect to earn on their claims on residents of the countries studied, capital flight may not be related to international yield differentials be-

* Mr. Dooley, Chief of the External Adjustment Division in the Research Department, is a graduate of Duquesne University, the University of Delaware, and the Pennsylvania State University.

tween these countries and the rest of the world or to other determinants of net international capital movements.

The plan of study of the remainder of the paper is as follows. Section I develops the analysis of the aggregate gross external asset position of the seven countries studied.[1] It is shown that this position is large relative to the stock of external debt. Measured investment income receipts, however, have been well below the level that would be consistent with market interest rates. The below-market yields measured on external claims are interpreted as the result of capital flight. That is, since residents have chosen to acquire financial assets that are outside the control of domestic authorities, it is reasonable to assume that the associated investment income is not reported in balance of payments data.

It follows that measured yields on external claims that are below risk-free market rates reflect many of the same considerations as risk premiums on external debt. The latter reflect nonresident creditors' assessment of risks associated with claims on residents of particular countries. The former reflect residents' assessment of risks associated with domestic financial assets. Thus, a country's *net* investment income payments to nonresidents would be quickly affected by a change in the attitudes of residents and nonresidents toward the country's prospects, particularly in cases where gross assets and liabilities are large relative to the net position. For example, in cases where risk premiums on liabilities have been large and variable, a successful adjustment effort can lead to a substantial reduction in such premiums and to a significant fall in investment income payments as debts are renegotiated. Moreover, in cases where a significant share of external claims reflects capital flight, an improvement in the outlook can be expected to lead to an increase in the realized yield on external assets and to an increase in investment income receipts.

The analysis of yields on external claims suggests a definition of capital flight that is somewhat different from alternatives suggested in previous studies.[2] In Section II, capital flight is defined as that part of the estimated stock of external claims that yields no recorded investment income to the creditor country. It is shown that this alternative measure supports the view that capital flight has been an important aspect of recent economic developments. Section III reports the results of empirical tests of the hypothesis that capital flight can be explained by variables that influence residents' and nonresidents' perceptions of risk asso-

[1] Argentina, Brazil, Chile, Mexico, Peru, the Philippines, and Venezuela.

[2] See Cumby and Levich (1987) for a discussion of alternative definitions of capital flight. See also Lessard and Williamson (1987) for case studies and further discussions of these issues.

ciated with claims on the subject country. It is also suggested that, although capital flight may have been related to increases in external debt in recent years, there is no necessary relationship between these two phenomena. For example, in circumstances in which residents face incentives to acquire claims on nonresidents but opportunities for external borrowing are limited, capital flight may still be financed through current account transactions. Section IV provides some conclusions and possible extensions of the paper.

I. Stocks of External Claims and Rates of Return

In this section estimates of stocks of gross claims on nonresidents and yields on these claims are presented and analyzed. The major problem in estimating stocks of external claims held by residents of the countries studied is that balance-sheet data for such claims may be much less reliable compared with similar data for external debts.[3] Indeed, to the extent that claims result from capital flight, these positions might be difficult to identify. For this reason, aggregate data for the countries studied are presented.

Column 1 of Table 1 shows the aggregate stock of private and official claims on nonresidents calculated from cumulated balance of payments data. The initial value was estimated by capitalizing investment income receipts in the initial year. Column 2 shows the aggregate stock of errors and omissions in the cumulated balance of payments accounts. This item is often associated with accumulations of financial claims on nonresidents in that it could reflect, among many other things, unrecorded capital outflows.[4] The sum of columns 1 and 2, shown in column 3, can be taken as an aggregate cumulative capital outflow from the seven countries considered, or an increase in gross claims, as estimated from balance of payments data.

These data, however, could seriously underestimate the stock of external claims. Balance of payments data seem to underestimate the aggregate accumulation of external debt. For the countries studied, external debt in 1984 as estimated from balance of payments data was only 60 percent as large as external debt as estimated by World Bank data.[5] If it is assumed that the World Bank data are accurate, there must have been

[3]For a recent effort to provide balance-sheet data on claims for groups of developing countries, see Williamson (1986).

[4]The sign of this item is reversed compared with balance of payments data; a positive value therefore indicates an increase in claims. See also Appendix I.

[5]See Dooley (1986) for further data and analysis.

Table 1. *Alternative Measures of Aggregate External Claims, 1977–84*
(In billions of U.S. dollars)

Year	Recorded Claims on Nonresidents Other Than Direct Investment, Cumulated Balance of Payments (1)	Cumulated Balance of Payments Errors and Omissions (2)	Total External Claims, Cumulated Balance of Payments (1) + (2) (3)	Unrecorded Stock of External Claims (4)	Total Stock of External Claims (3) + (4) (5)
1977	37.2	2.4	39.6	24.2	63.8
1978	46.8	0.7	47.5	32.7	80.1
1979	58.0	−2.3	55.8	43.4	99.1
1980	63.4	3.3	66.7	56.0	122.8
1981	71.5	14.1	85.7	56.4	142.1
1982	62.7	23.1	85.8	60.2	146.0
1983	68.1	25.1	93.2	66.9	160.1
1984	88.0	23.7	111.7	71.1	182.8

Note: See Appendix I for definitions and sources.

a corresponding underestimate of "balancing transactions." These could include any type of transaction with a nonresident. For example, imports of goods and services or purchases of financial claims on nonresidents could have been financed by the accumulation of external debt not captured in the balance of payments reporting system.[6] In order to proceed, it is assumed that all of the balancing transactions shown in column 4 represent increases in private financial claims on nonresidents. The sum of columns 3 and 4, shown in column 5, is an estimate of the total cumulated *stock* of financial claims on nonresidents for the countries considered.

With this in mind, it is natural to compare investment income receipts as recorded in the balance of payments with alternative measures of the stock of claims on nonresidents. Columns 1, 2, and 3 of Table 2 show the weighted-average ratio of investment income receipts to the three measures of gross external claims presented in columns 1, 3, and 5 of Table 1. These implicit yields can be compared with the weighted-average "market yield" shown in column 4 of Table 2.[7] Investment income re-

[6]For example, Gulati (1987) argues that systematic errors in international trade data may have generated a positive bias in the estimates of capital flight presented in this paper.

[7]Because relatively little information exists concerning the composition of external claims, the market yield reported here is equal to the yield on external liabilities to private creditors calculated in Dooley (1986, Appendix D).

Table 2. *Weighted-Average Yields on Alternative Measures of External Claims and a Measure of Capital Flight, 1977–84*

(In percent)

Year	Weighted-Average Yield on Recorded External Claims, Cumulated Balance of Payments (1)	Weighted-Average Yield on Total External Claims, Cumulated Balance of Payments (2)	Weighted-Average Yield on Total Stock of External Claims (3)	Weighted-Average Market Yield on External Claims (4)
1977	4.3	4.0	2.5	6.6
1978	5.7	5.6	3.3	8.2
1979	7.3	7.6	4.3	10.6
1980	10.3	9.7	5.3	12.7
1981	11.6	9.7	5.9	14.4
1982	10.5	7.7	4.5	12.0
1983	6.8	4.9	2.9	9.7
1984	7.4	5.9	3.6	10.7

Note: See Appendix II for definitions and sources.

ceipts seem to imply plausible yields relative to market yields if it is assumed that only cumulated recorded external claims yield income that is captured in the balance of payments accounts. In Table 2, the addition of cumulated errors and omissions to the level of external claims (column 2), or this plus unrecorded claims (column 3), reduces the implied yield to implausibly low levels.

These very low calculated yields on the more inclusive measures of external claims raise two interesting possibilities. One is that the data on investment income receipts are systematically understated. In some cases (the United States is an example), balance of payments data on such receipts are not measured directly but instead are estimated by applying a market yield to an estimated stock of gross claims. When this procedure is followed, the use of cumulated balance of payments statistics to estimate the stock of claims could understate the value of receipts. It would seem reasonable in this event to revise the balance of payments statistics, particularly in cases where liabilities reported by nonresidents to the subject country are substantial. Indeed, a satisfactory resolution of important problems in evaluating the role of external claims would seem to require an effort to gather data on the stock of external claims held by residents of developing countries comparable to the data available on the external debt of these countries. In cases where the under-

stating of investment income receipts is due simply to an inappropriate estimate of the stock of external claims, a correction of this error would result in a substantial "improvement" in the current accounts of some of the countries studied.[8]

The second possibility is that unreported earnings on external claims reflect residents' preferences for assets that are beyond the control of domestic authorities. In this case repatriation of external claims, or of the earnings thereon, might depend on a reduction of expected taxation or on other penalties on residents' financial wealth. It is worth noting, however, that additional steps may be necessary in cases where domestic financial markets are relatively underdeveloped and not fully recovered from the difficulties experienced in recent years.

II. Capital Flight

There is no widely accepted definition of capital flight. Capital outflows per se, types of outflows such as increases in private short-term claims, or errors and omissions are not a reasonable measure of this phenomenon. The United States, for example, by the end of 1983 had accumulated roughly US$700 billion in gross claims on nonresidents. These claims were offset by a roughly equivalent stock of external debt, and, because the average yield on claims is estimated to be higher than that on liabilities, the investment income accounts showed a substantial net credit in the U.S. balance of payments accounts.[9] The situation would of course be immediately altered if, for some reason, these foreign exchange earnings were not available to U.S. residents who had to make investment income payments.[10] This possibility suggests a useful defini-

[8] Such a procedure could be misleading, since the debtor country's government cannot use such earnings to make payments on external debt unless the income is repatriated. Nevertheless, it seems inappropriate to exclude these private external claims from balance of payments estimates. Retained earnings on foreign direct investment, for example, are recorded as investment income regardless of their effect on government revenues.

[9] See Dooley (1980–81) for a discussion of some of the factors that have generated gross capital inflows and capital outflows in the U.S. balance of payments in recent years.

[10] Diaz Alejandro (1984), for example, has argued that this problem occurs in cases where private external debt has been socialized, or payments subsidized, by the government but where private external assets and earnings remain private. Khan and Haque (1985) have provided a formal analysis of circumstances in which an individual will simultaneously acquire external debt and invest at home as well as abroad. Ize and Ortiz (1987) emphasized the fiscal implications of capital flight.

tion of capital flight. In cases where all or some portion of the income receipts are lost to the country, it is assumed that the capital outflow was motivated by the desire of residents to obtain financial assets, and earnings on those assets, that remain outside the control of the domestic authorities. Capital flight, by this definition, does not require a change in the stock of total claims on nonresidents but only that earnings on existing claims be placed outside the control of the domestic authorities. Thus, capital flight can occur, and be reversed, quite rapidly.[11]

The appropriate way to measure capital flight by this definition is to calculate the stock of claims implied by investment income receipts and market interest rates. This stock of "interest-earning" claims is interpreted as the stock of external claims not attributable to capital flight but attributable to normal foreign portfolio investment. The difference between this estimate and the estimate of the total stock of external claims (for the latter, see Table 1, column 5) is presented in column 1 of Table 3. It should be emphasized that whether such holdings are "beyond the reach" of the authorities is a matter of fact that cannot be established here. Thus, the estimates of aggregate capital flight shown in Table 3 are, at best, approximations.[12] One can gain some perspective on the possible importance of capital flight by comparing capital flight with other financial aggregates. By these estimates, the stock of flight capital accounted for about two thirds of the total stock of external claims of the seven countries studied and was equal to about one third of their external debt.

III. Empirical Tests

This section presents empirical tests of the view that capital flight occurs in circumstances in which residents and nonresidents have perceived different risks associated with domestic investments. The existing empirical literature on risk premiums and on capital flight emphasizes the idea that residents and nonresidents do face different risks. For

[11] For this reason the timing and amount of capital flight, as measured in this paper, can be quite different compared with the estimates of other studies. In most cases the measures of capital flight reported in this paper are much larger than those reported elsewhere. See, for example, Cuddington (1986) and Dornbusch (1985).

[12] For example, this measure could provide a misleading pattern of capital flight to the extent that investment income receipts recorded in the balance of payments reflect repatriated earnings rather than reported (or estimated) accrued earnings. In such cases analysis of year-to-year changes in this measure of capital flight would be less useful, although the trend over longer periods might still provide useful information.

Table 3. *Aggregate Capital Flight in Ratio to External Claims and External Liabilities, 1977–84*

Year	Aggregate Calculated Capital Flight (1)	Ratio of Aggregate Capital Flight to External Claims (2)	Ratio of Aggregate Capital Flight to External Debt (3)
1977	38.90	0.61	0.33
1978	48.18	0.60	0.33
1979	59.03	0.60	0.33
1980	71.63	0.58	0.32
1981	84.53	0.59	0.31
1982	91.62	0.63	0.30
1983	112.19	0.70	0.35
1984	121.19	0.66	0.36

Note: See Appendix III for definitions and sources.

example, of the nine independent variables in Edwards's (1984) paper on risk premiums and of the three independent variables in Cuddington's (1986) paper on capital flight, only one, the rate of domestic inflation, appears in both studies. The different regression hypotheses in these studies reflect the view that capital flight is motivated primarily by inflation or exchange rate risk, whereas risk premiums charged by nonresidents reflect mainly default risk.

The different regression hypotheses can be explained by the assumption that residents of a debtor country do not fear default in the sense that financial contracts between residents (including the government) will be enforced in cases where nonresident creditors are penalized. Residents are exposed to taxation of income derived from domestic investments, and in many cases domestic instruments may offer incomplete protection from domestic inflation and exchange rate depreciation.

In contrast, nonresident investors have typically acquired claims on debtor countries that are denominated in foreign currencies and therefore are not subject to domestic inflation and exchange rate depreciation. For this reason evaluations of risks faced by nonresident investors have focused on comparisons between the country's ability to generate the foreign exchange necessary to service its external debt and the stock of that debt.

In this section the conjecture is tested that both inflation and other taxes on residents and the default "tax" must be considered together to obtain a satisfactory explanation of capital flight. Governments faced

with a fiscal deficit may be forced to impose taxes on domestic and nonresident investors in a variety of forms that tend to reduce the value of such investments. Thus, any shock to the system that increases government expenditure or reduces revenues raises the possibility that new sources of revenue will be exploited. This possibility does not imply that all financial positions will be affected equally or at the same time. The revenue generated by a change in the tax on a financial position depends on the size of the position to be taxed and on the opportunities that asset holders have to avoid the tax. An unexpected increase in the inflation rate, for example, provides revenue to the government that is dependent on the stock of fixed-interest domestic currency liabilities of the government and on the ability of holders of such liabilities to shift into alternative assets to avoid the tax. In cases where the ability to shift into a less exposed position is limited—for example, by a limitation on the right to purchase foreign exchange at the official rate—the government might be expected first to "tax" domestic residents through money creation and inflation. In response, residents will seek to acquire financial assets denominated in foreign currency. If these are not available domestically, residents will attempt to acquire claims on nonresidents denominated in foreign currency.

The incipient capital outflow must be matched either by some offsetting transaction or by a change in other variables that discourages the outflow. One possible outcome is that nonresident investors will be attracted by the fall in prices (increases in yields) that will be generated as residents attempt to liquidate domestic securities. Nonresident investors, however, will not be interested in liabilities of the debtor country that are denominated in domestic currency because such investment also exposes them to the inflation tax. Unlike residents, nonresident investors may be able to purchase claims on residents denominated in foreign currency. Moreover, external creditors may have access to explicit or implicit government guarantees not available to residents.[13] In this case the capital outflow to avoid the inflation tax would be matched by a capital inflow that is protected from the inflation tax because of its currency denomination and that also enjoys a government guarantee.

The capital flows generated in the above example can best be understood as attempts by asset holders to arbitrage a yield differential that is generated by the inflation tax on residents. Because nonresidents can avoid this tax in ways not available to residents, capital flight is the expected outcome of this incentive structure. Similar types of capital

[13]This source of capital flight is emphasized by Eaton (1987).

flows among industrial countries have been called "round-trip" capital flows (see Dooley (1980–81)). In some cases fairly small tax incentives in terms of differential reserve requirements on bank deposits have led to very large gross capital inflows and outflows in the U.S. balance of payments. Such flows are not referred to as capital flight in this context, although they are analytically equivalent to what is called capital flight in developing countries.

There are limits to round-trip capital flows. At some point the stock of liabilities to nonresident investors becomes sufficiently large to make it worthwhile for the government to consider a tax on these financial positions. The inflation tax will not be effective, but other more direct measures that reduce payments on such liabilities will yield greater revenues as the stock of liabilities to nonresidents grows. As the perceived risk faced by nonresident creditors grows, the opportunities for arbitraging the differential risks faced initially by residents compared with nonresidents will be reduced and may eventually be eliminated. At this point the round-trip capital flows will end.

In this new situation resident investors could avoid domestic taxes on financial assets only by net transfers of goods and services to nonresidents. Thus, although capital flight as measured in this paper will be much reduced or eliminated, the consequences of continued actual or expected taxation of residents' financial assets remain serious. Indeed, the consequences will be more serious because the flow of real saving available for increases in the domestic stock of real capital is threatened, since taxes on domestic financial saving can no longer be avoided by round-trip capital flows.

The discussion above suggests that capital flight should be evident in circumstances in which residents perceive risks to income derived from domestic claims, but in which nonresidents have perceived relatively smaller risks on credits to the residents of the country studied. To evaluate this idea, the relationships between capital flight defined above, and several possible determinants of capital flight, are tested empirically. The variables considered that might be determinants of capital flight include the following.

- Domestic inflation (*INF*) is interpreted as measuring the extent to which the government has resorted to taxing domestic financial assets through money creation. This easily observed measure of "tax" policy can be considered a proxy for the difficulty the government is experiencing in generating revenue. Thus, even domestic investments that are indexed to the inflation rate may be considered at risk when the government is using the inflation tax.

• Financial repression (*FR*) measures the difference between interest rates paid on short-term assets denominated in U.S. dollars and time deposits denominated in the subject country's domestic currency and adjusted for actual exchange rate changes. A positive value for this differential indicates that such deposits are vulnerable to the inflation tax. A differential in favor of dollar-denominated assets would suggest more capital flight.

• The risk premium (*RP*) is the risk premium on external debt reported in Dooley (1986) and interpreted as measuring nonresidents' perception of the risk of being "taxed" by the subject country's government. As this perceived risk increases, it is expected that capital flight will be reduced because the differential risk faced by resident and nonresident investors is reduced.

As discussed above, these are clearly endogenous variables that reflect basic changes in the environment faced by governments of debtor countries. For this reason, instrumental variables were used in the regressions reported below. The pressure for domestic inflation is assumed to be related to the fiscal deficit of the debtor country, expressed as a share of gross national product. The default risk faced by nonresidents is assumed to be related to the ratio of debt to exports and to the ratio of interest payments to exports.

The results of a pooled regression for annual observations for six countries for the 1976–83 time period are shown in Table 4. In general, these results support the view that capital flight is related to the relative perceptions of risks that residents and nonresidents associate with claims on residents of the countries studied.

The positive coefficient for domestic inflation (*INF*) in the estimation results in Table 4 can be interpreted as indicating that, where the authorities are "taxing" domestic money balances through inflation, residents prefer to hold a larger share of their financial assets in a form that is outside the control of the domestic authorities. The coefficient on the political risk premium (*RP*) indicates the expected negative impact on capital flight. Thus it appears that the increases in this premium in recent years, and the associated reluctance of nonresident investors to "recycle" flight capital, have tended to limit the scale of such transactions. Finally, the positive relationship between financial repression (*FR*) and capital flight suggests that residents acquire foreign assets in cases where the yield on domestic currency time deposits is less than short-term rates in international markets (adjusted for exchange rate changes). In some cases this may reflect controls on domestic interest rates, or it may simply reflect relatively slow adjustment of domestic interest rates to rapidly changing economic conditions.

Table 4. *Determinants of Capital Flight: Pooled Data, 1976–83*

Symbol	Definition and Source
	Variables
CF	Capital flight, from data reported in Table 3
INF	Domestic inflation, calculated as differences in logarithms of consumer price indices, from International Monetary Fund, *International Financial Statistics (IFS)* (Washington, various issues), line 64
FR	Financial repression, $FR = [(1 + r_{US})/(1 + r)]/[1 + \ln X - \ln X(-1)]$, where r is the domestic time deposit interest rate (national sources), r_{US} is the U.S. Treasury bill interest rate (*IFS* line 60c), and X is domestic currency per dollar (*IFS* line ac)
RP	Risk premium r, from data reported in Dooley (1986)
	Instruments
Fiscal deficit/gross national product (GNP)	Fiscal deficit from *IFS* line 80zf and Fund staff estimates; GNP from *IFS* line 99a
Debt/GNP	Debt from World Bank, *World Debt Tables* (Washington, various issues)
Interest payments/GNP	London interbank offered rate (LIBOR) and *IFS* line 60eb, times Debt/GNP
	Estimation results[a]
CF	$= 23.10\,INF + 18.91\,FR - 1.12\,RP$ (4.34) (4.09) (−1.66)
$\bar{R}^2$	$= 0.83$

[a] "Estimation results" were obtained by instrumental variables and ordinary least squares with country-specific dummy variables. Annual observations for six countries (Argentina, Brazil, Chile, Mexico, the Philippines, and Venezuela) were pooled for the regression; Peru was excluded because data on deposit rates were unavailable. Numbers in parentheses are t-statistics; $\bar{R}^2$ is the coefficient of determination, corrected for degrees of freedom.

IV. Conclusions

An important challenge for analyses of external debt is to identify factors that are quantitatively important in shaping individual countries' current and prospective external positions. Existing studies have in general compared a very simple representation of a country's external finan-

cial position—the ratio of interest payments or debt service payments to some measure of the country's ability to make such payments—under alternative assumptions about economic growth in Organization for Economic Cooperation and Development countries, interest rates in creditor countries, and several other domestic and external variables. The analysis developed in this paper suggests that in many cases a more complete representation of individual debtor countries' financial positions would be useful.

Review of historical experience for selected debtor countries suggests that different incentives faced by nonresident and resident investors seem to be an important determinant of gross external claims and liabilities of the countries studied. Moreover, the size of such gross positions may have little to do with the determinants of net external debt of the country studied. Finally, because investment income payments associated with external debt and investment income receipts associated with external claims appear to have been quite sensitive to residents' and nonresidents' attitudes toward claims on debtor countries, a better understanding of the behavior that determines such positions might be an important element in evaluating and improving a country's external position.

APPENDIX I

Definitions and Sources, Table 1

These series are aggregates derived from the seven individual countries studied: Argentina, Brazil, Chile, Mexico, Peru, Philippines, and Venezuela.

Column 1, "Recorded Claims on Nonresidents Other Than Direct Investment, Cumulated Balance of Payments," is the cumulated sum of *BOPY* (International Monetary Fund, *Balance of Payments Statistics Yearbook,* Washington, various issues) lines 62–64, 69–71, 77–79, 84, 85, 89, 93, 94, and 98–109 plus the estimated value at the end of the first year shown. The estimated value is equal to the value of nondirect investment income receipts in that year, *BOPY* lines 15, 17, and 19 divided by line 60c of the International Monetary Fund's *International Financial Statistics* (Washington, various issues).

Column 2, "Cumulated Balance of Payments Errors and Omissions," is the cumulated sum of *BOPY* line 112. Value at the end of the first year shown is zero.

Column 4, "Unrecorded Stock of External Claims," is the difference between debt reported by the World Bank, which includes data for public and publicly guaranteed long-term and short-term debt (source: World Bank, *Statistics on External Indebtedness: External Liabilities of Individual Countries and Territories,* Washington, various issues), and *BOPY*.

APPENDIX II

Definitions and Sources, Table 2

These series are weighted-average yields derived from the seven individual countries studied: Argentina, Brazil, Chile, Mexico, Peru, Philippines, and Venezuela.

Column 1, "Weighted-Average Market Yield on Recorded External Claims, Cumulated Balance of Payments," is the ratio of investment income receipts (*BOPY* lines 15, 17, and 19) to "Recorded Claims on Nonresidents Other Than Direct Investment, Cumulated Balance of Payments" (see Appendix I, note on column 1).

Column 2, "Weighted-Average Yield on Total External Claims, Cumulated Balance of Payments," is the ratio of *BOPY* lines 15, 17, and 19 to "Total External Claims" (Table 1, column 3).

Column 3, "Weighted-Average Yield on Total Stock of External Claims," is the ratio of *BOPY* lines 15, 17, and 19 to "Total Stock of External Claims" (Table 1, column 5).

Column 4, "Weighted-Average Market Yield on External Claims," is from Dooley (1986), Appendix D.

APPENDIX III

Definitions and Sources, Table 3

These series are aggregates derived from the seven individual countries studied: Argentina, Brazil, Chile, Mexico, Peru, Philippines, and Venezuela.

Column 1, "Aggregate Calculated Capital Flight," is the difference between the capitalized value of investment income receipts—that is, investment income receipts (see Appendix II, note on column 1) divided by the "Weighted-Average Market Yield on External Claims" (see Appendix II, note on column 4)—and "Total Stock of External Claims" (see Appendix I, note on column 5).

Column 2, "Ratio of Aggregate Capital Flight to External Claims," is the ratio of column 1, as defined above, and "Total Stock of External Claims" (see Table 1, column 5).

Column 3, "Ratio of Aggregate Capital Flight to External Debt," is the ratio of column 1, as defined above, to debt reported by the World Bank (see Appendix I, note on column 4).

REFERENCES

Cuddington, John T., "Capital Flight: Estimates, Issues, and Explanations," Princeton Studies in International Finance (Princeton, New Jersey: International Finance Section, Department of Economics, Princeton University, 1986).

Cumby, Robert, and Richard Levich, "On the Definition and Magnitude of Recent Capital Flight," in Donald R. Lessard and John Williamson, eds., *Capital Flight and Third World Debt* (Washington: Institute for International Economics, 1987).

Diaz Alejandro, Carlos F., "Latin American Debt: I Don't Think We Are in Kansas Anymore," *Brookings Papers on Economic Activity: 2* (1984), The Brookings Institution (Washington), pp. 335–403.

Dooley, Michael P., "The Implication of the Internationalization of Banking for the Definition and Measurement of U.S. Credit and Monetary Aggregates," *Economic Forum* (Salt Lake City, Utah), Vol. 11 (Winter 1980–81), pp. 54–68.

———, "Country-Specific Risk Premiums, Capital Flight, and Net Investment Income Payments in Selected Developing Countries" (unpublished; Washington: International Monetary Fund, 1986).

Dornbusch, Rudiger, "External Debt, Budget Deficits, and Disequilibrium Exchange Rates," in *International Debt and Developing Countries,* ed. by John T. Cuddington and Gordon W. Smith (Washington: World Bank, 1985), pp. 213–35.

Eaton, Jonathan, "Public Debt Guarantees and Private Capital Flight," *World Bank Economic Review* (Washington), Vol. 1 (May 1987), pp. 377–95.

Edwards, Sebastian, "LDC Foreign Borrowing and Default Risk: An Empirical Investigation, 1976–80," *American Economic Review* (Nashville, Tennessee), Vol. 74 (September 1984), pp. 726–34.

Gulati, Sunil K., "A Note on Trade Misinvoicing," in Donald R. Lessard and John Williamson, eds., *Capital Flight and Third World Debt* (Washington: Institute for International Economics, 1987).

Ize, Alain, and Guillermo Ortiz, "Fiscal Rigidities, Public Debt, and Capital Flight," *Staff Papers,* International Monetary Fund (Washington), Vol. 34 (June 1987), pp. 311–32.

Khan, Mohsin, and Nadeem Ul Haque, "Foreign Borrowing and Capital Flight: A Formal Analysis," *Staff Papers,* International Monetary Fund (Washington), Vol. 32 (December 1985), pp. 606–28.

Lessard, Donald R., and John Williamson, eds., *Capital Flight and Third World Debt* (Washington: Institute for International Economics, 1987).

Williamson, Martin, "Acquisitions of Foreign Assets by Developing Countries, 1970–83: Empirical Evidence and Implications for the Debt Crisis" (unpublished; Washington: International Monetary Fund, December 1986).

Dual Exchange Rates in the Presence of Incomplete Market Separation

Long-Run Effectiveness and Policy Implications

DANIEL GROS*

The literature on dual exchange rate regimes assumes that the separation between the two foreign exchange markets is perfect. In this paper a divergence between the two exchange rates induces a flow of arbitrage activity, the magnitude of which depends on both the costs of evading exchange controls and the size of the exchange rate differential. These arbitrage flows lead to a gradual convergence of the two exchange rates. In the long run, therefore, a dual exchange rate regime with a fixed commercial rate imposes the same constraints as a fixed unified exchange rate.

A LARGE NUMBER of countries maintain separate exchange rates for commercial and financial transactions. In these so-called dual exchange rate regimes, the central bank usually intervenes only in the market for commercial transactions to maintain a stable exchange rate for exports and imports. The absence of official intervention in the market for financial transactions implies that the financial exchange rate has to move to ensure capital account equilibrium. Dual exchange rates are therefore often used by countries with balance of payments difficulties as a substitute for direct capital controls.

Dual exchange rate regimes have been discussed extensively in the economic literature, where it has been emphasized that the separation of the goods and financial markets that can be achieved with different

* Mr. Gros, an economist in the Research Department, is a graduate of the University of Chicago. This paper was revised while the author was a Research Fellow at the Centre for European Policy Studies in Brussels. He wishes to thank colleagues in the Fund for helpful discussions and comments.

exchange rates for commercial and financial transactions may be useful in insulating the goods markets of the home economy from the effects of disturbances in the financial markets.[1] For the most part, however, the existing models of the effects of dual exchange rates have not challenged the assumption that the authorities succeed in separating the two markets or, equivalently, that private arbitrage activity has no important consequences.[2] In contrast, this paper develops a model of the scale of private arbitrage activity and focuses on the important consequences of such activity for the viability and effectiveness of dual exchange rate regimes.

The "leakage" from the commercial to the financial exchange market is not just an interesting phenomenon of little practical significance. Bhandari and Decaluwe (1987) refer to the evidence available indicating that for Belgium the estimated trade balance functions show a statistically significant response to the exchange rate differential. Lanyi (1975) also emphasizes the importance of these leakages and describes the instruments used by traders to arbitrage between the two foreign exchange markets.

Taking private arbitrage activity into account leads to the conclusion that dual exchange rates (as well as capital controls) could succeed only temporarily in dampening the effects (on the domestic goods market) of shocks to financial or other markets. To offset such effects permanently, the authorities would have to induce a steadily increasing differential between the two exchange rates. But a steadily increasing differential would also lead to a steadily increasing incentive for private operators to circumvent the regulations that separate the two markets by buying foreign exchange at the lower rate (usually the controlled or commercial rate) and selling it at the higher rate (usually the free or financial rate). Such arbitrage activity would expand as the differential between the two exchange rates widened and would thus limit the size of the differential that the authorities could support.

For the same reason, these considerations suggest that any differential between the two exchange rates would tend to disappear over time unless the authorities were prepared to offset continuously the effects of

[1] See, for example, Dornbusch (1976) and (1985). The insulating effects of dual exchange rate regimes are also discussed in Flood and Marion (1982). Lizondo (1987) discusses the balance of payments effects of switching from a dual exchange rate regime to a unified floating or crawling peg system.

[2] The only exception seems to be Bhandari and Decaluwe (1987). But many other authors acknowledge that the difference between the two exchange rates should not become large enough to create problems in enforcing the separation between the two markets.

private arbitrage activity by adjusting one of the exchange rates or by using domestic credit policy.[3] Data from Belgium, which has had a dual exchange rate since 1946, confirm that the differential between the commercial and the financial exchange rates has usually been very small (Figure 1). However, substantial differentials have emerged during periods of tension in the financial markets (especially when a readjustment of the parities inside the European Monetary System (EMS) was expected). A similar pattern can be observed in the Mexican two-tier exchange market (Figure 2). The discount of the peso on the free exchange market (relative to the controlled exchange rate) started at nearly 100 percent at the end of 1982 and then declined continuously until June 1985.

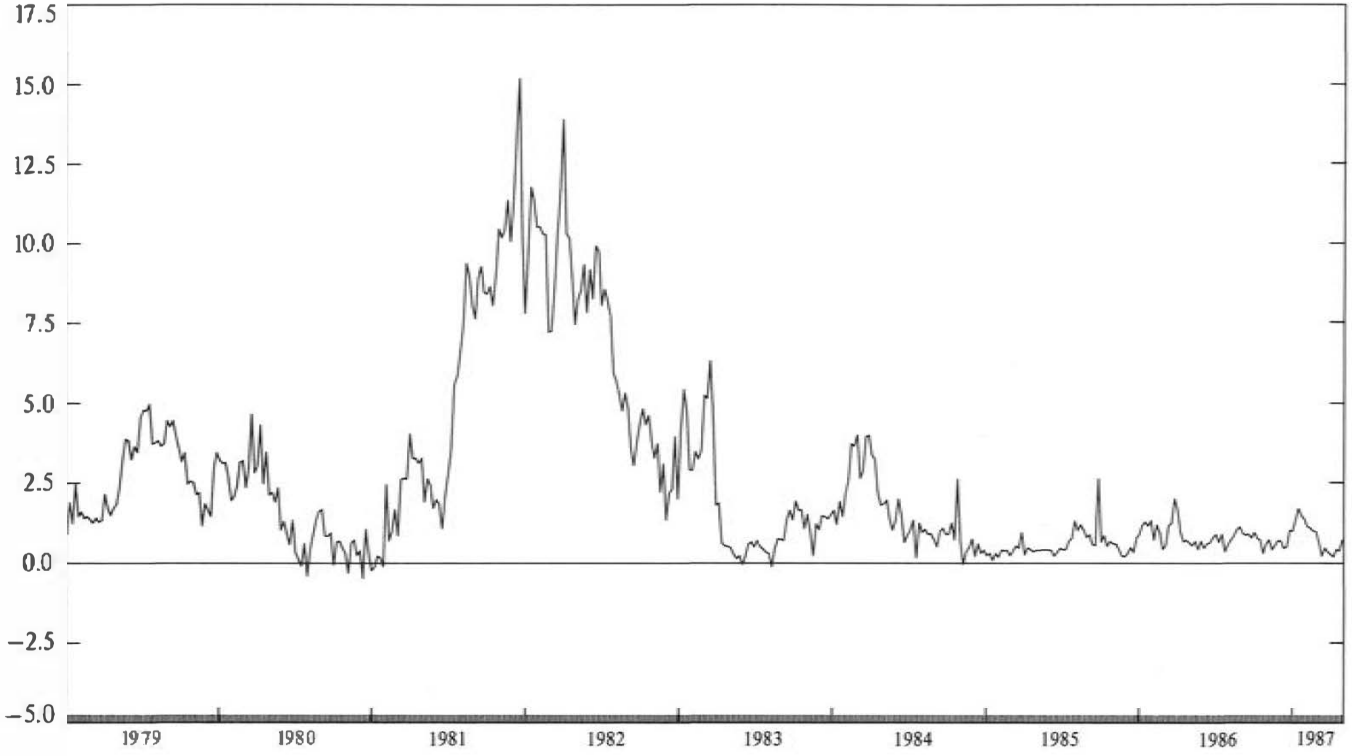
By combining the simple model of arbitrage activity with an equally simple macroeconomic model, this paper also analyzes the consequences of various shocks such as changes in the international interest rate or devaluations of the commercial exchange rate. The analysis suggests that an unanticipated devaluation of the commercial rate would lead to a jump in the financial rate that could either overshoot or undershoot the amount of the devaluation and might thus increase or decrease the differential. The analysis also suggests that the anticipation of a future devaluation of the commercial rate would lead to an immediate depreciation of the financial rate and would thus give rise to a differential between the two rates; this differential would be eliminated, or at least diminished, only when the devaluation of the commercial rate actually took place. This result seems intuitively plausible because the financial rate is a forward-looking variable.

Another more surprising result is that an increase in the degree of separation of the two markets, as represented by the costs for potential arbitrageurs, would increase the effects of disturbances in the financial markets (for example, changes in the international interest rate) on the financial exchange rate and thus, given the commercial exchange rate, on the exchange rate differential. If the aim of a dual exchange rate regime is to insulate the goods markets from disturbances in the financial markets, this result implies that the price for this insulation might be a higher degree of variability of some shadow prices in the financial markets and an increased deadweight loss because of the increase in private arbitrage and evasion activity.

The long-run ineffectiveness of the controls that have to support the dual exchange rate regime is also illustrated by the discussion of the

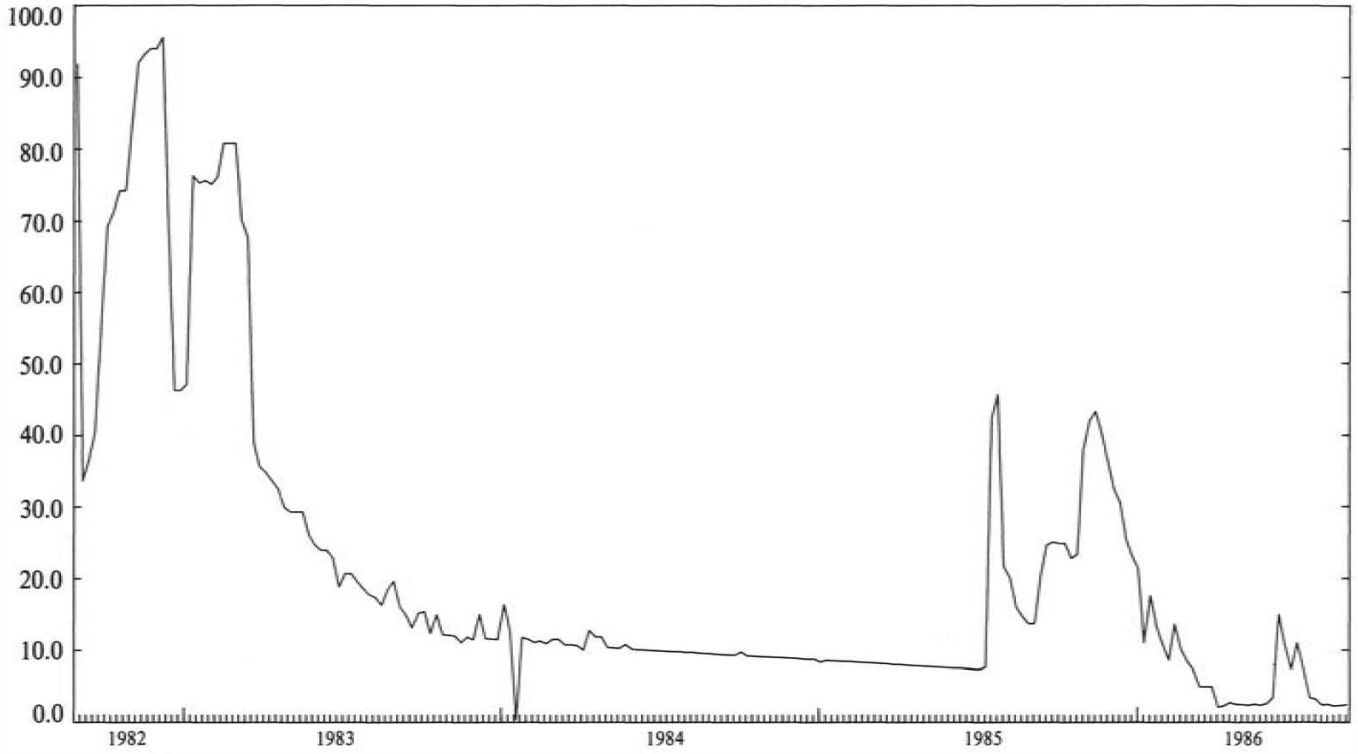
[3] A series of shocks would tend to raise the differential each time a shock hit the system, but after each shock the differential should, on average, decline.

Figure 1. *The Belgian Two-Tier Exchange Market: Percentage Discount on the Financial Franc*



Note: The discount is computed as the natural logarithm of the ratio of the financial exchange rate to the commercial exchange rate.

Figure 2. *The Mexican Two-Tier Exchange Market: Percentage Discount on the Free Peso*



Note: The discount is computed as the natural logarithm of the ratio of the peso price of one U.S. dollar on the free exchange market to the peso price of one U.S. dollar on the controlled market.

effects of monetary policy: with a fixed commercial exchange rate, no independent monetary policy is possible in the long run. Any increase in domestic credit would be offset in the long run by an equal loss of reserves. In terms of the so-called offset coefficient this implies that the impact offset coefficient is different from (negative) unity, but the long-run offset coefficient is always equal to unity.

By considering the relevance of private arbitrage activity, the paper also highlights the importance of two points that are often overlooked in policy discussions about dual exchange rate regimes. The two points can be developed by first considering the proposition that, in the absence of private arbitrage activity, a dual exchange rate regime is operative (in the sense that it induces a different capital and current account than would a unified floating or fixed exchange rate) only if the expected rates of change of the two exchange rates differ. This proposition can be shown to hold in a variety of models and has been emphasized recently by Adams and Greenwood (1985) and Frenkel and Razin (1986). The intuition behind the proposition is that the savings and investment decisions that determine the capital account depend on the expected intertemporal terms of trade, which in turn are a function of the rates of change of the two exchange rates rather than their levels. But the proposition must be modified in the presence of private arbitrage activity, as in the model developed in this paper; in that case a dual exchange rate regime can also affect the equilibrium current and capital accounts when the differential between the two rates is constant because of the arbitrage activity induced by the differential.

A first corollary of this modified proposition, given the tendency of the differential between the two exchange rates to disappear over time, is that a dual exchange rate regime cannot effectively constrain the (cumulative) capital or current account in the long run. Any effect that is obtained when the differential between the financial and the commercial rates is widening would be offset by an opposite effect when the differential is falling back to zero.

A second corollary of the modified proposition, contrary to the arguments of Adams and Greenwood (1985) and Frenkel and Razin (1986), is that in the presence of incomplete market separation no exact equivalence exists between a regime of capital controls that attempts to insulate domestic interest rates from international interest rates and a regime of dual exchange rates. Private arbitrage activity reacts differently to the two regimes. Under capital controls, the incentive for private arbitrage activity exists only if the capital controls are effective in creating a differential between international and domestic interest rates. Under dual exchange rates, however, the incentive for private arbitrage activity

exists even if the differential between the two exchange rates is constant and there is no differential between international and domestic interest rates.

A more technical result of the paper is that taking private arbitrage activity into account serves to determine the *level* of the financial exchange rate even in the context of models in which it would otherwise be indeterminate, such as the basic Dornbusch (1976) model. In most other models of dual exchange rates, the level of the financial exchange rate is determined by considerations of wealth effects. When private arbitrage activity is introduced as a function of the difference between the levels of the financial and the commercial rates, however, the level of the financial rate is determined even in the basic Dornbusch model.

Section I sets up the model that links arbitrage flows to the exchange rate differential, and Section II then incorporates this model of arbitrage flows into a simple general equilibrium model. Section III describes the effects of anticipated (future) and unanticipated devaluations of the commercial rate. It shows that the differential between the financial and the commercial exchange rates could either increase or decrease in response to an unanticipated devaluation, but that an anticipated future devaluation would always have an unambiguous effect on the differential by leading to an immediate devaluation of the financial exchange rate. Section IV describes the effects of anticipated and unanticipated changes in monetary policy. Given the fixed commercial exchange rate, monetary policy consists in changes in the domestic credit of the central bank; accordingly, this section shows that an increase in domestic credit, whether anticipated or not, would lead to an immediate depreciation of the financial rate. The magnitude of this depreciation is an increasing function of the degree of separation of the two exchange markets. Section V contains some concluding remarks.

I. A Model of Arbitrage Activity in Dual Exchange Rate Regimes

Under a dual exchange rate regime, the exchange rate applicable to current account transactions (the commercial rate) may differ from the exchange rate applicable to capital account transactions (the financial rate).[4] The commercial rate is usually fixed by the authorities, whereas the financial rate is sometimes set in a free market and sometimes set by

[4] It is assumed here that interest payments are also converted at the financial exchange rate; thus the commercial exchange rate does not apply to all current account transactions.

the authorities. Because dual exchange rate regimes have mostly been used by countries trying to limit capital outflows, the financial rate, whether determined by a free market or by the authorities, has usually priced foreign exchange at a premium compared with the commercial rate.

Dual exchange rate regimes and capital controls can be regarded as analogous, since under a dual exchange rate the authorities can either set (a path for) the financial rate to achieve a given capital account target or impose controls on capital movements and let the financial rate be determined in the market. Often the implicit target for the capital account (excluding any change in official reserves) is zero; in this case the authorities maintain the commercial rate at a certain level, but the financial rate is determined in a free market without government intervention in that market. This situation occurs in Belgium, where the authorities maintain the commercial franc within the limits imposed by the EMS, whereas the financial franc floats freely.

The existence of a dual exchange rate offers arbitrage opportunities to economic agents in the sense that they would like to buy foreign exchange at the lower rate and sell it at the higher rate. Because the financial rate usually prices foreign currency at a premium compared with the commercial rate, it is henceforth assumed that the arbitrage opportunity consists in buying foreign exchange at the commercial rate and selling it at the financial rate. Given this arbitrage opportunity, all dual exchange rates have to be complemented by a set of rules that define current account (that is, commercial) and capital account (that is, financial) foreign exchange transactions.

One way to circumvent the rules designed to keep the two markets separate would be for an exporter to underinvoice foreign clients and invest the unrecorded payments in foreign financial assets.[5] The proceeds from the sale of these assets could then be repatriated at the financial rate. An importer would correspondingly let himself be overinvoiced and would use the foreign exchange bought at the commercial rate to acquire foreign assets, which could likewise be repatriated at a profit. It is assumed here that importers and exporters incur costs each time they over- or underinvoice because each transaction (export or import) is subject to the control of the enforcement agencies.[6] (Costs

[5] See Lanyi (1975) for a description of the various possibilities for traders to use foreign-trade-related financial transactions to circumvent the controls.

[6] This assumption is in contrast to the model of capital controls developed elsewhere. With capital controls, arbitrageurs incur costs only once, when they shift funds to the international market. Keeping the funds on the international market, where they earn higher interest rates, does not involve any additional costs. With a dual exchange rate regime, however, arbitrageurs have to make the entire round-trip each time to earn the arbitrage profit.

might also be incurred in repatriating the profits at the financial exchange rate.) These costs might take the form of penalties assessed by the enforcement agencies and of side payments that have to be made to foreign suppliers and clients to induce them to collaborate in the over- or underinvoicing. The potential arbitrageur maximizes the difference between the arbitrage profits and his costs, given by $g(y)$:

$$\max\ y_t\left(\frac{E_t}{\overline{C}}-1\right)-g(y_t), \tag{1}$$

where y_t represents the amount of over- or underinvoicing that the arbitrageur undertakes, measured as the amount of domestic currency that the arbitrageur effectively uses to buy foreign exchange at the fixed commercial rate, $\overline{C}$, to earn a profit by selling at the financial rate, E_t. (Both these exchange rates are in terms of domestic currency per unit of foreign currency.) It is apparent that equation (1) is not an intertemporal problem because the arbitrage opportunity, which arises if $(E_t/\overline{C})$ exceeds unity,[7] and the costs occur at the same time. The arbitrageur, therefore, maximizes profits when the marginal cost of increasing the amounts over- or underinvoiced is equal to the proportional discount of the financial exchange rate:

$$\left(\frac{E_t}{\overline{C}}-1\right)=g'(y_t). \tag{2}$$

An interior equilibrium exists only if $g''(y_t)>0$: that is, if $g(\cdot)$ is convex or has increasing marginal costs. The assumption of increasing marginal costs can be defended in two ways. First, in practice dual exchange rate regimes have not collapsed immediately on account of large-scale evasion activity, contrary to what might be expected if evasion activity showed decreasing marginal costs. Second, it seems plausible that in reality importers and exporters would find it increasingly difficult to justify their official prices or terms of payment as they increased the scale of their over- or undervoicing.

These considerations suggest that, as long as the financial rate is at a discount, funds will flow to take advantage of the arbitrage opportunity, but the magnitude of the arbitrage flows will depend on the exact form of the "cheating" function $g(y_t)$. A small positive exchange rate differential might be compatible with no arbitrage flows if the marginal cost of even small amounts of arbitrage is positive; formally, the threshold at which arbitrage flows will stop is given by

$$\left(\frac{E_t}{\overline{C}}-1\right)=g'(0). \tag{3}$$

[7] As mentioned above, only situations in which the financial rate is at a discount compared with the commercial rate are considered here.

In the remainder of the paper, it was convenient to use a specific form of $g(\cdot)$ that yields a linear relationship between the discount and the scale of arbitrage flows. Such a linear relationship results if the arbitrage cost function is quadratic, that is, $g(y_t) = (\phi/2)y_t^2$; in this case the arbitrage flows are given by

$$y_t = \left(\frac{E_t}{\bar{C}} - 1\right) \Big/ \phi. \tag{4}$$

This specific functional form implies that arbitrage flows stop only if the discount disappears. This assumption seems appropriate for a country such as Belgium, which lies in the heart of Europe, is well integrated in international markets (exports are equal to 75 percent of gross national product (GNP)), and thus has considerable scope for arbitrage flows through over- or underinvoicing. The assumption also seems consistent with the fact that in Belgium, during periods of calm in the foreign exchange markets (presumably associated with few or no arbitrage flows), the discount on the financial rate has always been close to zero.

II. The General Equilibrium

This section discusses the general equilibrium consequences of the model of arbitrage activity developed here in the context of a streamlined macroeconomic model. The model is kept as simple as possible to illustrate the main channel through which the arbitrage flows affect conditions in the financial markets.

The model consists of three equations. The first equation describes the trade balance and thus the intervention activity of the government, which purchases any excess supply of foreign exchange resulting from commercial transactions or provides for any excess demand, to keep the (nominal) commercial rate constant. The first important point about the model is that the arbitrage flows induced by the dual exchange rate regime influence the intervention activity of the government because the overinvoicing of imports (underinvoicing of exports) leads to an increased demand for foreign exchange on the commercial market. The "true" trade balance is a function of the (commercial) real exchange rate, denoted by $(\bar{c} - p - \bar{k})$, where $\bar{k}$ denotes the long-run equilibrium real exchange rate. This implies that the flow of intervention activity (in units of domestic currency), denoted by $D(F_t)$, needed to keep the commercial exchange rate at $\bar{c}$ is given by

$$D(F_t) = \beta(\bar{c} - p_t - \bar{k}) - (e_t - \bar{c})/\phi. \tag{5}$$

In this expression, e_t and $\bar{c}$ represent the natural logarithms of the financial and commercial exchange rates, and p_t represents the logarithm of the price level. The second expression on the right-hand side of equation (5) is equivalent to equation (4) for small values of the discount.

The financial market for foreign exchange is not controlled; this implies that risk-neutral arbitrageurs ensure that uncovered interest rate parity holds.[8] The domestic interest rate, i, is therefore equal to the foreign interest rate, denoted by i^*, plus the expected rate of depreciation, $E[D(e_t)]$. Combining this expression with a simple money demand function yields the second equation of the model

$$i_t = E[D(e_t)] + i^* = -\lambda[\ln (F_t + \overline{DC}) - p_t], \tag{6}$$

where F represents the foreign assets of the central bank, $\overline{DC}$ represents domestic credit, $F + \overline{DC} = M$ represents the money stock, and λ is unity divided by the semi-elasticity of money demand. $E[D(e_t)]$ represents the expected rate of change of the financial exchange rate; in this perfect-foresight model no distinction is made between actual and expected values. The notation $E(\cdot)$ is therefore suppressed for the rest of the paper. The model is closed with a conventional price adjustment rule:

$$D(p_t) = \alpha(\bar{c} - p_t - \bar{k}). \tag{7}$$

The simple model consisting of equations (5) through (7) has three dynamic variables: the price level, p_t; the foreign assets of the central bank, F_t; and the financial exchange rate, e_t. These variables are functions of the commercial exchange rate, $\bar{c}$, and the stock of domestic credit, $\overline{DC}$, which represent the policy variables. It is apparent from equations (5) through (7) that in the long run the dual exchange rate regime is equivalent to a fixed unified exchange rate regime if the effects of private arbitrage activity are taken into account. Indeed, at the steady state the fixed commercial exchange rate has to be equal to the financial exchange rate; in addition, the commercial exchange rate determines the price level, the money supply, and hence, given domestic credit, the foreign exchange reserves of the central bank.[9]

It is also apparent from equations (5) and (6) that without the arbitrage flows the level of the financial exchange rate is not determined. Without arbitrage flows (that is, if ϕ is equal to infinity), the model would determine only the rate of change of the financial exchange rate.

[8]See Reding (1985) for some empirical evidence for Belgium that shows that for the financial franc covered interest rate parity holds.

[9]Formally this is implied by the steady-state conditions $p = \bar{c}$, $e = \bar{c}$, and $F = \exp - (i^* + \lambda\bar{c}) - \overline{DC}$.

This result has already been noted in Dornbusch (1976); in general it depends on the absence of wealth effects, which are used in Dornbusch (1976, 1985) and Frenkel and Razin (1986) to determine the level of the financial exchange rate.

The net foreign assets owned by the private sector have not been mentioned so far because they are determined by the exchange rate differential but do not influence any of the endogenous variables. Under the assumption that interest payments have to go through the financial foreign exchange market, the evolution of the net foreign assets of the private sector is given by the condition that the capital account plus interest receipts has to balance. Denoting the net foreign assets of the private sector by A_t, this implies

$$D(A_t) = -i^* A_t - (e_t - \bar{c})/\phi. \tag{8}$$

Because the variable A_t does not enter the system of equations (5)–(7), the net foreign assets are simply determined by the initial value of A_t, and the path of the exchange rate differential is as determined by the system of equations (5)–(7). Note that without leakage—that is, if ϕ is equal to infinity—the net foreign assets of the private sector would have to fall toward zero over time.

The viability of the dual exchange rate regime depends on the objectives of the government. If the government is concerned only with the level of the differential between the financial and commercial rates, it can maintain any constant value of the differential if it is prepared to offset the impact of the arbitrage activity by continuously reducing domestic credit, $\overline{DC}$, at the rate $(e_t - \bar{c})/\phi$. However, the purpose of a dual exchange rate regime is often to evade the consequences of an inflationary policy by keeping domestic interest rates artificially low. This purpose would not be achieved by a policy that keeps the premium constant and thus keeps $D(e_t) = 0$.[10]

Despite the extremely simple macroeconomic relationships used here, the general equilibrium is characterized by a system of three differential equations in the variables F_t, e_t, and p_t. The only other model known to the author that takes illegal leakage into account, Bhandari and Decaluwe (1987), also arrives at a third-order system, in the same variables (the financial exchange rate, the stock of reserves, and the price level). But Bhandari and Decaluwe were unable to obtain an analytical solution because their system is not recursive. They therefore had to limit themselves to numerical examples. In contrast, the present frame-

[10]This result depends on the assumption that interest receipts on foreign assets have to pass through the financial markets.

work can be solved analytically, as can be seen by writing equations (5) through (7) as[11]

$$\begin{bmatrix} \mu & 1/\phi & \beta \\ \lambda/M & \mu & -\lambda \\ 0 & 0 & \mu+\alpha \end{bmatrix} \begin{bmatrix} F_t \\ e_t \\ p_t \end{bmatrix} = 0. \tag{9}$$

The system has three roots, given by

$$\begin{aligned} \mu_1 &= -\alpha \\ \mu_{2,3} &= \pm(\lambda/\phi M)^{1/2}. \end{aligned} \tag{10}$$

The first, stable root ($\mu_1 = -\alpha$) determines the path of the price level, which then acts as an exogenous forcing variable in the remaining dynamic system of F_t and e_t because the complete system of equations (5)–(7) is recursive. The dynamic subsystem consisting of F and e has two roots (μ_2, μ_3) of opposite sign; thus, the subsystem exhibits the usual saddle-path instability. The absolute value of these roots is determined only by the parameters that characterize the financial markets, ϕ and λ. The complete solution of the model that also uses the initial condition is presented in the Appendix.

III. The Effects of Devaluations of the Commercial Exchange Rate

In most dual exchange rate regimes, the path of the nominal commercial exchange rate is fixed by the authorities; in many cases the authorities keep this exchange rate fixed for some time and devalue by discrete amounts whenever it is judged unavoidable because of developments in the trade account. This section, therefore, analyzes the effects of an unanticipated devaluation of the commercial rate, $\bar{c}$, on the exchange rate differential and domestic prices. Moreover, since the behavior of the authorities can often be predicted, this section also analyzes the effects of anticipated future devaluations of the commercial rate.

The reaction of the financial exchange rate to changes in the commercial exchange rate is of particular interest for Belgium because the com-

[11] Changing the assumption that interest receipts on foreign assets have to pass through the financial market would imply for this model that the general equilibrium consists of four differential equations, with the foreign assets of the private sector in addition to the three variables of the system of equations (5)–(7). It would then become impossible to find an analytical solution. The gain in terms of tractability of the analysis might explain why the assumption that interest receipts pass through the financial market is commonly used in the literature. (See, for example, Dornbusch (1976) and Frenkel and Razin (1986).)

mercial franc is fixed within the EMS (neglecting the ±2.25 percent bands) in relation to the deutsche mark and other European currencies. From the beginning of the EMS until September 1986 there were ten realignments, six of which involved changes in the central parity of the Belgian commercial franc in relation to the deutsche mark.[12] Figure 1 shows that around the dates of the realignments, the financial franc was at a discount. The discount on the financial franc usually rises in periods of tension in the foreign exchange markets—that is, before each EMS realignment—and even before those realignments at which the parity of the Belgian commercial franc was not adjusted. During 1981 and 1982 the discount sometimes reached 15 percent and oscillated in general between 7 percent and 10 percent. This pattern in the financial exchange rate of the Belgian franc can be explained in terms of the model presented here, as will be shown in the remainder of this section, which discusses the reaction of the financial exchange rate to unanticipated and future anticipated changes (devaluations) of the commercial rate.[13]

An unanticipated change in $\bar{c}$ of $\Delta\bar{c} = \bar{c}' - \bar{c}$ would lead to a jump in the financial exchange rate, Δe_t, equal to[14]

$$\Delta e_t = \left[\phi \frac{\beta - \alpha M}{1 + (\alpha/\mu)} + 1\right] \Delta\bar{c}. \tag{11}$$

Whether the financial exchange rate would jump by more, less, or the same amount as the commercial exchange rate depends on the relative magnitudes of β and αM (assuming that the economy was initially at its steady-state equilibrium). The financial rate would jump by the same amount as $\bar{c}$ to a new steady state, and the money market would remain in equilibrium (after the change in $\bar{c}$), with the financial rate constant (that is, with $D(e_t) = 0$) only if the devaluation of the commercial rate

[12]The dates and the rates of devaluation of the Belgian franc against the deutsche mark (as measured by changes in central bilateral parity between the deutsche mark and the Belgian franc) were: September 24, 1979 (−2 percent); October 5, 1981 (−5.5 percent); February 22, 1982 (−8.5 percent); June 14, 1982 (−4.25 percent); March 21, 1983 (−4 percent); and April 7, 1986 (2 percent). No changes occurred in the central parity between the deutsche mark and the Belgian franc during the realignments of November 30, 1979, March 23, 1981, July 22, 1985, and August 4, 1986.

[13] In Belgium, a certain class of exporters has the option of using either the financial or the commercial exchange rate. This means that the financial rate can never be lower than the commercial rate because the arbitrage costs apply only to outflows. This restriction is at variance with the model developed above, which implicitly assumes that arbitrage costs apply symmetrically to outflows and to inflows. The consequences of the asymmetry in the Belgian system will be taken into account whenever they become relevant in the remainder of the paper.

[14] See the Appendix, equations (18)–(24), for derivations.

had no effect on real balances. This result could occur only if the effects of the rising price level on the demand for money were exactly offset by the increased inflows through the trade account. (Note that αM describes the effect of the rising level of the logarithm of prices on the real demand for money (not expressed as a logarithm), and β describes the effect of the devaluation on trade flows.) In this special case, the potential for arbitrage activity has no impact because no difference emerges between the financial and the commercial rates.

By contrast, if the trade account reacts strongly to a devaluation of the commercial rate (that is, if β is relatively high), as would be expected in Belgium where the proportion of tradables in GNP is very high, the financial exchange rate would overshoot its long-run equilibrium. In this case arbitrage activity would emerge and would influence both the initial jump of the financial rate and its entire future path.[15] Because the (absolute value of the) root is a decreasing function of the parameter ϕ that describes the severity of the controls used by the authorities to separate the two exchange rate markets, it is not apparent from equation (11) whether an increase in the severity of the controls would tend to reduce or increase the over- or undershooting of the financial exchange rate. However, equations (27) and (28) in the Appendix prove that an increase in the severity of the controls would tend to increase the observed differential between the financial and the commercial exchange rates after unanticipated devaluations.

It has often been argued that if the financial exchange rate is at a discount, the authorities should devalue the commercial rate to eliminate the differential and thus the incentive for arbitrage activity. However, the above results imply that sometimes an unexpected devaluation might increase the differential between the financial and commercial rates.

In equations (5) and (7) the equilibrium real exchange rate, $\bar{k}$, has the same effect (but with the opposite sign) on inflation and the trade balance as does the commercial exchange rate, $\bar{c}$. This result implies that a change in the equilibrium real exchange rate of $\Delta\bar{k} > 0$ (which implies a long-run real depreciation of the domestic currency) would have similar effects on the financial exchange rate as a devaluation of the commercial exchange rate ($\Delta\bar{c} > 0$). The Appendix shows that this result is indeed true; the initial jump in the financial exchange rate, Δe_t, in response

[15] For Belgium, in the case of $\alpha M > \beta$, an undershooting would not be possible because the financial rate cannot be lower than the commercial rate. Instead there would be instantaneous inflows of money that would force the financial rate to be equal to the commercial rate.

to an unanticipated depreciation of the equilibrium real exchange rate is equal to

$$\Delta e_t = -\phi \left[\frac{\beta - \alpha M}{1 + (\alpha/\mu)} \right] \Delta \bar{k}. \tag{12}$$

If equations (11) and (12) are compared, they show that the same parameters that determine whether the financial exchange rate over- or undershoots in response to an unanticipated devaluation also determine whether the financial exchange rate appreciates or depreciates in response to an unanticipated depreciation of the equilibrium real exchange rate. This result implies that if $\beta > \alpha M$, an unanticipated depreciation of the equilibrium real exchange rate would lead to an overshooting of the financial exchange rate (that is, an immediate discrete depreciation followed by a continuous appreciation that leads the financial exchange rate back to its original level). The similarity between equations (11) and (12) also implies that an increase in the severity of the controls would tend to increase the magnitude of the over- or undershooting.

Dual exchange rate regimes are supposed to isolate the real sector from the effects of disturbances in the financial sector. It is interesting to note that this result implies that a disturbance in the real sector (that is, a change in the equilibrium real exchange rate) has effects on the financial sector (that is, on the financial exchange rate and therefore on interest rates).

In contrast to the effects of a current unanticipated devaluation of the commercial rate, it is clear that a future anticipated devaluation has to raise the discount on the financial rate at the time the market begins to expect the future devaluation of the commercial rate. By how much the financial rate has to jump at the time the news about the future devaluation reaches the market is determined by the condition that no jump in the financial rate should occur at the time the commercial rate is actually devalued (provided that the amount and the timing of the devaluation were correctly anticipated by the market). However, the differential that arises as the financial rate depreciates tends to lead to a loss of reserves because of the arbitrage flows. This loss of reserves implies (see equation (6)) that the expected rate of devaluation has to increase, and the financial exchange rate will therefore continue to depreciate increasingly until the devaluation of the commercial rate takes place. At this time, the arbitrage flows should stop, or may even be reversed, depending on how far ahead the devaluation was anticipated and on the value of $(\beta - \alpha M)$. In Belgium, the differential cannot be negative, and the initial jump of the financial rate is therefore determined also by the condition that the financial rate is expected to be at least equal to the

new (devalued) commercial rate. But if β is so much larger than αM that an unanticipated devaluation would result in overshooting, the financial rate might remain at a discount even compared with the new commercial rate after the devaluation. In this case the differential would only be reduced by the impact of the devaluation of the commercial rate, and the differential would disappear only gradually over time.

The results of this section may be compared with those obtained in the usual no-leakage models of the literature on dual exchange rate regimes. In these models (see, for example, Lizondo (1984)), a steady-state exchange rate differential usually exists that is different from zero and that is determined by wealth effects for the reasons mentioned above. The steady-state exchange rate differential is therefore not affected by the level of the commercial rate. In Lizondo (1984), a devaluation of the commercial rate will therefore initially reduce the exchange rate differential (a result that is possible but not guaranteed in the present model), but the financial rate has to continue to depreciate until the former exchange rate differential is re-established. The results described in this section are therefore similar to the results of all those models that contain a steady-state differential that is not affected by the level of the commercial exchange rate; the essential difference remains that in this model the steady-state differential is equal to zero.

IV. The Effects of Monetary Policy on the Exchange Rate Differential

The imposition of a dual exchange rate regime is often justified by the authorities' desire to insulate the domestic economy from the effects of disturbances in international financial markets. This section therefore considers the effects of an increase in the international interest rate, i^*, on the exchange rate differential and on the domestic interest rate. A further reason for the imposition of dual exchange rate regimes has often been the authorities' desire to acquire some independence for the conduct of monetary policy without having to let the exchange rate float freely. This section therefore considers the effects of anticipated and unanticipated increases in domestic credit on the exchange rate differential and calculates the extent to which reserve flows would offset any change in domestic credit.

For the purpose of the discussion, it is convenient to begin with an examination of the effects of an expansionary monetary policy—that is, an unanticipated increase in domestic credit. The long-run effects of such a policy in the context of the model used here can be seen immedi-

ately from the condition that domestic credit of the steady-state money supply is determined by the fixed commercial exchange rate. The increase in domestic credit must therefore be offset, in the long run, by an equivalent loss in reserves. Since the behavior of prices and thus the trade account is also determined by the fixed commercial exchange rate, the reserve loss can be caused only by arbitrage activity, which in turn has to be induced by a differential between the financial and commercial exchange rates. Formally, this condition can be shown by computing the initial jump in the financial exchange rate, Δe_t, as a function of the change in domestic credit, $\Delta\overline{DC}$:[16]

$$\Delta e_t = (\phi\lambda/M)^{1/2}\Delta\overline{DC}. \tag{13}$$

This expression implies that the initial jump in the financial exchange rate is an increasing function of the parameter ϕ that characterizes the degree of separation of the two exchange rate markets. A higher degree of separation (in the sense of a high cost for potential arbitrageurs and a high value of ϕ) implies a larger initial jump in the financial exchange rate because the required outflow of reserves can take place only if there is a large differential that causes the arbitrage flows in spite of the high costs associated with them. After the initial discrete depreciation of the financial rate, the rate of depreciation becomes negative (that is, the financial rate appreciates). It continues to appreciate at a decreasing rate until the differential disappears.

These results may also be compared with those of the no-leakage models in the literature, such as Cumby (1984). In that paper a monetary expansion is not fully offset by an outflow through the balance of payments because a monetary expansion leads to a permanently higher financial exchange rate—which implies a higher real value of the net foreign assets held by the private sector and, through portfolio effects, a higher demand for money. Such a result would not be possible in the context of this model even if it did include portfolio effects because the leakage ensures that in the long run the financial and commercial rates have to be equal.

An equivalent result can be easily obtained for the effects of a change in the international interest rate. The money market condition of equation (6) implies that a change in domestic credit of ΔDC is exactly equivalent to a change in the international interest rate of $\Delta(M/\lambda)i^*$. Using this fact and equation (12) implies immediately that an un-

[16] See equation (26) in the Appendix for a derivation of this result.

anticipated increase in the international interest rate has to lead to a jump in the financial exchange rate equal to

$$\Delta e_f = (\phi M/\lambda)^{1/2}\,\Delta i^* . \tag{14}$$

The adjustment pattern after the initial jump is then parallel to the one following an increase in domestic credit: the initial differential disappears over time as the financial rate appreciates (at a decreasing rate) until it is equal to the commercial rate. The reserve flows induced by this differential reduce the domestic money supply and thus raise the domestic interest rate until it is equal to the international interest rate. This result implies that dual exchange rates cannot protect domestic financial markets from the effects of disturbances in international financial markets.

The effectiveness of monetary policy has often been discussed in terms of the so-called offset coefficient, which measures by how much any given change in domestic credit is offset by reserve flows. In this framework, the offset coefficient is a function of time, as can be shown by computing the change in reserves induced by a given change in domestic credit (see the Appendix for details of computations):

$$F_t - F_0 = \Delta\overline{DC}\,(e^{-\mu t} - 1). \tag{15}$$

Given the continuous time formulation, the impact coefficient (that is, for $t = 0$) is equal to zero. However, over time the (absolute value of the) offset coefficient rises and goes to (negative) unity in the long run (that is, as t goes to infinity).

The use of dual exchange rate regimes has often been advocated on the grounds that a dual exchange rate isolates goods markets from the effects of disturbances in financial markets. In this model, the goods market (characterized by the level of the real exchange rate) is indeed isolated from the effects of disturbances in the financial markets, but because the effects of disturbances in financial markets cannot be eliminated, they show up in the shadow price associated with the differential between the financial and the commercial exchange rates. If the cost of moving from one market to the other is high, this shadow price has to move by more for a given disturbance, because only in this way can the shadow price induce the flows of funds that ultimately neutralize the effects of the original disturbance. Because the deadweight loss associated with the arbitrage flows is an increasing function of the differential between the two exchange rates, a better separation of the two markets might not improve welfare. In general, this consideration thus contributes another argument against the use of dual exchange rate regimes.

The effects of an anticipated future increase in domestic credit can be derived from the requirement that no discrete jump in the financial exchange rate should occur at the time that the increase in domestic credit actually occurs. This requirement implies that the financial exchange rate has to jump at the time the news of the future increase in domestic credit policy is received. The exchange rate differential that is then created leads immediately to capital outflows, the money supply starts to fall, and the rate of depreciation of the financial exchange rate becomes positive. The financial rate will then continue to depreciate increasingly until the anticipated increase in domestic credit actually occurs. At that point in time, the depreciation of the financial rate becomes negative (that is, the financial rate appreciates), and it continues to be negative until the exchange rate differential disappears.

Because the domestic interest rate is equal to the international interest rate, i^*, plus the expected rate of depreciation of the financial rate, the results on the effects of monetary policy have important consequences for the viability and usefulness of dual exchange rate regimes. If the aim of the dual exchange rate regime is to keep domestic interest rates below international interest rates, the results shown in this section suggest that an increase in the money supply (that is, an increase in domestic credit) could achieve this objective only if it comes as a surprise; but even then the effects would be only temporary. If a one-time expansion of domestic credit was anticipated, it would have the undesired effect of raising domestic interest rates initially. A continuously expansionary domestic credit policy that was correctly anticipated would have no effects on domestic interest rates in the long run either. Such a policy would lead to a constant premium of the financial exchange rate that was just large enough to induce arbitrage outflows that neutralized the effect of the domestic credit expansion. Such a situation would be viable, but it would also imply that the only effect of the dual exchange rate system was to offer arbitrageurs the opportunity to make profits.

Similarly, if in response to an increase in international interest rates the authorities tried to neutralize the effects of arbitrage flows on domestic interest rates by increasing the rate of domestic credit expansion, the differential between the two exchange rates would widen, thus increasing the arbitrage flows. Any attempt by the authorities to conduct an independent money supply policy would then lead to an unstable spiral of depreciations of the financial rate and increasing arbitrage flows. A dual exchange rate regime (with a fixed commercial exchange rate) imposes, therefore, essentially the same constraints on the conduct of monetary policy as a unified fixed exchange rate.

V. Concluding Remarks

This paper has analyzed the consequences of incomplete market separation in dual exchange rate regimes by developing an explicit model of the arbitrage flows that occur when the two exchange rates differ. While it has often been argued informally that the differential between the two exchange rates should not be allowed to become too large, it appears that the consequences of arbitrage flows in response to the exchange rate differential have not previously been analyzed formally within a macroeconomic framework.

The main result of the paper is to show that these arbitrage flows would eliminate the exchange rate differential over time. This result is in contrast to the existing literature in which there is no force that reduces the exchange rate differential to enable it to settle down to any value in the long run. However, the data on the Mexican dual exchange rate market show that, in the absence of new shocks, the discount on the financial exchange rate tends to decline over time. The experience of the Belgian dual exchange rate regime, which has been in place for almost forty years, also shows that the discount on the financial rate tends to stay close to zero except in times of turbulence when devaluations of the commercial rate are expected. The data from these two countries thus support the framework proposed here.

An important corollary of the tendency of the two exchange rates to converge is that any attempt by the authorities to neutralize permanently the effects of these arbitrage flows would lead to an unstable situation. The use of dual exchange rate regimes to offset permanently the effects of permanent shocks should therefore not be attempted.

By using a simple macroeconomic model, the paper also analyzed the effects of various disturbances on the exchange rate differential. An unanticipated devaluation of the commercial exchange rate (or a fall in the long-run equilibrium real exchange rate) would lead to an overshooting of (and thus to a discount on) the financial exchange rate if the trade balance was sufficiently elastic with respect to changes in the commercial exchange rate. An unanticipated increase in domestic credit would always lead to a discount on the financial exchange rate; moreover, this discount is an increasing function of the degree of severity of the regulations that are used to keep the two exchange markets separate.

Because arbitrage flows link the two exchange rates, the financial exchange rate, which is forward looking, reacts to the anticipation of future disturbances. A future devaluation of the commercial rate or a future increase in domestic credit would therefore lead to an immediate

discount on the financial rate. This result is confirmed by the data for Belgium, which show that the discount on the financial rate tends to increase sharply before each devaluation of the commercial franc inside the EMS.

APPENDIX

Derivations

The system of equations (5)–(7) in the text is recursive; the path of prices is determined by equation (7) alone and acts as an exogenous, forcing variable in the reduced system of equations (5) and (6). The particular solution for the two variables F and e, therefore, contains not only a constant but also a term in [exp $(-\alpha t)$]. The sum of the particular and the homogenous solutions for F_t and e_t can therefore be written as

$$F_t = Ae^{-\mu t} + F_{p,t} = Ae^{-\mu t} + Be^{-\alpha t} + \bar{F} \tag{16}$$

$$e_t = Ce^{-\mu t} + e_{p,t} = Ce^{-\mu t} + De^{-\alpha t} + \bar{c}, \tag{17}$$

where A, B, C, and D are constants to be determined by the initial conditions. Equation (17) already uses the result that the stationary state implies that $\bar{e} = \bar{c}$.

The Reaction of the Financial Rate to a Devaluation of the Commercial Rate

To calculate the jump in the financial exchange rate that results from an unanticipated devaluation of the commercial exchange rate by $\Delta\bar{c} > 0$ from $\bar{c}$ to $\bar{c}'$, it is convenient to assume that initially the economy was at a stationary state with $\bar{p} = \bar{c} - \bar{k}$. The particular solution (the last two terms in equations (16) and (17)) has to hold at each point in time. Using this fact and the solution for the price level $p_t = -\Delta\bar{c}e^{-\alpha t} + \bar{c}' - \bar{k}$ in equations (5) and (6) yields

$$-\beta\Delta\bar{c} = \alpha B - D/\phi \tag{18}$$

$$-\lambda\Delta\bar{c} = -\alpha D + \lambda B/M. \tag{19}$$

Solving equation (18) for B and substituting into equation (19) yields

$$-\lambda\Delta\bar{c} = -\alpha D + \lambda[-\beta\Delta\bar{c} + (D/\phi)]/M\alpha \tag{20}$$

$$D = \frac{\beta - M\alpha}{\left(1 - \alpha^2\frac{M\phi}{\lambda}\right)}\Delta\bar{c} = \phi\frac{\beta - M\alpha}{1 - (\alpha/\mu)^2}\Delta\bar{c}, \tag{21}$$

where the second equality sign comes from the result that $\mu = \pm(\lambda/M\phi)^{1/2}$. At the time of the jump in the commercial exchange rate, the money supply is given because it is a slowly adjusting variable. Since the system is assumed to start from a stationary state, this implies that $\dot{e}_0 = 0$. Using this initial condition in the complete solution for the exchange rate (17) yields

$$0 = -\mu C - \alpha D. \tag{22}$$

Combining equation (22) with equation (21) implies that the exchange rate at time zero is given by

$$e_0 = \left(1 - \frac{\alpha}{\mu}\right) \phi \frac{\beta - M\alpha}{1 - (\alpha/\mu)^2} \Delta\bar{c} + \bar{c}', \tag{23}$$

or, since the financial rate was equal to $\bar{c}$ before the devaluation,

$$\Delta e = e_0 - \bar{c} = \left[\phi \frac{\beta - M\alpha}{1 + (\alpha/\mu)} + 1\right] \Delta\bar{c}. \tag{24}$$

Effects of a Change in the Equilibrium Real Exchange Rate

The effects of a change in the equilibrium real exchange rate from $\bar{k}$ to $\bar{k}'$ by $\Delta\bar{k} < 0$ can be calculated using equations (18)–(22), where $\Delta\bar{c}$ has been replaced by $-\Delta k$. The only difference is in equations (23) and (24); the new steady rate for the financial rate is still given by $\bar{c}$, which implies that the initial value of the financial exchange rate is given by

$$e_0 = \left(1 - \frac{\alpha}{\mu}\right) \frac{\beta - M\alpha}{1 - (\alpha/\mu)^2} (-\Delta\bar{k}) + \bar{c}. \tag{23a}$$

The initial jump in the financial exchange rate is therefore given by

$$\Delta e = e_0 - \bar{c} = \left(\frac{\beta - M\alpha}{1 + (\alpha/\mu)}\right) (-\Delta\bar{k}). \tag{24a}$$

The Effects of Domestic Credit Policy

The effect of an unanticipated jump in domestic credit of $\Delta\overline{DC}$ (from $\overline{DC}$ to $\overline{DC}'$) is more straightforward to calculate. In this case $p = \bar{p} = \bar{c}$, and the particular solutions consist only of a constant term. The initial condition in this case is that $\Delta e = -i - \lambda[\ln (F - \overline{DC}') - \bar{p}]$. From equations (16) and (17), this implies

$$\frac{\Delta\overline{DC}}{\bar{M}} \lambda = \mu C. \tag{25}$$

The initial jump of e_t is equal to C; this implies that

$$e_0 - \bar{c} = \Delta\overline{DC} \frac{\lambda}{\bar{M}\mu}. \tag{26}$$

Using $\mu = (\lambda/M\phi)^2$ yields equation (12) of the text.

Effects on the Initial Jump of the Financial Rate of Making Evasion Activity More Costly

The effect of a change in the parameter ϕ on the absolute value of the initial jump in the financial exchange rate that occurs in response to an unanticipated devaluation of the commercial exchange rate can be calculated from text equation (10):

$$\frac{\partial \Delta e}{\partial \phi} = [|\beta - \alpha M|] \left\{ \frac{1 + \alpha \frac{\phi M}{\lambda} - \phi\alpha \frac{1}{2} \left(\frac{\phi M}{\lambda}\right)^{-1/2} \frac{M}{\lambda}}{\left[1 + \alpha \left(\frac{\phi M}{\lambda}\right)^{1/2}\right]^2} \right\}, \tag{27}$$

where the term (α/μ) in the denominator of equation (10) has been substituted using the result $\mu = (\lambda/\phi M)^{1/2}$. The expression above can be simplified to

$$\frac{\partial \Delta e}{\partial \phi} = [|\beta - \alpha M|]\left[1 + \alpha\left(\frac{\phi M}{\lambda}\right)^{1/2}\right]^{-2}\left[1 + (1/2)\alpha\left(\frac{\phi M}{\lambda}\right)^{1/2}\right], \tag{28}$$

which is always positive.

REFERENCES

Adams, Charles, and Jeremy Greenwood, "Dual Exchange Rate Systems and Capital Controls: An Investigation," *Journal of International Economics* (Amsterdam), Vol. 18 (February 1985), pp. 43–63.

Bhandari, Jagdeep S., and Bernard Decaluwe, "A Stochastic Model of Incomplete Separation Between Commercial and Financial Exchange Markets," *Journal of International Economics* (Amsterdam), Vol. 22 (February 1987), pp. 22–55.

Cumby, Robert, "Monetary Policy Under Dual Exchange Rates," *Journal of International Money and Finance* (Guildford, England), Vol. 3 (August 1984), pp. 195–208.

Dornbusch, Rudiger, "The Theory of Flexible Exchange Rate Regimes and Macroeconomic Policy," *Scandinavian Journal of Economics* (Stockholm), Vol. 78 (May 1976), pp. 255–75.

———, "Special Exchange Rates for Capital Account Transactions," NBER Working Paper 1659 (Cambridge, Massachusetts: National Bureau of Economic Research, July 1985).

Flood, Robert P., and Nancy Peregrim Marion, "The Transmission of Disturbances Under Alternative Exchange Rate Regimes with Optimal Indexing," *Quarterly Journal of Economics* (Cambridge, Massachusetts), Vol. 97 (February 1982), pp. 43–68.

Frenkel, Jacob A., and Assaf Razin, "The Limited Viability of Dual Exchange Rate Regimes," NBER Working Paper 1902 (Cambridge, Massachusetts: National Bureau of Economic Research, April 1986).

Lanyi, Anthony, "Separate Exchange Markets for Capital and Current Transactions," *Staff Papers*, International Monetary Fund (Washington), Vol. 22 (November 1975), pp. 714–49.

Lizondo, José Saúl, "Exchange Rate Differential and Balance of Payments Under Dual Exchange Markets" (unpublished; Washington: International Monetary Fund, 1984).

———, "Unification of Dual Exchange Markets," *Journal of International Economics* (Amsterdam), Vol. 22 (February 1987), pp. 57–77.

Reding, Paul, "Interest Parity in a Two-Tier Exchange Rate Regime: The Case of Belgium 1975–84," Cahiers de la Faculté Des Sciences Economiques et Sociales de Namur, Serie Recherche No. 65 (March 1985).

Costly Trade Liberalizations

Durable Goods and Capital Mobility

GUILLERMO A. CALVO*

The social costs of a temporary liberalization policy in the context of an economy with infinitely lived individuals and no intertemporal consumption substitution are studied. Importable goods, however, can be stored, and storability is the central source of distortions. Possible welfare costs of the induced inventory accumulation are shown to be significant. It is also argued that temporariness is formally equivalent to "lack of credibility." Because international capital mobility is the vehicle that magnifies these distortions, the results suggest that an optimal trade liberalization without full credibility could call for controls on international capital mobility.

THIS PAPER DISCUSSES some simple examples intended to develop intuition about the costs of a temporary trade liberalization policy or, what amounts to very much the same thing, the costs of policies that are *expected* to be temporary. The motivation for this line of research is the observation that policies fall into the "temporary policy" category any time that their credibility is less than perfect. Because perfect credibility is likely to be more the exception than the rule, it follows that the case of temporary trade liberalization should be more relevant for actual policymaking than the free-trade-forever paradigm that has played such a central role in trade theory.

As elaborated in a previous study (Calvo (1987)), some of the costs of temporary policy are due to the intertemporal substitution in consump-

* Mr. Calvo is Professor of Economics and Co-Director of the International Economics Research Center at the University of Pennsylvania and holds a Ph.D. from Yale University. This paper was written while he was a visiting scholar in the Research Department.

tion that it generates. In contrast, the present paper will completely ignore those effects by assuming that there is no intertemporal substitution of that kind; instead, it will be assumed that consumption goods are durable and storable, thus opening the door for "supply-side" intertemporal substitution.

The examples assume the existence of a representative, infinitely lived individual; there are two goods—exportables and importables—and the economy is endowed with a fixed amount of each. Most of the paper assumes that consumption takes the form of importable goods, although extensions to fixed proportions are discussed. Moreover, the country is assumed to be "small" in both goods and capital markets; the latter are perfectly competitive. To highlight the role of durability, it is further assumed that individuals always choose a constant consumption path so that, given wealth, consumption is not affected by changes in rates of interest.

Section I discusses the basic and simplest example, in which goods are perfectly durable and storable at no cost and in which trade liberalization is announced at time zero ("initial time") and is expected to be instantaneously reversed. This case is that of "one-instant liberalization"—a policy that is costly because it leads the private sector to accumulate inventories in anticipation of the future tariff. More interesting, however, is that these costs (measured as the proportional loss of consumption in relation to free trade) could be quite significant. The model is extended in Section II to consider finite-horizon liberalizations and imperfect durability and to allow for the consumption of exportable goods. It is shown that the quantitative results are not significantly changed.

A common feature of the above examples is that they imply unrealistically high inventory levels, in part because of the assumption that the economy can accumulate indefinitely large amounts of inventories at a point in time at no extra cost. Section III deals with this issue by discussing the implications of quantitative constraints and of storage costs. Examples are given to show that the presence of the latter succeeds in lowering the ratio of inventories to gross national product (GNP) to more realistic levels, but still the cost of the liberalization policy continues to be sizable. The reason is that, although storage costs help to reduce the social opportunity costs associated with speculative inventory accumulation, such costs contribute significantly to the total cost of temporary policy.

In Section IV the paper closes with some brief notes on temporariness, credibility, and capital mobility.

I. The Basic Model: One-Instant Liberalization

Assume that the economy is populated by identical individuals, each of whom receives an exogenous path of "mana-type" income in terms of exportable goods. Utility, however, depends only on the consumption of importables (extensions are discussed in Section II). For the sake of contrast with a previous study (Calvo (1987); see also Section IV), it is assumed that there is no intertemporal substitutability and that, therefore, in the absence of uncertainty (as assumed in this paper), consumers choose a level of consumption that is constant over time.

Individuals are perfect competitors in goods and bond markets. For simplicity, it is assumed that the terms of trade are constant—and equal, by normalization, to unity—and that the rate of interest r (that is, the own-rate of interest of tradable goods in international markets) is a positive constant through time. Furthermore, it is assumed that consumer goods can be stored with no direct cost to the consumer and are perfectly durable. These assumptions will be relaxed in Sections II and III.

The domestic relative price of importables, in terms of exportables, is denoted by p, which will differ from unity because of tariffs ($p - 1 =$ the tariff rate). It is assumed that the proceeds of the tariff are given back to the public in a lump-sum (egalitarian) manner.

The "present" is time $t = 0$. The experiment examined in this section consists of setting the tariff equal to zero for $t = 0$, with the understanding (or the expectation) that it will be greater than zero (and constant) forever after. In other words, this is a case of a "one-instant" trade liberalization. Thus, more formally,

$$p_0 = 1 \tag{1a}$$

$$p_t = p, \qquad \text{for } t > 0, \tag{1b}$$

where p is a constant greater than unity.

Under the present circumstances, it is clear that if a utility-maximizing individual purchases inventories of importable goods at $t = 0$ (denoted by Z), he will plan to consume from his stock until it becomes depleted. Moreover, it is also clear that after $t = 0$ he would have no incentive to accumulate stocks any further because the interest rate is positive and p is expected to be constant.[1]

[1] This, incidentally, shows that if p is expected to be always constant (that is, including $t = 0$), then there would be no incentive for storing importable goods.

Thus, without loss of generality, the relevant budget constraint for the representative consumer may be expressed as follows:

$$K + Y + G - Z - pc\int_{\tau}^{\infty} e^{-rt}dt = 0, \tag{2}$$

where r is the own-rate of interest on exportables; K, Y, and G, also in terms of exportables, denote the present values of the initial holdings of the international bond, claims on future endowments of the exportable good, and (tariff-related) government lump-sum transfers, respectively; c stands for the constant level of consumption; and τ denotes the time at which the stock of importables, Z, is depleted. The integral in equation (2) runs from time τ because, as noted before, the individual is not going to import anything until his stock of importables reaches zero. Subsequently, he will import only what he requires for consumption.

Perfect durability implies that the inventory depletion time, τ, satisfies

$$\tau c = Z. \tag{3}$$

Hence, by equations (2) and (3),

$$r(K + Y + G) - c(x + pe^{-x}) = 0, \tag{4}$$

where x is the product of the rate of interest, r, and the inventory depletion time, τ; that is,

$$x = r\tau. \tag{5}$$

It is assumed, of course, that the representative individual attempts to maximize his consumption level, c, subject to his budget constraint, equation (4). This assumption obviously leads him to choose τ so that, by equation (5), the associated x maximizes

$$x + pe^{-x}. \tag{6}$$

Thus, x is set such that

$$e^{x} = p. \tag{7}$$

Hence, by equations (5) and (7), the individually optimal value of τ is determined solely by r and p, and it is positive if and only if $p > 1$ (that is, the future tariff is positive).

Now, given the assumption about the distribution of the tariff proceeds, one has

$$G = (p - 1)\hat{c}\int_{\hat{\tau}}^{\infty} e^{-rt}dt, \tag{8}$$

Table 1. *Costs of Liberalization, Variable* rT

(In percent)

Price, p	$rT = 0$[a]	$rT = 0.04$	$rT = 0.12$	$rT = 0.2$
1.03	0.04	0.04	0.04	0.04
1.27	2.57	2.47	2.29	2.12
1.60	8.68	8.36	7.71	7.21
2.20	19.55	18.92	17.73	16.59
2.89	28.94	28.12	26.53	25.00
3.40	34.12	33.23	31.48	29.78
4.00	38.88	37.94	36.08	34.25

[a] Corresponds to ϕ (Section I).

where a circumflex ($\hat{}$) on a variable denotes the equilibrium market value. Because each individual is identical, however, $c = \hat{c}$, and from equations (4), (5), (7), and (8) one has

$$\hat{c} = \frac{r(K + Y)}{\log p + 1/p}. \tag{9}$$

Clearly, a planner who maximizes welfare of the representative individual will choose c to be constant and equal to $r(K + Y)$. Thus, it is natural to define the cost of the present policy, ϕ, as follows:

$$\phi = 1 - \frac{\hat{c}}{r(K + Y)}. \tag{10}$$

Hence, by equations (9) and (10), one has

$$\phi = 1 - \frac{1}{\log p + 1/p}. \tag{11}$$

An attractive feature of the present case is that the cost of the policy is merely a function of p; that is, it does not depend on the interest rate.[2] Table 1 (second column) displays the value of ϕ for some relevant values of p when $rT = 0$. Note that the 1 percent cost mark, a relatively large number in this literature (see Harberger (1959)), is quickly reached when the import tariff ($p - 1$) is less than 27 percent. As indicated in Table 2 (last column), however, the value of τ is also very sensitive to p,

[2] The calculations would also apply to an economy that produces the importable good, such that the output of exportables and importables is not affected by the tariff.

Table 2. *Inventory Depletion Time, τ*

(In years)

Price, p	$\rho = 0.1$	$\rho = 0.5$	$\rho = 1.5$	$\rho = 2.5$	$\rho = \infty$ [a]
1.03	0.07	0.24	0.44	0.53	0.73
	(0.07)	(0.23)	(0.44)	(0.56)	(0.74)
1.27	0.50	1.99	3.58	4.27	5.97
	(0.60)	(2.14)	(3.70)	(4.33)	(5.82)
1.60	1.07	3.92	7.05	8.39	11.75
	(1.32)	(4.23)	(7.27)	(8.37)	(10.73)
2.20	1.79	6.57	11.80	14.08	19.55
	(2.53)	(7.75)	(11.79)	(13.15)	(15.86)
2.89	2.41	8.84	15.92	18.95	28.94
	(3.79)	(10.57)	(15.00)	(16.35)	(18.53)
3.40	2.78	10.20	18.36	21.85	34.12
	(4.65)	(12.17)	(16.59)	(17.87)	(20.16)
4.00	3.15	11.55	20.79	24.75	39.89
	(5.57)	(13.67)	(17.96)	(19.15)	(21.18)

Note: It is assumed that $r = 0.04$ per year. Numbers in parentheses indicate the ratio of inventories, Z, to annual GNP at time zero, which equals $r(K + Y)$.
[a] Corresponds to perfect durability; that is, $\delta = 0$.

implying perhaps implausibly long periods before the stock of importables is totally depleted and unrealistically high ratios of inventories to GNP. These limitations, and common sense as well, dictate that the model be extended to account for realistic features such as imperfect durability.

II. Imperfect Durability, Finite Liberalization, and Consumption of Exportable Goods

An easy extension of the above results is to assume that the liberalization lasts for $T > 0$ periods. It is quite intuitive that individuals will wait until time $t = T$ before storing importable goods. At that point they will solve exactly the same problem as discussed in the previous section, except that now the inefficient accumulation of inventories happens at a later time, so one would expect the cost of liberalization to be lower than before. In fact, if the cost of this policy is defined as earlier, and denoted by ϕ_T, one can show that

$$\phi_T = 1 - \frac{1}{e^{-rT}/(1-\phi) + 1 - e^{-rT}}. \tag{12}$$

Clearly, ϕ_T is a function of the product rT. Table 1 shows some relevant numbers for this relationship.[3] To confirm the intuition, the picture that emerges shows that costs are smaller than if $T = 0$.[4] Most important, the quantitative significance of the costs is not shown to be substantially diminished.

A slightly less obvious extension would be to allow for imperfect durability. Consider, for example, the case in which durable goods depreciate at the constant rate $\delta \geq 0$. Clearly, equation (3) becomes

$$\frac{c}{\delta}(e^{\delta\tau} - 1) = Z; \tag{3a}$$

hence, equation (4) becomes

$$r(K + Y + G) - c[r(e^{\delta\tau} - 1)/\delta + pe^{-r\tau}] = 0. \tag{4a}$$

Thus, consumption maximization is equivalent to minimizing the bracketed expression in equation (4a) with respect to τ, which yields

$$\tau = \frac{\log p}{\delta + r}. \tag{13}$$

An implication of the above analysis is that

$$-\frac{\partial\tau}{\partial\delta}/\tau = (\delta + r)^{-1}. \tag{14}$$

Consequently, if equation (14) is evaluated at $\delta = 0$ (the case of perfect durability) and $r = 0.04$ per year is assumed, it follows that a 1 percent per year increase in the rate of depreciation results in a 25 percent shortening in the depletion time of inventories, τ. This relationship shows, in the simulations exhibited in Table 2, why there is such a sharp decrease in the depletion time as δ is increased above zero—the case of perfect durability.

Denoting by ϕ^δ the cost of a one-instant liberalization policy (that is, equation (1)) when the rate of depreciation is $\delta > 0$, by equations (8), (4a), and (13) one obtains[5]

[3]Note that if $r = 0.04$ per year, Table 1 covers the cases of $T = 0$, 1, 3, and 5 years.

[4]This would not necessarily be so, however, if there were some intertemporal substitutability (Calvo (1987)).

[5]Costs are measured as in equation (10), with $\hat{c}$ standing now for the equilibrium consumption level when inventories depreciate at the rate δ.

$$\phi^{\delta} = 1 - \frac{1}{p^{1/(1+\rho)}(\rho + 1) - \rho - (p - 1)/p^{1/(1+1/\rho)}}, \tag{15}$$

where ρ is the ratio of the rate of interest, r, to the rate of depreciation, δ; that is,

$$\rho = r/\delta. \tag{16}$$

Thus, contrary to the results of Section I, costs are a function of the ratio of the rate of interest to the rate of depreciation.

Notice that, given r, there is no a priori reason to expect the cost to be a monotonic function of durability. When goods are instantly perishable ($\delta \to \infty$), costs would be zero because there will be no inventory accumulation; but if stocks of durable goods reproduce themselves at the rate r (that is, $\delta = -r$), then, once again, costs would be zero because the economy as a whole would be indifferent between holding a foreign bond and carrying inventories of importable goods. Because intermediate cases exhibit positive costs, it follows that it would not be possible to say in general whether more durability would increase or decrease the costs of a trade liberalization policy such as that specified in equation (1).

In Table 2 one sees a sharp fall in the depletion time, τ, and in the ratio of inventories to GNP with respect to the case of perfect durability, making the results somewhat more realistic. In Table 3 one notices that, even when a wide variety of depreciation rates is tried (if the annual $r = 0.04$, the simulations cover the cases of $\delta = 0.4$, 0.08, 0.027, 0.016, and 0 per year), the 1 percent cost mark continues to be reached in most cases when the tariff ($p - 1$) is less than 30 percent. Furthermore, costs of more than 10 percent are still quite possible except for the rather extreme case in which durable goods depreciate at the rate of 40 percent per year (see the column corresponding to $\rho = 0.01$ in Table 2; for this computation it is still assumed that $r = 0.04$ per year).

Finally, the assumption that only importable goods are consumed domestically is relaxed. A simple way to do this is to assume that individuals consume both goods in fixed proportions.[6] Thus, if the ratio of the consumption of exportables to the consumption of importables is α, the budget constraint (2) becomes

$$K + Y + G - Z - c(pe^{-r\tau} + \alpha)/r = 0. \tag{2a}$$

Hence, if in this case the costs under perfect durability are denoted by $\phi(\alpha)$, one can show, on the basis of equations (3), (4), and (2a), that

[6]Given the purpose of this paper, this assumption is much less restrictive than it sounds; in allowing for substitutability between importables and exportables, one would imply the existence of gains from trade—from which the present paper is trying to abstract.

Table 3. *Costs of Liberalization, Variable Durability*
(In percent)

Price, p	$\rho = 0.1$	$\rho = 0.5$	$\rho = 1.5$	$\rho = 2.5$	$\rho = \infty$ [a]
1.03	0.00	0.01	0.03	0.03	0.04
1.27	0.27	0.97	1.66	1.94	2.57
1.60	1.13	3.75	6.06	6.90	8.68
2.20	3.40	10.29	15.20	16.73	19.55
2.89	6.58	17.80	24.37	26.15	28.94
3.40	9.00	22.81	29.92	31.68	34.12
4.00	11.83	28.05	33.56	36.97	38.89

[a] Corresponds to perfect durability; that is, $\delta = 0$.

$$\phi(\alpha) = 1 - \frac{1 + \alpha}{\log p + 1/p + \alpha}, \tag{17}$$

which, as expected, reduces to ϕ in equation (11) when $\alpha = 0$. Clearly, costs are a decreasing function of α. As shown in Table 4, however, costs continue to be sizably large, even when individuals are assumed to consume equal values of importable and exportable goods.

III. Quantity Constraints and Direct Costs

The numerical examples presented thus far are not yet fully persuasive because they tend to imply levels of inventories that are several times annual GNP. Without this enormous accumulation of inventories, costs would probably not exceed the usual 1 percent or 2 percent levels.

The equilibrium level of inventories may change substantially if the accumulation of inventories is quantity constrained. Quantitative limits could be due to the existence of physical constraints, such as maximum

Table 4. *Costs of Liberalization,* $\alpha = 1$
(In percent)

Price, p	$\phi(1)$	$Z/r(K + Y)$
1.03	0.02	0.74
1.27	1.30	5.89
1.60	4.53	11.21
2.20	10.83	17.58
2.89	16.91	22.04
3.40	20.57	24.30
4.00	24.14	26.30

Note: The last column assumes that $r = 0.04$ per year.

port capacity, or to the existence of import quotas. The first type of constraint is likely to play some role under extreme circumstances, but it would probably be difficult to argue that, as a general rule, physical constraints will be the dominant force in limiting the size of inventories of international goods.[7] The second possibility, quotas, is only one of the policies that could be employed to reduce the costs of the lack of a stable tariff policy and therefore does not affect the relevance of the previous results.[8]

Another important factor that may have a sizable effect on the level of inventories and social welfare is the existence of direct costs of holding inventories—for example, warehousing costs. Consider a case in which the cost of accumulating a stock of inventories Z is βZ, where $\beta \geq 0$. The budget constraint for the representative individual becomes

$$r(K + Y + G) - rZ(1 + \beta) - cpe^{-r\tau} = 0. \tag{4b}$$

For the sake of brevity, one-instant liberalization policy such as in equation (1) will be examined when goods are perfectly durable. Recalling equations (3) and (8), and the procedure followed in Section I, one obtains the following expression for the cost of policy (1)—defined in the same manner as in the introduction to the paper:

$$1 - \frac{1}{(1 + \beta)[\log p + 1/p - \log(1 + \beta)]}. \tag{18}$$

The inventory depletion time, τ, however, satisfies

$$\tau = \frac{\log p - \log(1 + \beta)}{r}. \tag{19}$$

Table 5 shows results of some experiments in which β was chosen high enough so that the ratio of inventories to GNP is approximately equal to unity—a substantial reduction with respect to the corresponding numbers in Table 2. Note, first, that the required β is very large relative to the tariff-related gross revenue per unit of inventories ($p - 1$). Second, despite these enormous disincentives, the welfare cost hovers very near the 1 percent mark for a 27 percent tariff, and quickly rises above it. Finally, note that the depletion time of inventories, τ, is always less than five quarters.

[7] A more plausible constraint would be international credit market rationing. This aspect will not be examined here, however, because the focus of the paper is on the harmful effects of capital mobility when it is combined with the existence of storable goods; rationing, of course, represents a constraint on capital mobility.

[8] In this respect note that, in any application of the "quotas solution," the cost-reducing effect of a quota would have to be weighed against the rent-seeking costs that it may generate (Krueger (1974)).

Table 5. *Costs of Instant Liberalization when* $Z\Pi(K + Y) \simeq 1$

Price, p	β	Cost (percent)	τ
1.27	0.22	0.95	1.00
1.60	0.53	2.41	1.12
2.20	1.11	4.51	1.04
2.89	1.77	7.06	1.06
3.40	2.25	9.30	1.12
4.00	2.82	11.57	1.15

Note: τ is number of years; $r = 0.04$ per year.

The main lesson from these experiments is that, although direct costs of holding inventories may induce a drastic reduction in their size and therefore result in a sharp decrease in the total opportunity cost of the funds devoted to acquire them (in previous sections, the only source of costs), the total costs, including direct costs, may still be quite sizable.

IV. Reinterpreting the Results: Credibility

The above framework is general enough to give some insight into the costs of credibility as regards liberalization policy. In common parlance, expressions such as "incomplete credibility" of policy are used to denote situations in which the public believes that there is a positive probability that policy announcements will not be carried out. Consequently, a policy that is not fully credible is one that elicits the expectation that it is going to be modified in the future. With this interpretation in mind, therefore, the examples in previous sections would correspond to situations in which the free-trade policy is not credible, and the public expects that it will be replaced by a constant-tariff policy after time T.

Obviously, the behavior of the economy during the transition (from zero to T) will be the same as in the previous examples. Furthermore, the actual cost of the noncredible liberalization policy would also be the same if inventories cannot be resold in international markets, and if a constant permanent tariff policy was expected after time T. Under these circumstances, let us consider the interesting case in which policymakers announce a free-trade-forever policy beginning at time zero, but, before time T, the public does not believe in its continuation after time T (the next elections, say), when they expect (with probability 1) that a constant and permanent tariff will be imposed. Furthermore, assume that if the public sees free trade continuing after time T, their disbelief will vanish altogether, and full credibility in the free-trade-forever policy will be ensured. Clearly, therefore, a free-trade-forever policy will be associ-

ated with exactly the same paths and the same welfare costs as in our examples. This illustrates the central point of the paper: *a trade liberalization policy that is not fully credible may be costly, and its costs may not be negligible, particularly when they are compared with the usual estimates for the gains from trade.*[9]

In closing, I would like to stress an obvious but important point. In the models developed here the distortions of credibility or temporariness were shown to be harmful because it was assumed that capital was perfectly mobile. In the context of the models, if there were no international capital mobility, the accumulation of inventories simply could not occur, and hence the social costs would be nonexistent. This observation provides some grounds for controlling international borrowing and lending during a trade liberalization program that is not fully credible. Such control could be accomplished by the sheer imposition of quantity controls (for example, quotas on durable goods or on foreign borrowing) or by a tax on international capital mobility. When credibility is at stake, however, there is a case for preferring direct quantity controls over taxes, since the latter require a good knowledge about the exact nature of the credibility problem. In the examples given here, for instance, the policymaker should be able to ascertain the expected future tariff, $p - 1$, and the expected timing of its implementation, T. This information is not required for quantity constraints.[10]

REFERENCES

Calvo, Guillermo A., "Incredible Reforms" (unpublished; Philadelphia: University of Pennsylvania, November 1986).

———, "On the Costs of Temporary Policy," *Journal of Development Economics* (Amsterdam), Vol. 27 (1987), pp. 245–62.

Edwards, Sebastian, and Sweder van Wijnbergen, "The Welfare Effects of Trade and Capital Market Liberalization," *International Economic Review* (Philadelphia), Vol. 27 (February 1986), pp. 141–48.

[9]To avoid any confusion, I would like to point out that in the context of the model, a noncredible tariff—a tariff that is expected to be phased out in the future—would have no welfare effects because there would be no incentive to accumulate inventories of importable goods. This asymmetry disappears immediately, however, when one allows for inventories of exportable goods. Consequently, the central problem is the lack of credibility of policy announcements, rather than the noncredibility of liberalization policies.

[10]This point is further elaborated in Calvo (1987). The role of capital mobility in the staging of economic liberalization policies has been recently discussed by Edwards and van Wijnbergen (1986).

Harberger, Arnold C., "Using the Resources at Hand More Effectively," *American Economic Review* (May 1959), pp. 134–46; reprinted as Chapter 5 of Arnold C. Harberger, *Taxation and Welfare* (Boston: Little, Brown, 1974), pp. 108–19.

Krueger, Anne O., "The Political Economy of the Rent-Seeking Society," *American Economic Review* (Nashville, Tennessee), Vol. 64 (June 1974), pp. 291–303.

Financial, Exchange Rate, and Wage Policies in Singapore, 1979–86

ICHIRO OTANI and CYRUS SASSANPOUR*

The unique aspects of Singapore's financial, exchange rate, and wage policies during the period 1979–86 are discussed. A simple short-term model is formulated and estimated, and various policy simulations conducted, to quantify the impact of alternative policies on major macroeconomic variables. It is found that the wage policy pursued by the authorities in the early 1980s played a significant role in influencing output and prices and that an appropriate wage policy is complementary to exchange rate policy in maintaining external competitiveness in Singapore.

SINGAPORE'S recent experience with the conduct of financial and exchange rate policies suggests that the authorities' ability to implement an independent monetary policy has been constrained. Because the economy is small and highly integrated with international markets, movements in the money supply have been influenced significantly by developments in the external sector, and domestic interest rates have been determined largely by foreign rates.[1] At the same time, the authorities' exchange rate policy has been geared toward the often conflicting objectives of mitigating external inflationary pressures and sustaining economic activity by increasing external competitiveness.

The cornerstone of financial policies in Singapore has been budgetary discipline, evidenced by the government's budget surpluses that have been recorded every year since the late 1960s. These surpluses have been

*Mr. Otani, Senior Economist in the Research Department when this paper was written, is now Chief of the English Division in the IMF Institute. He is a graduate of the University of California, Berkeley, and the University of Minnesota.

Mr. Sassanpour, an Economist in the Middle Eastern Department, received his doctorate from Brown University.

[1] Edwards and Khan (1985) found that over 90 percent of movements in domestic interest rates in Singapore were explained by changes in foreign rates.

substantially augmented by savings generated by the social security scheme (Central Provident Fund) and government-controlled financial institutions (for example, the Post Office Savings Bank), which have been deposited in government accounts with the Monetary Authority of Singapore (MAS). Consequently, government financial operations have drained liquidity from the banking system, creating persistently tight liquidity conditions.

Because the banking sector has easy access to international capital markets, the tight conditions of domestic liquidity have been eased by the inflow of funds from abroad. In the process, the authorities have provided domestic liquidity to the banking system by purchasing foreign exchange from, or arranging swap transactions with, the commercial banks.[2] As a result, the short-run increases in the net foreign assets of the MAS have been closely associated with increases in government deposits with the MAS. Reflecting these financial policies, monetary aggregates have exhibited significant short-term fluctuations, but over the longer run the upward trend in foreign assets has sustained the underlying growth of domestic liquidity.

Exchange rate movements have reflected the authorities' reserve management policies. During 1979–84, increases in foreign reserves were rather moderate, and the effective exchange rate of the Singapore dollar appreciated substantially.[3] Since early 1985, however, the level of reserves has risen sharply, and the effective exchange rate depreciated markedly, reversing most of the appreciation that had taken place earlier.

To benefit from the emerging pattern of external demand, the authorities pursued a high wage policy during 1979–85 with a view to inducing a shift in production away from labor-intensive to capital-intensive and "high-tech" activities. During this period labor costs increased at an annual average rate of 15 percent as nominal wages increased rapidly and the employers' (as well as workers') contributions to the Central Provident Fund were raised significantly. This wage policy had important implications for Singapore's external competitiveness and, thus, for economic activity. During the first half of the 1980s, Singapore's external competitiveness, measured in relative unit labor costs adjusted for the

[2] Under these swap arrangements, the MAS typically purchases foreign exchange from commercial banks at the prevailing spot exchange rate with the provision to unwind the swap on a future date at a rate determined at the time of the arrangement.

[3] This rate is defined as a weighted average of the bilateral exchange rates between the Singapore dollar and the currencies of Singapore's major trading partners and competitors.

exchange rate, deteriorated by more than 60 percent, contributing to the slowdown in economic activity and to the severe, but short-lived, recession in 1985—the first decline in more than two decades. In 1986, concerned with the loss of external competitiveness, the authorities implemented several measures to reduce labor costs, including imposing limits on nominal wage increases and reductions in contributions to the Central Provident Fund. The reduction in unit labor costs, together with the depreciation of the Singapore dollar, helped to regain some of the lost external competitiveness and contributed to economic recovery.

Although the causal links between key macroeconomic variables described above have been well identified, there has been little attempt to formalize such relationships.[4] The present paper is an attempt to fill this gap. The major objectives are to analyze transmission processes and to quantify the importance of alternative policy instruments in influencing key macroeconomic variables, including output, prices, the exchange rate, and foreign reserves. For this purpose, a simple short-term model is formulated and estimated, and various policy simulations are conducted.

The remainder of the paper is organized as follows. Section I discusses the underpinnings to the specification of the model, taking into account the unique features of the Singapore economy. Section II presents the estimates of the behavioral equations, and Section III analyzes the simulation results based on alternative policy scenarios. Section IV makes some concluding remarks and draws policy implications for macroeconomic management in Singapore. Data sources and definitions are given in the Appendix.

I. Model Specification

In this section a simple macroeconomic model is formulated to describe the movements of foreign reserves, the exchange rate, prices, real output, and government revenue and expenditure.

Foreign Reserves

Because financial operations have consistently resulted in a significant drainage of liquidity, as explained earlier, a familiar feature of financial operations in Singapore has been increases in net foreign assets to elim-

[4]Different aspects of financial and exchange rate policies in Singapore have been reviewed by studies contained in Singapore (1981).

inate the domestic excess demand for money. To describe this aspect of adjustment, we assume that actual holdings of real money balances (M/P) adjust with a lag to the difference between the desired holdings $(M/P)^d$ in the current period and the actual holdings in the previous period.[5] The rationale behind this equation is that, if the current demand exceeds the actual holdings in the previous period, the public adjusts its holdings by adding to its balances, and vice versa. This partial adjustment mechanism, expressed in terms of natural logarithms, can be described as

$$\Delta \ln (M/P)_t = k [\ln (M/P)_t^d - \ln (M/P)_{t-1}], \tag{1}$$

where k is the coefficient of adjustment and is expected to be positive and less than unity, and Δ is the first-difference operator.

The demand for real balances is specified as a function of real output (Y) and the domestic interest rate (r_d), with the latter measuring the opportunity cost of holding money:

$$\ln (M/P)_t^d = \alpha_0 + \alpha_1 \ln Y_t + \alpha_2 (r_d)_t, \tag{2}$$

where α_1 denotes the income elasticity and is expected to be positive, and α_2 is the semi-interest-rate elasticity and is expected to be negative.

From the money supply identity, changes in the stock of money are equal to the sum of changes in net foreign assets (R) and changes in net domestic assets (D):

$$\Delta M_t = \Delta R_t + \Delta D_t. \tag{3}$$

This relation can be log-linearized about the sample means and expressed in terms of the rate of change of variables in real terms as

$$\Delta \ln (M/P)_t = m_0 + m_1 \Delta \ln (R/P)_t + m_2 \Delta \ln (D/P)_t, \tag{4}$$

where $m_1 = (\overline{R}/\overline{M})$ and $m_2 = (\overline{D}/\overline{M})$ are, respectively, ratios of the sample means of net foreign assets and net domestic assets to the money supply, lagged by one period.

Substituting equations (2) and (4) in equation (1) and rearranging terms result in an equation that expresses the rate of change of net foreign assets in real terms as a function of variables in the demand for money equation, the lagged level of real balances, and the rate of change of domestic assets in real terms:

[5] In analyzing the money demand in Singapore, Khan (1981) found that the empirical results rendered considerable support for this type of dynamic specification.

$$\Delta \ln (R/P)_t = (m_0 + k\alpha_0)/m_1 + k\alpha_1/m_1 \ln Y_t + k\alpha_2/m_1 (r_d)_t - k/m_1 \ln (M/P)_{t-1} - m_2/m_1 \Delta \ln (D/P)_t. \quad (5)$$

According to equation (5), reserve accumulation occurs as long as the demand for money is left unsatisfied at the beginning of the period and is not met by domestic credit creation during this period.[6]

Exchange Rate

Exchange rate policy in Singapore should be viewed as an integral part of the authorities' financial management. By providing domestic liquidity to the commercial banks through swap arrangements or outright purchases of foreign exchange, the authorities have accumulated foreign assets to relieve the upward pressure on the exchange rate. At the same time, however, the authorities have had to consider the implications of the intervention policy for monetary stability and thus face a trade-off between the exchange rate and monetary objectives.[7]

It may be argued that the change in the exchange rate is a function of the gap between the authorities' desired level of the exchange rate (E^d) and the actual rate prevailing in the preceding period (E_{t-1}), as shown in equation (6):

$$\Delta \ln E_t = \phi [\ln E_t^d - \ln E_{t-1}], \quad (6)$$

where ϕ is the adjustment coefficient and is expected to be positive and less than unity; E is defined as the value of the domestic currency in terms of the foreign currency, such that an increase in E denotes an appreciation of the Singapore dollar.

The desired level of the exchange rate in a given period is assumed to be set by the authorities, who take into consideration the ratio of foreign assets to total assets at the end of the preceding period. As the level of foreign assets increases relative to domestic assets, reflecting the sales of domestic currency to the banking system, the desired level of the exchange rate would be lowered, other things being equal. In addition, the desired level of the exchange rate is assumed to respond to the discrepancy between the foreign inflation rate (Π_f) and the domestic rate (Π), with the lagged exchange rate as an anchor. This inflation rate differen-

[6] This formulation is consistent with the monetary approach to the balance of payments, as highlighted by Frenkel and Johnson (1976) and International Monetary Fund (1977), and as applied to small open economies by Otani and Park (1976) and Sassanpour and Sheen (1984).

[7] Branson (1981) emphasized the importance of this type of trade-off facing the policymakers in a small open economy such as Singapore.

tial can be interpreted as an indicator representing the premium or discount on the Singapore dollar, or as the *expected* rate of change in the exchange rate:[8]

$$\ln E_t^d = \beta_0 + \beta_1 \ln (R/D)_{t-1} + \beta_2 [(\Pi_f - \Pi)_t + \ln E_{t-1}], \tag{7}$$

where β_1 is expected to be negative and β_2 to be positive. In this equation, if the role of foreign reserves becomes negligible in the authorities' exchange rate policy, the desired level of exchange rate would be influenced more by purchasing power parity, and thus the value of β_2 would be close to unity. In this case, the real exchange rate would be constant in the long run. In contrast, if the role of foreign reserves becomes significant, β_2 would be close to zero, and the real exchange rate may not be constant in the long run.

Combining equations (6) and (7), we can express the dynamic adjustment of the exchange rate as

$$\begin{aligned}\Delta \ln E_t = {} & \phi\beta_0 + \phi\beta_1 \ln (R/D)_{t-1} + \phi\beta_2 (\Pi_f - \Pi)_t \\ & + (\beta_2 - \phi) \ln E_{t-1}.\end{aligned} \tag{8}$$

This functional specification is the authorities' reaction function and reflects their foreign reserve management policy as well as other factors influencing the exchange rate.[9]

Prices

The wholesale price index in Singapore, which is composed entirely of price indices of traded goods, reflects price developments in the world markets as well as movements in the exchange rate. Therefore, variations of the wholesale price index (P_w) are specified as a function of changes in import prices (P_m) and export prices (P_x), both denominated in terms of domestic currency:

$$\Delta \ln (P_w)_t = w_0 + w_1 \Delta \ln (P_m)_t + w_2 \Delta \ln (p_x)_t, \tag{9}$$

[8] Had information been available on the premium or discount of the Singapore dollar in relation to the currencies of Singapore's major trading partners and competitors, such information could have been used in place of the inflation rate differential. This information is available, however, only for the exchange rate between the Singapore dollar and the U.S. dollar.

[9] Equation (8) could be only one of many plausible specifications to represent the authorities' reaction function. It is not the purpose of this paper, however, to identify their true reaction function; such an attempt would require testing alternative reaction functions.

where w_1 and w_2, respectively, measure the contribution of import and export prices to changes in P_w. Both P_m and P_x are exogenously determined world market prices of Singapore's imports and exports, adjusted for the exchange rate.[10]

The rate of change in the consumer price index—$\Pi = \Delta \ln P$—has diverged from that in the wholesale price index, mainly because of changes in prices of nontraded goods in response to domestic demand and supply conditions. Thus the rate of change in the consumer price index is specified in equation (10) below as a function of the excess supply of domestic liquidity, which may serve as a counterpart to excess demand for nontraded goods. The domestic rate of inflation is also expected to respond positively to the excess of real output above its potential level. The potential output path, in logarithms, is defined as $Y_0 e^{\mu T}$, where Y_0 is the initial output level, μ is the potential (or steady-state) growth rate, and T is the time-trend term. In addition to demand and supply forces described above, the rate of change in the wholesale price index is included as an explanatory variable to capture the influence of traded-goods prices on the consumer price index:

$$\Delta \ln P_t = \gamma_1 [\ln (M/P)_{t-1} - \ln (M/P)_t^d] + \gamma_2 \ln (Y/Y_0 e^{\mu T})_t + \gamma_3 \Delta \ln (P_w)_t, \tag{10}$$

where γ_1, γ_2, and γ_3 are expected to be positive. Combining equations (2) and (10) and rearranging terms result in

$$\Delta \ln P_t = -(\gamma_1 \alpha_0 + \gamma_2 \ln Y_0) + \gamma_1 \ln (M/P)_{t-1} + (\gamma_2 - \gamma_1 \alpha_1) \ln Y_t - \gamma_1 \alpha_2 (r_d)_t + \gamma_3 \Delta \ln (P_w)_t - \gamma_2 \mu T. \tag{11}$$

Real Output

Given the openness of the economy and the limited size of the domestic market,[11] overall economic activity in Singapore is dominated by the activity in the external sector, which in turn is influenced in large part both by external demand conditions for Singapore's output and by do-

[10] That is, $P_m = P_m^*/E$ and $P_x = P_x^*/E$, where P_m^* and P_x^* are world market prices of imports and exports, respectively.

[11] The value of Singapore's exports has averaged about 140 percent of gross domestic product (GDP) in recent years.

mestic supply factors. As a result, the behavioral equation for real output needs to be derived from the demand and supply functions.

The world demand for Singapore's output could be posited as a function of the world real income (Y^*) and the real exchange rate, representing the ratio of Singapore's output price relative to that of its competitors in the world market ($E \cdot P/P^*$):[12]

$$\ln Y_t^d = \lambda_1 + \lambda_2 \ln Y_t^* + \lambda_3 \ln (E \cdot P/P^*)_t, \tag{12}$$

where λ_2 is assumed to be positive and λ_3 to be negative.

The supply of output in the short run is assumed to depend on the inverse of real wages (P/W) and the utilized capital stock (UK).

$$\ln Y_t^s = \delta_1 + \delta_2 \ln (P/W)_t + \delta_3 \ln (UK)_t, \tag{13}$$

where δ_2 and δ_3 are expected to be positive. Solving equation (13) for P, substituting the solution in equation (12), and assuming that the actual output equals the demand, we derive the output equation as

$$\ln Y_t = \psi_0 + \psi_1 \ln Y_t^* + \psi_2 \ln (E \cdot W/P^*)_t + \psi_3 \ln (UK)_t, \tag{14}$$

where

$$\psi_0 = \frac{\lambda_1 \delta_3 - \lambda_3 \delta_1}{\delta_2 - \lambda_3}, \qquad \psi_1 = \frac{\lambda_2 \delta_2}{\delta_2 - \lambda_3}, \qquad \psi_2 = \frac{\lambda_3 \delta_2}{\delta_2 - \lambda_3}, \qquad \psi_3 = \frac{-\lambda_3 \delta_3}{\delta_2 - \lambda_3}.$$

Given the sign conditions in equations (12) and (13), ψ_1 and ψ_3 are expected to be positive and ψ_2 to be negative. The term $E \cdot W/P^*$ is an indicator of external competitiveness and, for estimation purposes, is proxied by the relative unit labor costs adjusted for the exchange rate ($E \cdot RULC$), where $RULC$ is the ratio of Singapore's unit labor costs to that of its major trading partners.[13] The utilized capital stock is proxied by the real value of bank credit extended to the private sector (C_{ps}/P).[14] The first difference of equation (14), together with the proxy variables discussed above, yields the following equation:

$$\Delta \ln Y_t = \psi_1 \Delta \ln Y^* + \psi_2 \Delta \ln (E \cdot RULC)_t + \psi_3 \Delta \ln (C_{ps}/P)_t. \tag{15}$$

[12]The domestic consumer price index was used as a proxy for the GDP deflator; although the latter would have been more appropriate in this equation, the former was chosen to limit the number of endogenous variables and provide a link with the rest of the model.

[13]Branson and Love (1988) also used the index of unit labor costs, adjusted for changes in the exchange rate, to represent external competitiveness.

[14]In many developing and newly industrialized countries, the availability of bank credit plays an important role in influencing the rate of capacity utilization. The role of credit in output functions has been emphasized by Kapur (1976) and Keller (1980).

Government Finances

Government revenue has been increasing in Singapore because of high rates of economic growth and a buoyant tax structure. Public expenditure has also been rising in recent years, reflecting government efforts to meet the need for housing and infrastructure. Owing to budgetary discipline, however, expenditures have been maintained successfully within available domestic resources; indeed, since 1968 the government's fiscal operations have continuously resulted in surpluses.

Within this framework, the government expenditure (GE) function is formulated on the assumption that the government attempts to adjust its expenditure with a lag to increases in revenue. In other words, the more government revenue (GR) exceeds expenditure, the faster the latter rises:

$$\Delta \ln GE_t = g(\ln GR_t - \ln GE_{t-1}), \tag{16}$$

where g denotes the adjustment coefficient and is expected to be positive and less than unity. Rearranging terms in equation (16) leads to

$$\ln GE_t = g \ln GR_t + (1 - g) \ln GE_{t-1}, \tag{17}$$

which formulates government expenditure as a weighted average of current government revenue and the lagged level of expenditure.[15]

Government revenue is specified simply as a log-linear function of nominal income:

$$\ln GR_t = q_0 + q_1 \ln (P \cdot Y)_t, \tag{18}$$

where q_1 measures the elasticity of total revenue with respect to nominal income.[16]

Domestic Assets

With the money supply identity defined as in equation (3), the monetary sector is completed by defining the net domestic assets (D) of the banking system. The change in net domestic assets is equal to changes

[15] In the estimation process, a constant term was added to this equation to capture the budgetary surpluses that have been recorded.

[16] In this formulation, the actual and desired levels of revenue are assumed to be the same. Aghevli and Khan (1977) assumed that revenue adjusts with a lag to its desired level. This formulation was suggested for high-inflation countries because the authorities attempt to adjust revenue in the face of its declining value in real terms.

in stock of credit to the private sector and changes in net credit to the government (GC), including other items (net):[17]

$$\Delta D_t = \Delta(C_{ps})_t + \Delta GC_t. \tag{19}$$

The change in net credit to the government (GC) is defined as

$$\Delta GC_t = (GE_t - GR_t) - GNBB_t, \tag{20}$$

where the first term represents budgetary operations and the second term indicates changes in net government borrowing from the banking system for nonbudgetary operations ($GNBB$). As a result of budgetary surpluses and funds generated by the social security scheme and other operations, the government has been a net depositor of funds with the banking system.

Working of the Model

The essence of the model,[18] which conforms with the variety of formulations suggested by the monetary approach to the balance of payments, is that reserve accumulation occurs as long as the demand for money is not satisfied by increases in net domestic assets (equation (5)). In this model, the excess demand for money arises from the government's contractionary financial operations, which reflect traditional budget surpluses and sizable buildups of deposits by the government with the monetary authorities (equations (19) and (20)).

Because reserve accumulation reflects the monetary authorities' foreign exchange purchases from (or swaps with) the commercial banks, a significant reserve accumulation would prevent the exchange rate from appreciating (equation (8)). This effect would reinforce the direct influence of rising foreign prices on the traded-goods prices (equation (9)), which would in turn increase domestic consumer prices at a faster pace if they were rising, or decrease them at a slower pace if they were falling (equation (11)).

A more depreciated exchange rate would safeguard the external competitiveness and improve growth performance, provided that wage costs remain unchanged (equation (15)). A higher real output level, in turn,

[17] Private sector credit is treated as an exogenous variable in this model. It could be assumed that the stock of credit to the private sector is, in principle, demand determined, and this demand is always met by the supply of credit from the banking system.

[18] The process described below is presented in a static framework, but the actual process should be viewed in a dynamic setting in which interactions among variables occur simultaneously. This is done in Section III.

would increase the demand for money and reduce the pressure on domestic prices. Thus, the direct impact of exchange rate changes on these prices would be counteracted by the influence of output, and the net impact would depend on the relative size of the relevant coefficients.

Developments in domestic prices and real output have implications for the level of government revenue (equation (18)). To the extent that expenditure grows in line with revenue (equation (17)), however, the net monetary impact of budgetary operations should be minimal.

The upshot of the discussion is that a tight financial policy could trigger foreign exchange intervention that might induce a depreciation of the domestic currency and might lead eventually to an expansion of output. This possibility, however, depends on the values of the parameters of the behavioral equations and is thus an empirical question.

II. Empirical Results

Table 1 summarizes the estimates of the complete model, which comprises seven behavioral equations and three identities. The model is estimated by the two-stage least-squares method for the period 1979–86, using quarterly data.[19] The behavioral relationships are in general well estimated, and the model as a whole appears to capture the essential features of the Singapore economy.

The estimated equation for foreign reserve movements suggests that the extent of these changes is indeed systematically related to the excess demand for money that is not met by changes in domestic credit. From this equation, the long-run income and interest rate elasticities of the demand for money are calculated to be about 1.2 and -0.1, respectively. Furthermore, using the average sample value of $m_1 = 0.7$, the adjustment coefficient, k, is estimated to be about 0.6, which implies that about 60 percent of the excess demand for money would be satisfied by reserve accumulation within one quarter and over 80 percent of this excess would be met within two quarters. The estimated speed of adjustment of the money market is therefore rather fast, which is consistent with the observation that Singapore's financial markets are highly integrated with the foreign markets.

Although movements in foreign reserves are found to be the principal mechanism for meeting changes in the excess demand for money, the level of these reserves relative to domestic assets is a significant factor

[19] The beginning of the observation period corresponds to the first full year following the total liberalization of exchange controls in June 1978. Thus the structural equations specified in this model may not be suitable for the period before mid-1978.

influencing the level of the exchange rate. The estimated coefficient of this explanatory variable suggests that, as the level of foreign assets increases relative to the domestic assets, the upward pressure on the exchange rate is relieved. At the same time, the estimated coefficient of the term for the inflation rate differential is found to be statistically significant, indicating that the authorities have implicitly taken into account changes in relative prices in managing the exchange rate. The coefficient of the lagged endogenous variable, together with the coefficient of the inflation rate differential term, suggests that about 76 percent of the gap between the desired level and the actual level of the exchange rate is adjusted within a given quarter, indicating a reasonably rapid exchange rate adjustment.[20]

The rate of change in wholesale prices, as expected, is explained almost entirely by changes in import and export prices denominated in the domestic currency, indicating the openness of the economy and the role of the exchange rate in the transmission of inflationary pressures from abroad.[21] The rate of change of consumer prices, however, is influenced more by overall domestic conditions, which suggests that the demand and supply conditions for nontraded goods are important determinants of consumer prices.

The coefficients of the estimated output function have the expected signs and are statistically significant. It is noteworthy that the availability of credit to the private sector, as observed in many developing and newly industrialized countries, is a significant factor for influencing short-term fluctuations in output. Moreover, it is found that the relative unit labor costs adjusted for the exchange rate play an important role in growth performance, as expected for a highly open economy such as Singapore.

The estimated equations for government revenue and expenditure capture the essential characteristics of the budgetary operations. The large and statistically significant elasticity of revenue with respect to nominal income reflects the progressive income tax structure in Singapore. The sum of two estimated coefficients in the expenditure equation is not significantly different from unity.[22] This finding indicates that

[20] The estimates of ϕ and β_2 involve a second-degree polynominal in β_2. The roots of the polynominal are -0.76 and 0.49. Because β_2 is assumed to be positive and less than unity, only the latter root is the meaningful solution. Given $\beta_2 = 0.49$, ϕ is calculated to be 0.76.

[21] Although the sum of the point estimates of w_1 and w_2 exceeds unity, the variance-covariance matrix for these estimates suggests that, with a 95 percent probability, one cannot reject the hypothesis that the sum of w_1 and w_2 equals unity.

[22] The variance-covariance matrix for the two coefficients in the government expenditure equation indicates that, with 95 percent probability, one cannot reject the hypothesis that the sum of the coefficient of ln (GR) and that of ln $(GE)_{t-1}$ equals unity.

Table 1. *Estimated Model, 1979–86*

Item	Estimated Equation
	Behavioral equations
Foreign reserves[a]	$\Delta\ln(R/P)_t = -4.55 + 1.01\ln(Y)_t - 1.1\,(r_d)_t - 0.81\ln(M/P)_{t-1} - 0.39\,\Delta\ln(D/P)_t$ $(3.4)^{**}\ (4.5)^{**}\qquad (2.2)^{*}\qquad (5.5)^{**}\qquad (5.5)^{**}$ $\bar{R}^2 = 0.73 \quad \text{SEE} = 0.04 \quad \text{DW} = 1.86$
Exchange rate[a]	$\Delta\ln(E)_t = 0.07 - 0.04\ln(R/D)_{t-1} + 0.37\,(\Pi_f - \Pi)_t - 0.28\ln(E)_{t-1}$ $(5.4)^{**}\ (5.2)^{**}\qquad (6.7)^{**}\qquad (3.4)^{**}$ $\bar{R}^2 = 0.76 \quad \text{SEE} = 0.01 \quad \text{DW} = 1.70$
Wholesale prices	$\Delta\ln(P_w)_t = 0.001 + 0.20\,\Delta\ln(P_x)_t + 0.90\,\Delta\ln(P_m)_t$ $(0.1)\quad (2.4)^{*}\qquad (8.2)^{**}$ $\bar{R}^2 = 0.96 \quad \text{SEE} = 0.01 \quad \text{DW} = 1.92$
Consumer prices[a]	$\Delta\ln(P)_t = 0.45 + 0.07\ln(M/P)_{t-1} - 0.09\ln(Y)_t + 0.10\,(r_d)_t + 0.07\,\Delta\ln(P_w)_t - 0.001\,T$ $(1.6)\quad (1.9)\qquad (2.2)^{*}\qquad (1.0)\qquad (1.3)\qquad (1.1)$ $\bar{R}^2 = 0.65 \quad \text{SEE} = 0.01 \quad \text{DW} = 1.97$

Output[a]	$\Delta \ln (Y)_t = -0.01 + 0.35\ \Delta \ln (Y^*)_t - 0.14\ \Delta \ln (E \cdot RULC)_t + 0.82\ \Delta \ln (C_{ps}/P)_t$ $(3.2)^{**}\ (4.4)^{**}\qquad (3.1)^{**}\qquad (7.5)^{**}$ $\bar{R}^2 = 0.78 \qquad SEE = 0.02 \qquad DW = 2.10$
Government revenue[a]	$\ln (GR)_t = -11.90 + 1.47 \ln (P \cdot Y)_t + 0.50\ Dummy^b$ $(5.5)^{**}\ (9.1)^{**}\qquad (3.0)^{**}$ $\bar{R}^2 = 0.78 \qquad SEE = 0.18 \qquad DW = 1.51$
Government expenditure	$\ln (GE)_t = 1.15 + 0.57 \ln (GR)_t + 0.28 \ln (GE)_{t-1} + 0.50\ Dummy^b$ $(1.4)\ \ (4.8)^{**}\qquad (2.5)^{*}\qquad (2.1)^{*}$ $\bar{R}^2 = 0.76 \qquad SEE = 0.19 \qquad DW = 2.28$
	Identities
Money supply	$\Delta M_t = \Delta R_t + \Delta D_t$
Domestic credit	$\Delta D_t = \Delta C_{ps,t} + \Delta CG_t$
Credit to government	$\Delta CG_t = GE_t - GR_t - GNBB_t$

Note: $\bar{R}^2$ is the adjusted coefficient of determination; SEE is the standard error of estimate; DW is the Durbin-Watson statistic; ** and * indicate that the estimated coefficient is significantly different from zero at a critical level of 1 percent and 5 percent, respectively.

[a] Corrected for autocorrelation in residuals by the Cochrane-Orcutt method.

[b] The dummy variable takes a value of unity for the first quarter of 1986, and of zero otherwise, to capture the once-and-for-all change in the accounting entries that resulted from the transaction of land ownership.

attempts have been made to keep the level of expenditure in line with total revenue. The coefficient of adjustment for the expenditure equation is estimated to be in the order of 0.7, implying that about 90 percent of adjustment occurs within two quarters.

III. Simulation Results

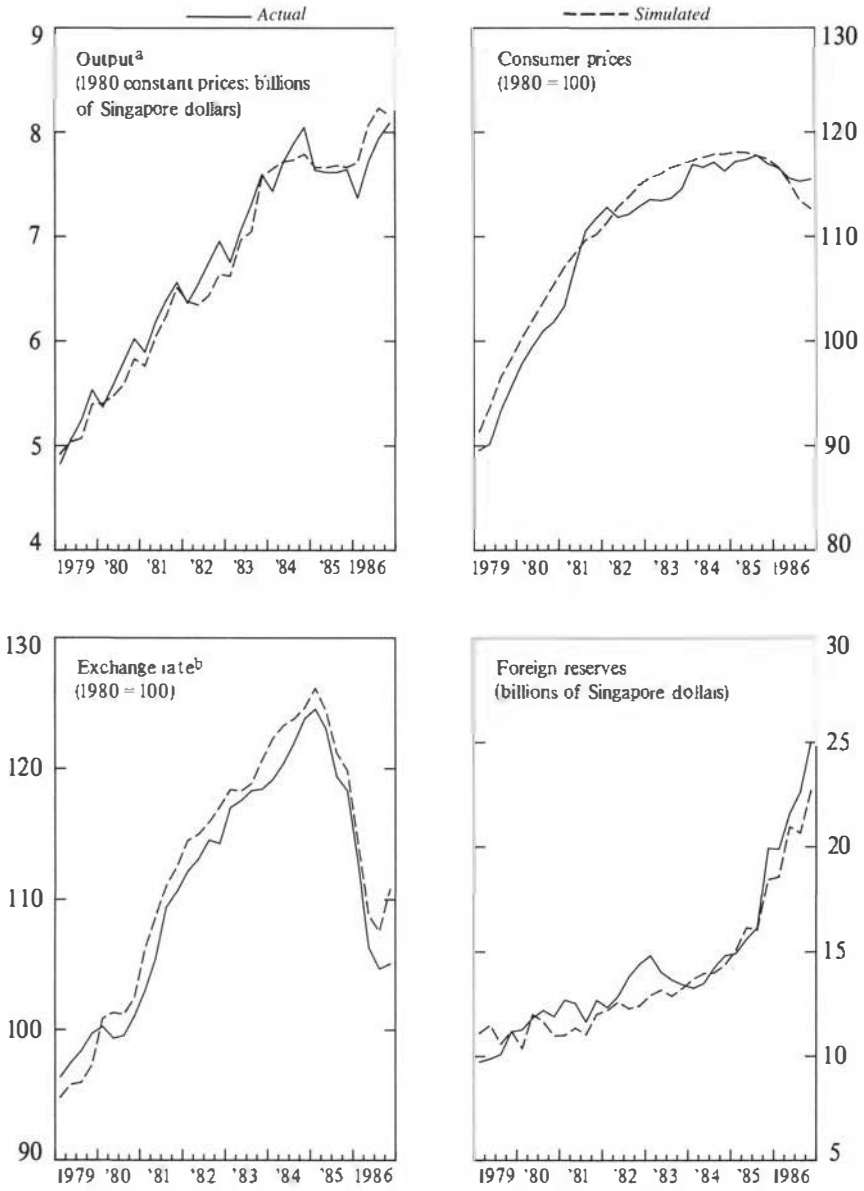
This section analyzes the results of various simulation exercises. The major objectives are, first, to examine the overall explanatory power of the model and, second, to explore the impact of alternative policy scenarios on key macroeconomic variables.

For the purpose of investigating the overall performance of the model, the system of the estimated behavioral and the definitional equations were simulated dynamically during the observation period, 1979–86. The simulation results for the key variables are depicted in Figure 1. The comparison between the actual and the simulated values of real output, consumer prices, the exchange rate, and foreign reserves all suggest that the model as a whole tracks the recent developments rather well, capturing even major turning points—that is, the decline of output in 1985, the decline in consumer prices from mid-1985, and the sharp turnaround in the exchange rate in late 1984. The gaps between the actual and the simulated values for the consumer prices and the exchange rate are serially correlated for certain periods, but they do not seem to be excessive, given the tendency that the errors in a dynamic simulation are accumulated over time and hence are more magnified than in a static simulation.

To examine the effects of policy actions and changes in exogenous factors on key macroeconomic variables, several simulation exercises were conducted.[23] One of the compelling issues facing the Singapore authorities in recent years has been the impact of the high-wage policy, pursued in the first half of the 1980s, on the performance of the economy in terms of output and employment. Although this is a complex issue, a partial answer may be provided by analyzing the results of a simulation exercise based on the assumption that the authorities' wage policy was rather conservative, such that the growth rate of nominal wages adjusted

[23] As suggested by Lucas (1976), however, the parameters of the behavioral relationships could change as the authorities pursue alternative policies. Thus, the results should only be interpreted as indicative of the direction of changes and of the rough magnitudes of the effects.

Figure 1. *Selected Macroeconomic Variables for Singapore, 1979–86*



Note: Simulated values represent the results of dynamic simulation, taking as given the actual values of exogenous variables.

[a]Output measured in terms of real GDP.

[b]An increase indicates an appreciation of the Singapore dollar.

Table 2. *Simulation Results, 1981–86*

(Period average)

Simulation	Simulation Period	Foreign Reserves (In billions of Singapore dollars)	Exchange Rate (1980 = 100)	Consumer Prices (1980 = 100)	Output (In 1980 constant prices; billions of Singapore dollars)
Base run	1981–86	14.7	116.7	114.7	7.2
	1984–86	17.1	118.9	116.7	7.8
Case 1	1981–86	15.6	115.5	114.2	7.4
		(6.1)	(−1.0)	(−0.4)	(2.8)
Case 2	1984–86	15.7	120.8	116.8	7.8
		(−8.2)	(1.6)	(0.1)	(−0.3)
Case 3	1984–86	16.2	119.4	117.2	7.6
		(−5.3)	(0.4)	(0.4)	(−2.6)

Note: Figures in parentheses represent percentage differences from the base-run results. Case 1: relative unit labor cost is fixed at the level of end-1980 for the period 1981–86. Case 2: the government deposits with the Monetary Authority of Singapore (MAS) are reduced by S$0.2 billion in each quarter during 1984–86. Case 3: the quarterly growth rate of world income in real terms is reduced by 1 percentage point.

for changes in labor productivity was kept at the same rate as the weighted average of those in Singapore's major competitor countries. To capture this hypothetical case, it is assumed that the relative unit labor costs (unadjusted for the exchange rate) were fixed at the level of 1980 for the period 1981–86. Thus the assumed level of relative unit labor costs were, on average, about 16 percent below the actual level; all other exogenous variables take the actual values.

The results of this scenario (case 1) and that of the base run (dynamic simulation) are contained in Table 2. According to this simulation, the level of output would have been about 3 percent higher than the base-run outcome, mainly because this wage policy would have contributed directly to the improvement in Singapore's external competitiveness. A higher output level would have enlarged the excess demand for money and increased the need to supply domestic liquidity by purchasing foreign exchange. In turn, foreign reserves would have increased, the exchange rate would have declined, and external competitiveness would have improved. As indicated, compared with the base run, the level of foreign reserves would have been higher by about 6 percent and the nominal exchange rate would have been lower by about 1 percent. Primarily because of the output effect, consumer prices would have been slightly below the result suggested by the base-run simulation.

Another important issue facing the authorities has been the effectiveness of financial and exchange rate policies. To analyze this issue, consider case 2, which represents the hypothetical situation wherein the public sector financial institutions' deposits in the government's accounts with the MAS had been reduced by S$0.2 billion in each quarter. In other words, the equivalent of about 1 percent of total liquidity had been injected in the banking system in each quarter. In this situation, with the excess demand for liquidity being partially alleviated, the need for the monetary authorities to acquire foreign exchange would have been reduced, leading to a reduction in foreign reserves to a level about 8 percent below the base-run result. In the process, the domestic currency would have been allowed to appreciate by nearly 2 percent compared with the base-run outcome. This, in turn, would have contributed to stabilizing prices, but at the same time would have worsened external competitiveness and thus output performance. This scenario exemplifies the severe constraint that the authorities have been experiencing with their conduct of financial and exchange rate policies.

To highlight the impact of external demand conditions, case 3 examines the hypothetical situation wherein the growth rate of world income in real terms had been 1 percentage point less than the actual growth rate, or the average level of that income had been about 6.5 percent

lower than the actual. In this case Singapore's output would have been about 3 percent lower than the base run, indicating that approximately half of the variation in the world economic activity would have been transmitted to Singapore. Because a lower output level would have reduced the demand for liquidity, the level of international reserves would have been about 5 percent below the level suggested by the base-run simulation. Lower foreign reserves would have contributed in turn to an appreciation of the Singapore dollar compared with the base-run result. Despite this appreciation, consumer prices would have risen, mainly as a result of a lower level of output relative to the liquidity position of the economy.

IV. Concluding Remarks

The purpose of this paper has been to examine the impact of financial, exchange rate, and wage policies in Singapore on key macroeconomic variables such as output, prices, and foreign reserves. Toward this end, a simple short-term model was formulated, estimated, and used to conduct several policy simulations. A few conclusions can be drawn from this analysis.

First, the exchange rate policy in Singapore has been largely influenced by the liquidity implications of the government's contractionary budgetary and other financial operations. As a result, there has been a trade-off that the authorities need to consider in their decision to adopt a particular mix of exchange rate and reserve levels.

Second, the authorities' high-wage policy pursued in the first half of the 1980s contributed to rapid increases in the domestic labor costs relative to those of Singapore's competitors. This, together with the sharp appreciation of the Singapore dollar, contributed to a significant loss of external competitiveness and a severe recession in 1985. Had the authorities adopted a more moderate wage policy, the output would have been significantly higher than the actual, and the 1985 recession probably would have been avoided, or at least its severity would have been reduced. Now that the relative unit labor costs adjusted for the exchange rate have declined to a level prevailing in the early 1980s, reflecting both the depreciation of the domestic currency and the sharp decline in wages and other labor costs, growth prospects should improve noticeably over the medium term.

Third, considering the openness of the economy, economic activity has been influenced by the external environment, particularly the level of world income in real terms. The model estimates suggest that financial

and exchange rate policies had relatively limited influence in insulating the domestic economic activity from external shocks, mainly because price stability was also one of the major objectives of the authorities.

The experience of Singapore, albeit unique in many respects, may offer some lessons for other countries with small and highly open economies and no impediments to international capital movements. Singapore's experience suggests that flexibility in financial, exchange rate, and wage policies is crucial in achieving noninflationary growth with external payments viability. Should exchange rate policy to improve external competitiveness jeopardize an inflation target, as is often the case in many open economies, other policy instruments merit consideration. As demonstrated in this paper, an appropriate wage policy may be an important complement to the exchange rate policy. Over the longer run, these policies could help to maintain external competitiveness and to adapt the structure of production to the changing pattern of external demand. The vulnerability of these economies to external developments would thus be lessened.

APPENDIX

Data Sources and Definitions

Data used for estimation were obtained from various issues of four primary sources:

A. International Monetary Fund, *International Financial Statistics* (Washington).
B. International Monetary Fund, *Direction of Trade Statistics* (Washington).
C. Singapore, Department of Statistics, *Monthly Digest of Statistics*.
D. Singapore, Monetary Authority of Singapore, *Monthly Statistical Bulletin*.

Variables were defined as follows (with data sources indicated as above):

R	Net foreign assets of the banking system (source A)
M	Broad money—currency plus demand deposits plus quasi-money (source A)
C_{ps}	Outstanding credit to the private sector by the banking system (source A)
D	Domestic credit (source A)
CG	Net claims on the government (source A)
GE	Government expenditure including net lending (source A)
GR	Government revenue (source A)
$GNBB$	Government nonbank borrowing (source A)
E	Nominal effective exchange rate (calculated from data provided in sources A and B)

P	Consumer price index (source A)
P_w	Wholesale price index (source A)
P_x	Export price index (source A)
P_m	Import price index (source A)
$RULC$	Ratio of Singapore's unit labor costs to that of its major trading partners and competitors (calculated from data provided in sources A, B, and C)
r_d	Three-month interbank deposit rate (source D)
Π	Percentage change in P (source A)
Π_f	Percentage change in the arithmetic average of export and import prices in foreign currency terms (source A)
Y	Real GDP (source C)
Y^*	Weighted average of real GDP or gross national product (GNP) in Singapore's major trading partner countries (source A).

REFERENCES

Aghevli, Bijan B., and Mohsin S. Khan, "Inflationary Finance and the Dynamics of Inflation: Indonesia, 1951–72," *American Economic Review* (Nashville, Tenn.), Vol. 67 (June 1977), pp. 390–403.

Branson, William H., "Monetary Stability and Exchange Rate Objectives in Singapore," in *Papers on Monetary Economics* (Singapore: Singapore University Press for the Monetary Authority of Singapore, 1981), pp. 112–29.

———, and James P. Love, "The Real Exchange Rate, Employment and Output in Manufacturing in the U.S. and Japan," NBER Working Paper 2491 (Cambridge, Mass.: National Bureau of Economic Research, January 1988).

Edwards, Sebastian, and Mohsin S. Khan, "Interest Rate Determination in Developing Countries: A Conceptual Framework," *Staff Papers,* International Monetary Fund (Washington), Vol. 32 (September 1985), pp. 377–403.

Frenkel, Jacob A., and Harry G. Johnson, eds., *The Monetary Approach to the Balance of Payments* (London: Allen & Unwin, 1976).

International Monetary Fund, *The Monetary Approach to the Balance of Payments* (Washington, 1977).

Kapur, Basant K., "Alternative Stabilization Policies for Less-Developed Economies," *Journal of Political Economy* (Chicago), Vol. 84 (August 1976), pp. 777–95.

Keller, Peter M., "Implications of Credit Policies for Output and the Balance of Payments," *Staff Papers,* International Monetary Fund (Washington), Vol. 27 (September 1980), pp. 451–77.

Khan, Mohsin S., "The Dynamics of Money Demand and Monetary Policy in Singapore," in *Papers on Monetary Economics* (Singapore: Singapore University Press for the Monetary Authority of Singapore, 1981), pp. 46–76.

Lucas, Robert E., Jr., "Econometric Policy Evaluation: A Critique," in *The Phillips Curve and Labor Markets,* Vol. 1 of the Carnegie-Rochester Conference Series on Public Policy, ed. by Karl Brunner and A.H. Meltzer (Amsterdam: North-Holland; New York: Elsevier, 1976).

Otani, Ichiro, and Yung C. Park, "A Monetary Model of the Korean Economy," *Staff Papers*, International Monetary Fund (Washington), Vol. 23 (March 1976), pp. 164–99.

Sassanpour, Cyrus, and Jeffrey R. Sheen, "An Empirical Analysis of the Effects of Monetary Disequilibria in Open Economies," *Journal of Monetary Economics* (Amsterdam), Vol. 13 (January 1984), pp. 127–63.

Singapore, Monetary Authority of Singapore, *Papers on Monetary Economics* (Singapore, 1981).

Social Security Issues in Developing Countries

The Latin American Experience

G.A. MACKENZIE*

The major economic, financial, and administrative issues that confront social security systems in Latin America are surveyed. The larger systems have contributed substantially to public sector financial disequilibria. Expenditures of the systems with more limited coverage could increase dramatically as the pension plan matures, life expectancy increases, and coverage is broadened, but the narrow revenue base will force a trade-off between broader coverage and the generosity of benefits. Most plans are pay-as-you-go, and the case for full or partial funding is not found to be compelling. The inflationary environment can have a substantial effect on the financial balance, even under full indexation.

SOCIAL SECURITY systems in Latin American countries mobilize significant quantities of resources.[1] In some countries, such as Argentina, Costa Rica, and Uruguay, the revenue and expenditure of social security systems account for a large share of general government revenue

* Mr. Mackenzie, Senior Economist in the Government Expenditure Analysis Division of the Fiscal Affairs Department, holds degrees from Dalhousie University, Canada, and Oxford University. He has published articles on public financial and macroeconomic issues. In addition to many colleagues at the Fund, the author thanks Miguel Urrutia and Professor Carmelo Mesa-Lago for their comments on this study.

[1] The International Labor Office (ILO) includes in its coverage of social security the schemes or services that have the following three objectives: the provision of curative or preventive health care; the maintenance of income in the case of involuntary loss of all or a large part of income, including income lost as a result of retirement; and the granting of a supplemental income to persons having family responsibilities. Benefits can be received under the following nine categories: medical care, sickness, unemployment, old-age, employment injury,

and expenditure, and their share of gross domestic product (GDP) approximates the shares of some member countries of the Organization for Economic Cooperation and Development (OECD). The coverage of the social security system in these countries is broad, and the level of benefits is generous. In most of the other countries of the region, social security systems do not represent as large a share of GDP but have the potential to do so, because coverage in general is low and could be expanded and because the pension component of the system is not yet mature.

Many problems and issues of financial, economic, and administrative policy arise in connection with the functioning of social security systems in Latin America; this paper surveys the most important of these issues. Section I describes some basic features of Latin American social security systems and prepares the ground for the more extensive Section II, which surveys the principal financial, economic, and administrative issues facing these systems. Section III presents conclusions and summarizes the paper's findings.

I. Basic Features

The relative size and degree of development of the social security system vary substantially among Latin American countries. In Argentina, Chile, Costa Rica, and Uruguay, the ratio of social security expenditure to GDP approached or exceeded 10 percent and accounted for between 25 percent and 40 percent of general government expenditure in 1978–83. At the other extreme, in Guatemala and other Central American countries, the ratio of total social security expenditures to GDP has been on the order of 2 percent or less (Table 1).

Each system's size can be related to different features and reflects important differences in the economic structure of each country as well as the political environment.[2] Of particular importance is the degree of coverage of the population (Table 2). For example, in Uruguay, the pension scheme is estimated to cover about 81 percent of the labor force (economically active population), and the sickness-maternity system about 68 percent. By contrast, the corresponding figures for Guatemala are 33 percent and 14 percent, and coverage is even lower in some other Central American countries.[3] Differences across countries in the number

family, maternity, invalidity, and survivor's benefit. These definitions were used to compile the data that underlie Table 1. See ILO (1985, pp. 1–2).

[2] This section draws heavily on Mesa-Lago (1985). The main results of this study are summarized in Mesa-Lago (1986a, pp. 127–52, and 1986b).

[3] The estimates of coverage are uncertain for several countries for a variety of reasons, including the lack of accurate records of the number of dependents and some duplication of coverage. See Mesa-Lago (1985, pp. 10–13, and Table 2, p. 270).

Table 1. *Total Expenditure of Social Security System as a Percentage of GDP in Selected Latin American Countries, 1975 and 1978–83*

Country	1975	1978	1979	1980	1981	1982	1983
Argentina	6.4	7.5	8.1	9.6	10.1	7.9	8.6
Bolivia	2.1	2.7	3.2	3.0	3.4	2.3	2.1
Brazil	...	...	...	...	6.0	6.4	5.7
Colombia	2.2	2.7	2.7	2.8	2.2	2.1	2.2
Costa Rica	4.2	6.0	6.7	7.1	7.2	5.9	6.3
Chile	10.6	10.6	10.4	13.2	10.5	15.7	14.4
Ecuador	...	2.3	2.4	3.5	3.7	3.5	4.2
El Salvador	2.3	1.2	1.3	1.7	1.8	1.8	1.8
Guatemala	1.2	1.1	1.1	1.2	1.0	1.0	1.0
Mexico	...	...	...	2.7	2.9	3.0	2.8
Nicaragua	2.6	2.4	2.5	2.3	1.1	1.3	1.2
Panama	6.5	6.9	6.7	6.1	6.3	6.4	7.8
Uruguay	9.7	9.1	7.7	9.9	10.9	13.1	11.0
Venezuela	1.2	1.3	1.1	1.3	1.4	1.4	1.5

Sources: International Labor Office (ILO) (1985, 1988); International Monetary Fund, *International Financial Statistics (IFS)*; and Fund staff calculations.

Table 2. *Coverage of Social Security Programs in Selected Latin American Countries, 1980*

Country	Sickness-Maternity (percent of total population)	Pensions (percent of economically active population)
Argentina	78.9	69.1
Bolivia	25.4	18.5
Brazil	96.3	95.6
Colombia	11.6	22.4
Costa Rica	76.0	68.3
Chile	67.3	61.7
Ecuador	7.9	23.2
El Salvador	6.2	11.6
Guatemala	14.2	33.1
Mexico	53.4	42.0
Nicaragua	9.1	18.9
Panama	50.3	45.6
Uruguay	68.5	81.2
Venezuela	45.2	49.8

Source: Mesa-Lago (1985).

of risks covered and the relative generosity of benefits per contributor are of lesser importance. All Latin American countries have old-age pension schemes, sickness-maternity health plans, and disability schemes. Some countries have no family allowance systems, and only a few have unemployment plans (Mesa-Lago (1985, p. 10)).

The size of social security is also related to its importance in total expenditure of the pension scheme; thus, the pension scheme accounted for 80 percent of total benefit expenditure in Uruguay in 1980, compared with 5 percent in Guatemala. An exception to this relationship is Costa Rica, where the pension scheme, although it has grown in importance, accounted for only 15 percent of total benefit expenditure in 1980.

The relationship between the size of the whole system and the importance of the pension component is probably attributable to the degree of maturity of the pension scheme. The pension schemes of Uruguay, Chile, and Argentina have been in place longer than those of the Central American countries, so that the share of the population at or above retirement age that would be eligible for a pension is greater in the former group of countries than in the latter. Finally, the older systems often permitted retirement at an early age and after a short period of participation in the plan.

The substantial variation in coverage among Latin American countries is related to differences in the size of the organized sector of the economy. In countries where coverage is low, a substantial share of the labor force will typically be found in the unorganized rural sector, with a high proportion of either self-employed or unpaid family workers. These groups are difficult to include in conventional social security programs financed through salary-based employer-employee contributions.

Those countries where the social security programs are the largest in relation to GDP introduced their programs much earlier than other countries. An important feature of certain of these countries' programs is the degree of stratification of the system. Certain occupational groups participated in social security programs early on, and other groups were gradually included, *not* in the original plans but in entirely different and separately administered schemes. Substantial inequalities existed across plans, and these have been the object of reform in both Chile and Uruguay in recent years.

Social security programs in Latin America, as in most countries, are financed mainly through payroll taxes. Typically, the employer's share is larger than that of the employee. Other taxes are an important source of finance in several countries; for example, in Argentina, a substantial share of revenues from the value-added tax (VAT) is earmarked for the social security system. In addition, transfers from the central govern-

ment or other levels of government are important in Chile and Uruguay. Neither of these countries can finance social security expenditures from payroll taxes alone, and without other sources of revenue the revenue shortfall would be substantial. The share in total revenue of income from capital is in general low, although it contributes more to revenues in some Central American countries where the pension schemes are relatively young and have been able to accumulate reserves.

Social security expenditures relative to GDP in the group of Latin American countries in the 1970s and early 1980s did not evolve uniformly, although expenditures as a percentage of GDP in the countries with larger systems showed some tendency to increase (Table 1). Taking a somewhat longer perspective, the expenditure ratio has risen since the 1960s. One motive for this secular trend has been the substantial increase in life expectancy that has taken place since 1960 in most countries, which has increased both pension expenditures and the demand for medical services.

The lack of a uniform pattern also characterizes the evolution of the ratio of revenue to GDP in the 1970s and early 1980s (Table 3). Revenue increased in Costa Rica, where the system was still expanding at a significant rate, and in Panama, but it declined in Uruguay. In view of the lack of a pronounced trend in either revenue or expenditure, it is not surprising that the deficit did not evolve in the same direction in most countries. Nonetheless, although comparable data for the period since 1983 are lacking, the available information suggests that in some countries the deficit has tended to grow.[4]

II. Issues

There are several principal financial, economic, and administrative issues that must be considered in a survey of these systems.

Pay-As-You-Go Versus Funded Pension Systems

The larger pension plans in Latin America are now financed on a pay-as-you-go basis, in which current contributions pay for current benefits, with any shortfall being met either out of reserves or by means of a transfer from other government resources. The original pension plans

[4] An earlier version of the paper contains an appendix discussing developments in Uruguay and is available from the author.

Table 3. *Total Revenue of Social Security System as a Percentage of GDP in Selected Latin American Countries, 1975 and 1978–83*

Country	1975	1978	1979	1980	1981	1982	1983
Argentina	7.1	7.8	8.1	9.2	7.2	5.7	6.4
Bolivia	2.4	3.2	3.4	3.0	4.1	2.8	2.2
Brazil	...	...	...	6.5	5.2	6.5	5.3
Colombia	2.2	2.5	2.7	2.8	2.3	2.1	2.3
Costa Rica	5.7	7.3	7.9	8.1	8.1	6.7	8.8
Chile	9.6	9.0	8.3	11.2	8.1	8.0	8.7
Ecuador	...	4.1	4.5	5.2	5.0	5.0	5.4
El Salvador	1.9	1.7	1.9	2.1	2.4	2.5	2.5
Guatemala	1.2	1.6	1.6	1.6	1.4	1.4	1.3
Mexico	...	...	...	3.1	3.4	3.6	2.9
Nicaragua	2.6	2.5	3.1	3.3	1.8	1.9	2.1
Panama	7.7	7.9	8.3	8.0	9.0	9.5	10.1
Uruguay	8.5	8.6	7.4	7.7	7.1	7.2	6.7
Venezuela	1.5	1.6	1.3	1.6	1.6	1.6	1.5

Source: ILO (1985, 1988); International Monetary Fund, *IFS,* and Fund staff calculations.

Note: Central government contributions are excluded from revenues.

were funded at the outset but suffered a decapitalization of their reserves that created strong pressure for the adoption of pay-as-you-go financing. One reason for this decapitalization was the erosion of the real value of financial investments, mainly government bonds, that took place when nominal returns failed to keep pace with inflation. Another reason was the extension of social security systems to encompass groups that were poor actuarial risks, and the introduction of benefits that could not be financed by payroll contributions from the intended beneficiaries (for example, minimum pensions for low-income workers). Finally, a substantial share of reserves was allocated to public investment projects—public housing, for example—where the real rate of return proved to be low or even negative.[5]

Would there be advantages for countries with well-developed systems, such as Uruguay, to convert their pension systems to funded systems? Should the countries whose systems are just now expanding and maturing eschew the pay-as-you-go method? Many economists have argued that pay-as-you-go plans depress saving and capital formation, but the extensive debate on this issue is inconclusive.[6] Nevertheless, among

[5]These points are made in a discussion of investment policy in Chile's social security system in Wallich (1983).

[6]Boskin (1986, p. 80) who argues that the pay-as-you-go system in the United States has depressed private saving and capital formation, has nonetheless stated

other considerations, the adoption of a pay-as-you-go system can obscure the long-run financial implications of a pension program, particularly when a country's demographic structure is changing.

The government that introduces a pension scheme commits itself to expenditures both now and in the long run. If the pension scheme is financed on a pay-as-you-go basis, and if an increase in the dependency ratio is projected, then contribution rates will have to increase to maintain the average pension as a given proportion of the average wage.[7] If a pension program relates the size of the pension to the number of years in which participants make contributions, expenditures can be expected to rise rapidly in the initial years of the program.

An advantage of a fully funded system is that the possible unpleasant surprise of an increase in the contribution rate is avoided. A related argument is that if it is difficult to change social security contribution rates once they are set, then they should be set high enough initially to generate sufficient resources for the expenditure programs over a long period. To put it another way, a funded system conveys the right kinds of signals about the future costs of a pension program.[8] The relevance of these arguments depends on the rapidity with which changes in rates are required.

An important consideration in the Latin American setting is the impact of inflation on a funded system. A high and variable rate of inflation will substantially increase the variance of the rate of return to the fund's reserves unless its investments are indexed. Furthermore, an unexpected surge in inflation could wipe out the real value of the reserves. The problem of inflation is mitigated by the development of sophisticated financial instruments and may not be a serious problem in a stable financial environment. Nonetheless, a pay-as-you-go system is not vulnerable to inflation in the same way.

---

that the "evidence is far from conclusive." Aaron (1982, p. 51) states that "using the best that economic theory and statistical techniques have to offer, [economists] have produced a series of studies that can be selectively cited by the true believers of conflicting hunches [about the effect of an unfunded social security system on saving] or by people with political agendas that they seek to advance." This work surveys the empirical problems involved in estimating the impact of social security on saving. A survey of empirical studies and a tabulation of their conclusions may be found in Break (1981).

[7]The dependency ratio is the ratio of the number of pensioners (P) to the number of contributors (C). The total value of pensions in a given period is P times the average value A. Under a pay-as-you-go system, $twC = PA$, where t is the contribution rate and w the average wage. If A/w is a constant, then t varies directly with P/C, the dependency ratio.

[8]This argument is discussed in Halter and Hemming (1987).

These various considerations do not constitute a strong case for a funded system in most Latin American countries. The argument that a pay-as-you-go system depresses saving is irrelevant if confidence in the pension system is low, because consumption will not be stimulated if individuals do not expect to have their contributions returned to them in the form of pensions. In any case, pay-as-you-go systems have proved to be politically attractive.

If a country opts for a fully or partially funded system but uses the system's reserves simply to finance additional current expenditure by the government, any possible advantages of the choice are lost. The appropriate policy for the government to adopt will depend, however, on how the private sector reacts to a funded scheme. If it views its contributions as saving, albeit forced, then the social security system's savings could substitute for other forms of saving, leaving private consumption unchanged. In this case, any additional expenditure financed by the surplus has an expansionary impact on the economy. Unless the stance of fiscal policy before the introduction of the program was overly tight, this use of the surplus would be inappropriate policy. If the stance was right, then the central government should act as if the surplus were not available to finance expenditure increases or tax reductions. One appealing option would be to invest the reserves in the central government debt and thereby reduce the deficit of the consolidated central government.[9]

If contributions to a funded system are viewed as a tax, then the introduction of the program should give rise to some offsetting reduction in taxation or increase in expenditure unless the stance of fiscal policy before the introduction of the program was too loose. If the rate of investment in the economy is deemed to be too low, the reserves should finance public investments chosen for their social rate of return.

Sources of Financing

Payroll taxation has traditionally been viewed as an equitable means of financing social security because it relates an individual's cost of

[9] For example, if the deficit of the central government before consolidating the financial operations of the social security system is 5 percent of GDP and a funded social security program is introduced that initially generates a surplus equal to 2 percent of GDP, the surplus could be invested in government bonds, on the assumption that the government's operations are unaffected by the presence of the social security system, so that the deficit of the consolidated central government declines to 3 percent of GDP. In a closed economy, the private sector would have a surplus of 5 percent before the introduction of the social security system and a surplus of 3 percent afterward.

participation in social security to the benefits he will receive in a manner analogous to the relationship between the value of premiums in a private pension plan and average benefits. This view, however, presupposes that the effective incidence of payroll taxation is on the insured; that is, that the introduction or increase in a payroll tax does not affect employers' labor costs but reduces employees' pay net of both the employer and employee portions of the tax. In any case, the relationship between the value of contributions made by, or on behalf of, different individuals and the expected value of future benefits can be tenuous. For example, often when social security programs are introduced they are immediately extended to the current generation of retirees, who cannot have made any contributions.

A related argument in favor of using payroll taxes to finance social security, either in part or in full, is that their use imposes financial discipline on the social security system. An undue expansion in benefits will be checked because the burden of the contributions is effectively borne by the contributors. Finally, the use of the payroll tax has been justified on the grounds of administrative ease.

In many Latin American countries, however, exclusive reliance on payroll taxation is simply not feasible for a social security system that aspires to broad coverage. The base of a payroll tax is effectively limited to the organized sector, and, as has been noted, this sector is often relatively small.

The assumption that the incidence of the payroll tax falls effectively on the insured is also open to question. Because of institutional rigidities in wage setting and oligopolistic market structures in the organized sector, the tax could simply be passed on to consumers. An example of institutional rigidity is the regulation in Mexico's social security laws that employers of workers receiving the minimum wage are obliged to pay both the employer's and the employee's share of payroll taxes (see Wilson (1985)). A study of social security and other government programs in Chile revealed that prices were determined by average cost plus a markup, suggesting that payroll costs are simply passed on to the consumer (cited in Wilson (1985, p. 262)).

If the payroll tax is not effectively borne by the contributor, then it must increase the cost of labor. In turn, this will have some impact on employment, on the assumption that labor and other factors of production are to some extent substitutable. The shift away from payroll taxes to value-added taxation that took place in Uruguay and Argentina in 1978–79 was in part prompted by the view that high rates of payroll taxation were depressing the level of employment. A recent study calculated a range of estimates for the impact on employment in Mexican

manufacturing that would be generated by a switch from payroll to value-added taxation. This impact ranged from 1.7 percent to 12.5 percent (Wilson (1985, pp. 267–69)).

If the burden of the payroll tax does not fall on those who benefit from the tax, then there is no reason, on purely economic grounds, that social security programs should not be financed from general revenues. Moreover, there may be a case for reducing payroll tax rates if an alternative financing source is available.

Financial Implications of Demographic Trends

The social security systems of OECD countries have had to confront the financial implications of an increase in the dependency ratio brought about by the increased life expectancy of older persons, the declining birthrate, and the systems' maturation. Similar developments are evident in Uruguay, and declining mortality rates in Chile and Costa Rica have contributed to an increase in the dependency ratio in these countries as well. Further increases in the dependency ratio in these countries could create serious financial pressures because pension expenditures are already a significant share of GDP. If life expectancy increases rapidly in those countries where it is now low, a substantial increase in pension expenditure is likely to result. At the same time, the aging of the population could require a substantial increase in medical expenditure for the treatment of the degenerative diseases associated with old age. In consequence, it is useful to consider what would happen to the composition of the population—and to the dependency ratio—if the present trends of a falling birthrate and an increasing life expectancy were to continue.

The World Bank's population projections for virtually all countries are based on the assumption that these trends will continue.[10] By computing the ratio of persons at or above a standard retirement age to the number of persons of working age, a crude proxy for the dependency ratio may be obtained. The ratio thus derived cannot take into account changes in the rate of labor force participation, but it gives a rough indication of the impact of demographic change on the financial requirements of a pension system.

The base projections of the World Bank do not show any great variation in the ratio of the retirement-age to working-age populations between 1980 and 2000 for any of the Latin American countries, with the

[10]The 1985 projections are contained in Vu (1985).

Table 4. *Projections of Dependency Ratio in Selected Latin American Countries, 1980–2030*

(In percent)

Country	1980	1990	2000	2010	2020	2030
Argentina	24.1	26.3	26.0	26.1	28.6	30.7
Bolivia	12.7	13.2	12.9	13.0	14.0	16.3
Brazil	13.8	13.8	14.6	16.4	21.5	27.9
Chile	16.7	17.2	19.0	22.8	29.3	36.3
Colombia	12.9	12.7	13.2	15.5	21.1	29.7
Costa Rica	12.6	12.9	14.2	16.6	23.9	33.2
Ecuador	13.4	12.3	12.0	12.5	15.4	20.3
El Salvador	13.1	12.6	11.8	11.7	14.3	18.9
Guatemala	11.1	11.5	11.8	12.5	15.3	19.4
Mexico	13.3	11.9	11.9	13.0	16.6	23.2
Nicaragua	10.5	10.7	10.3	10.3	12.7	16.2
Panama	14.8	14.5	14.8	17.3	22.7	30.4
Uruguay	29.5	32.9	33.9	32.3	33.5	38.7
Venezuela	10.6	11.4	12.3	14.2	19.4	25.5
Average[a]	14.5	14.3	14.6	16.0	20.3	26.2

Sources: Vu (1985), and Fund staff calculations.

Note: The dependency ratio is calculated by expressing the population aged 60 or more as a percentage of the population aged between 20 and 59.

[a] Calculated by weighting each country's dependency ratio by its share of the group's total projected population.

possible exceptions of Chile and Uruguay (Table 4). Thus, increases in the rate of growth of the elderly population, owing to increased longevity, and decreases in the rate of growth of the working-age population, stemming from a falling birthrate, would by themselves have little impact on the dependency ratio in the foreseeable future. Beyond the year 2000, however, the impact of the projected decline in the birthrate on the size of the working-age population begins to make itself felt. The increase is particularly marked in some of the Central American countries, as well as in Colombia and Venezuela. Moreover, the combination of a relatively rapid increase in the elderly population of these countries and the extension of coverage to the bulk of the population could cause expenditures to increase enormously.

Alternatives to Public Pension and Health Insurance Systems and the Experience of Chile

In 1979, Chile implemented reforms to its pension system that were intended to replace the old public pension system with what could be termed a compulsory private sector savings plan, under which the state's

role would be reduced to one of regulation. Under the new scheme the employer's contribution has been eliminated, and individuals are required to deposit 10 percent of their salary in one of a group of eligible financial institutions—the *Administradoras de Fondos de Pensiones,* or AFPs—that the individuals can choose themselves. These funds, which are required to specialize in the management of the savings entrusted to them, restrict their investments to a range of financial securities.[11]

Each participant in the savings plan has a personal account with his chosen fund, and his share of the net earnings of the fund—its investment income net of commissions—accrues to that account. Pensions are determined by the capitalized value of each person's contributions to the plan; at retirement, each participant may choose to invest the accumulated value of his contributions and their earnings in a life annuity purchased from an insurance company. The annuity is expressed in real terms—*unidades de fomento*—which are linked to the consumer price index.[12] The reforms also provided for a minimum pension for persons with at least twenty years of work experience. If the accumulated capital in an individual's account is insufficient to provide this minimum pension, the government makes up the difference.[13]

The measures implemented in 1979 included a provision that allowed Chileans covered by the old public system to switch to the new scheme at any time between 1981 and 1986. Subsequently, the deadline for transfer was extended indefinitely. Persons opting for the new system would have their accounts in the new scheme credited at the time they retired by an amount related to their accrued contributions under the old public system.[14] After December 1982, all new entrants to the labor market except the self-employed, whose participation was voluntary, were required to join the new scheme. To foster competition, and hence more efficient management, participants in the new scheme were allowed to switch from one fund to another.

The reforms also included increases in the retirement age, to 60 for women and 65 for men, which strengthened the finances of the old system. Its contribution rates, however, were not changed. Because the

[11] A discussion of various aspects of the new scheme can be found in Arellano (1985, Chapter 3).

[12] Individuals retiring under the new system may also pick the programmed retirement option, which entitles them to monthly payments directly from the AFP. This second option does not guarantee a constant real payment to the pensioner, as the first option does. The Chilean system is similar in some respects to a proposed reform of the U.S. social security system. See Boskin, Kotlikoff, and Shoven (1985).

[13] An additional obligatory deduction of about 3.5 percent is used to purchase disability and survivor's insurance from an insurance company.

[14] At an individual's retirement, his account is credited with a *bono de reconocimiento,* a "recognition bond," by the government.

percentage of income involuntarily saved under the new system did reflect the new retirement ages, it was less than the contribution rate of the old system, and a substantial incentive to switch to the new system was created. A majority of contributors in the old system switched to the new.[15]

The government plays an important regulatory role in the new system by overseeing the performance and functioning of the private funds, but it also ensures that their contributors receive a minimum rate of return. This rate of return is defined as 50 percent of the average rate of return for all the funds or the average rate less 2 percent, whichever is less.

The method of determining an individual contributor's pension differs radically between the two systems. Under the old system, an individual's pension was determined by the number of years he contributed and his average salary in the period immediately before retirement. Pensions were adjusted from time to time but were not subject to any automatic indexation rule, and their real value fluctuated substantially in the twenty years before the reform was instituted (Arellano (1985, p. 79)).

Under the new system, the value of a pension will be determined by a contributor's savings and the average rate of return of his fund, subject to a minimum rate of return determined by the government. Consequently, there is no guarantee that pensions will bear any given relationship to a contributor's earnings, nor that a pension will maintain its value in real terms over the retirement period. The funds conceivably may not be able to achieve positive real rates of return, so the government might be compelled to intervene.

The more stable the financial environment, the more successful management of the funds is likely to be, and the existence of the new system is also likely to create additional incentives for just such an environment. Nonetheless, although 1981–86 was not a stable period, the funds earned rates of return that were high in real terms.

Earlier in Section II, a funded system was compared with a pay-as-you-go system as a generator of savings, but neither system was found to be conclusively superior to the other. Many economists would conclude that Chile's new system should be preferred because of its ability to generate savings. Nonetheless, it remains critical that the contributions to funds not be used to finance additional current expenditure by the government.

[15] Arellano (1985) estimates that the number of persons contributing to the old system declined from 1,695,000 in 1980 to 449,000 in 1983, and that the number of individuals contributing to the new system reached 1,106,000 by the end of 1984 (Arellano (1985, p. 145)). Mesa-Lago has estimated that 83 percent of insured persons was in the new system at the end of 1986 (Mesa-Lago, personal correspondence with the author (1987)).

Table 5. *Social Security Administration Expenditures as a Percentage of Total Expenditure in Selected Latin American and OECD Countries, 1980*

Latin America		OECD	
Country	Percentage	Country	Percentage
Argentina	4.3	Belgium	4.1
Bolivia	19.3	Canada	2.5[a]
Brazil[b]	9.9	Denmark	2.7
Colombia	12.4	France	3.9
Costa Rica	6.9	Germany, Fed. Rep. of	3.1
Chile	6.2	Greece	4.5
Ecuador	23.7	Ireland	4.7
El Salvador	14.9	Italy	4.3
Guatemala	12.1	Japan	2.0[a]
Mexico	13.3	Netherlands	3.4
Nicaragua	11.6	Norway	2.1
Panama	5.3	Portugal	9.9
Uruguay	6.4	Spain	2.7
Venezuela	13.8	Sweden	2.5
		Turkey	4.1
		United Kingdom	2.8[a]
		United States	3.1

Source: ILO (1985, 1988).
[a] 1979–80.
[b] 1981.

The new system has been touted as more efficient than the old, although the existence of substantial economies of scale in the administration of public savings plans could mean that a unified public system would be potentially less costly than the new privatized system. The old system, however, was far from unified, and in 1980 administrative expenditure was estimated at about 6 percent of total expenditure by the Chilean social security system, a high figure by the standards of industrial countries but below average for Latin American countries (Table 5; see also "Administrative Issues," below).[16] Private sector operations would presumably have more incentive to minimize their operating costs than would a public bureaucracy, and the government's guarantee of a minimum rate of return to the contributors is not a guarantee that it will rescue a fund in financial difficulty. Hence, the regulation of the industry does not appear to create a disincentive to cost minimization.

Nonetheless, it is not clear that the new system is less costly than the old. One observer has drawn attention to the fund's apparently high level

[16]ILO (1988, Table 1). The average for the 14 countries in Latin America shown in Table 5 is 11 percent.

of promotional and advertising expenditures in its first few years of operation.[17] In addition, the coexistence of the two systems must substantially increase the amount of resources devoted to the administration of savings plans.

The switch to a regulated private and funded system has had the effect of substantially increasing the deficit of the financial operations of the consolidated central government because revenues of the public social security system fell with the transfer of contributors to the private system—although the old system still had to pay the pensions of persons who had retired under that system—and because the central government now had to pay the *bono de reconocimiento* of each transferee to the new system as he retired.[18]

Should the increase in the deficit resulting from the reforms, which has been estimated at about 5 percent of GDP in 1983, be regarded as an expansionary shift in the stance of fiscal policy (Yáñez (1985))? To the extent that payroll tax contributions to the public social security system are simply replaced by contributions to a private and compulsory savings plan, while the public social security system goes on making pension payments to retired contributors to the system, the increase in the deficit would not be expansionary. The private funds could invest their revenue in public debt, and the increase in the public sector deficit would be offset by an increase in the savings of the private sector, with no increase in expenditure either for investment or consumption. Only if the savings entrusted to the funds leads to an increase in expenditure—for example, by being invested in private securities that give rise to an increase in investment—would the increase in the deficit be associated with an expansionary shift in the fiscal stance.

The reforms of 1980–81 also affected the health and medical component of the social security system. Under the previous system public hospitals, whose services were largely free and which catered mainly to the blue-collar worker, existed side by side with a voucher system, whose participants could choose among a variety of health care providers. The social security system financed 50–70 percent of the cost of the vouchers, with users financing the remainder; the social security system's costs

[17] Arellano (1985, p. 170) has estimated that marketing and sales costs amounted to 39.8 percent of total operating costs exclusive of depreciation and amortization in 1982, and 30 percent in 1983. Expenditures in 1981 were even higher because the sales campaigns were launched in that year.

[18] Another measure taken in the reform, but not discussed previously, that had the effect of increasing the deficit was the reduction in the rate of contribution for the family allowance and *cesantia* (early retirement subsidy) component of social security, which continued to be publicly administered.

were financed through contributions of 4 percent of workers' taxable earned income. (Foxley, Aninat, and Arellano (1979, p. 106)).

The new system is also financed by a payroll tax, whose rate was raised to 6 percent in 1983 and, subsequently, to 7 percent. Individuals who do not wish to rely on the services of public hospitals, however, can opt to become members of providential health institutions (*Instituciones de Salud Previsional,* ISAPRES), which receive the 7 percent contribution from their membership. Each ISAPRES accepts only persons whose annual income is above a certain minimum. At the end of 1984 this minimum varied from 40,000 pesos to 120,000 pesos and was several times the average income of the contributors remaining with the old system.

These changes in the health component of the system have affected considerably fewer people than the reforms made in the pension component. At the end of 1983, some 108,000 persons were registered in some 15 different ISAPRES. The new system has some interesting implications. In particular, it has reduced the revenue of the public system, which no longer receives the contributions of the individuals who registered with the ISAPRES. Health expenditures have also been reduced; however, because persons registered at the ISAPRES were among the better off, they were undoubtedly subsidizing the group of persons remaining with the public system. The result is that the redistributive role of the health component has been lessened. (Arellano (1985, pp. 188–91)).

Effects of Social Security on Income Distribution

Two characteristics of Latin American social security systems are particularly relevant in any assessment of the effects of social security on income distribution: their coverage, which is often quite low, and the financing role played by transfers from the central government, in the form of either earmarked revenues or general budgetary support.

When coverage is not universal, the incidence of the social security system depends critically on the financing source and its incidence.[19] There is typically a correlation between the degree of urbanization and industrialization and the extent of coverage. In general, those excluded from coverage are in the rural and unorganized urban sectors, where the

[19]The incidence of the system as a whole is said to be regressive if the net transfer it creates increases in proportion to income as income increases. Typically, a system will be neither regressive nor progressive over the *entire* range of income.

incidence of extreme poverty is the highest. Those covered often form a relatively privileged group.[20]

Under these circumstances, the incidence of the system could be broadly neutral if social security expenditures were financed entirely by payroll taxes whose effective (regardless of legal) incidence fell on employees, and if an individual's benefits were related to his contributions. As noted previously, however, there is reason to believe that the consumer bears at least part of the burden of the incidence of payroll taxes. Thus, uninsured persons could be paying, in the form of higher prices, for the benefits that accrue to a relatively privileged minority. Among the insured, some transfer of income from one income class to another could also take place. If the system has a minimum pension, then those who are better off subsidize those who are worse off: this provision has a progressive incidence. Conversely, a cap on the absolute value of contributions has a regressive incidence if no similar cap affects pensions.

By contrast, if coverage is universal or at least does not exclude those at the lowest end of the income scale, then the incidence of the payroll tax is less important. If employers pass on the tax in higher prices, the group bearing the effective incidence of the tax, the general population, is also benefiting from the expenditures the tax finances.

When coverage is limited and social security expenditure is financed at least in part out of general revenues, then its effective incidence is likely to be regressive, even if the employed contributors bear the burden of the payroll tax component of the system's revenues. The general revenue component is partially borne by the insured, who benefit from the system's expenditures; the greater burden of general revenues is borne by the uninsured, who do not. But the uninsured are likely to be among the relatively poor. It can be tentatively concluded that the incidence of the social security system is regressive in most of Central America, and also in relatively affluent countries such as Colombia and Venezuela, because coverage in these countries is low.

Two recent studies (cited in Mesa-Lago (1983)) of countries where coverage is relatively broad by the standards of the region have found that the incidence of social security was either neutral or slightly progressive. A study of the Costa Rican system as of 1973 found that, even with 50 percent coverage, the system's incidence was slightly progressive. The progressivity of the incidence of the system's health component, which was found to be distributed quite evenly regardless of income level,

[20]Mesa-Lago (1983) notes that in the countries where social security was pioneered, such as Chile, Argentina, and Uruguay, the first groups to be covered were in general relatively well-off workers in the mining enclave, transport sector, or public sector.

offset the regressivity of a ceiling on income subject to contributions and the regressive impact of less than universal coverage. (The degree of coverage has increased substantially since 1973; see Wilner-Green (1977).) A study of Chile in the early 1970s found that the incidence of the system was broadly neutral for the insured group (Foxley, Aninat, and Arellano (1979)). But for the population as a whole, the incidence could be regressive.

Problems Posed by an Inflationary Environment

An inflationary environment complicates the administration of social security and the social security system itself. As an important component of general government, the social security system can play an important role in sustaining or generating inflationary pressures, or in resisting these pressures, depending on its indexation policies.

Consider first the impact of inflation on the financial balance of a social security system whose sole revenue source is the payroll tax. A revenue system that is neutral with respect to inflation is defined here to mean one in which a change in the rate of inflation does not affect the ratio of revenues to GDP. For neutrality to hold, the following set of conditions is sufficient: first, the built-in elasticity of tax revenues with respect to the tax base—the wage bill—must be unity; second, the elasticity of the wage bill with respect to GDP must also be unity. If the second condition holds, the built-in elasticity of payroll tax revenues with respect to the wage bill will be unity when the tax is a fixed percentage of wages, with no ceiling on contributions and no lag in collection.

Lags in collection reduce the ratio of taxes to their base when the rate of inflation increases because the base is raised by inflation before the yield of the tax is raised (Tanzi (1977)). In practice there will always be some lag, although in the case of the payroll tax it should be relatively short if the degree of contributors' compliance is sufficiently high. However, the inflationary process can create strong incentives for enterprises to delay their tax remittances unless the penalty for doing so varies directly with the rate of inflation and with the period of the delay.

If the penalty for late remittances cannot be adjusted promptly when inflation accelerates, it should be set at a very high level; otherwise, the acceleration of inflation can create serious problems for social security finances, even if the tax is a fixed percentage of wage income and there is no ceiling on taxable income.[21] If the financial position of enterprises

[21] Setting the penalty at a high level can cause problems too: when the rate of inflation is low, a penalty rate high enough to deter late remittances at high rates of inflation will be punitive.

deteriorates along with the acceleration of inflation, the incentive to delay remittances will be all the greater, especially if the compliance of contributors is not strong to begin with. Thus, in some circumstances accelerating inflation can significantly reduce the ratio of social security revenue to GDP.

The ratio of the social security system's expenditure to GDP will depend on the indexation policy the system follows. Assuming that the indexation formula relies on the consumer price index (CPI), that the CPI increases at the same rate as the GDP deflator, and that the elasticity of real expenditures with respect to real GDP is unity, the ratio of expenditure to GDP will vary directly with the frequency of adjustment, inversely with the lag in the adjustment, and inversely with the rate of inflation. In other words, accelerating inflation reduces the ratio more if the adjustment lag is long and if the frequency of adjustment is low. For example, with an initial ratio of expenditure to GDP of 10 percent, an increase in the rate of inflation from zero to 50 percent reduces the expenditure ratio to 8.7 percent even when expenditure is adjusted quarterly with a quarterly lag. When adjustment takes place once instead of four times a year, the ratio falls to 7.5 percent. Unless the adjustment lag is zero and the adjustment is continuous, an increase in the rate of inflation will lower the ratio of expenditure to GDP, even with full indexation.[22]

The choice of an automatic or discretionary approach to the indexation of the pension component of social security expenditures depends to some extent on the index used. A policy of full indexation to the CPI implies that one social group receives special protection from shocks to the economic system that entail a decline in real income; for example, when the economy suffers a substantial terms of trade loss. In such circumstances, this kind of indexation may forestall the adjustment in real incomes necessary to restore external balance, and it is worth adding that pensioners in Latin America can be a privileged group.

In consequence, a policy of indexation to the CPI may be misguided. Indexing pensions on wages would be a better rule for financial stabilization. But could a policy of complete discretion be even better? The argument for discretion is that no single rule would always be appropriate. If real wages are rising at an unsustainably rapid pace, then wage-based indexation would not be appropriate. Yet it may be argued that the long-term functioning of a social security system may require that contributors be able to expect a certain stability in their benefit enti-

[22] An earlier version of the paper contains an appendix with illustrative calculations of the impact of lags and frequency of adjustment on the ratio of pension expenditure to GDP and is available from the author on request.

tlements. The periodic adjustment of pensions in line with wages can thus be seen as the expression of an implicit contract that pensions will normally be related to earned incomes, even if pensioners will not always be protected from declines in their real income.

Administrative Issues

Evasion of payment of social security contributions appears to have been significant in several Latin American countries, although it is difficult to ascertain its extent. It is likely to be more prevalent in the rural and unorganized urban sectors because enterprises in these sectors typically do not keep accounting records and because, given the small size of these enterprises, administrative surveillance is often not cost effective. In addition, evasion is likely to be more prevalent in systems in which there is little or no relation between a participant's contributions (including contributions made on his behalf by his employer) and the benefits he enjoys. For example, the incentive for workers to collude with employers to evade contributions would be greater if pensions were not related to the value of contributions, or if the contributions financed medical care and payments for income lost while workers were sick.[23]

The temptation to evade may increase in periods of recession, and it must certainly increase when the rate of inflation increases if penalties for evasion are a fixed percentage of the value of contributions not remitted. Finally, the rate of the payroll tax itself is related to the incentive to evade.

Clearly, evasion will be a problem if the social security institution cannot command the administrative means to ensure a reasonable degree of compliance. In a system in which benefits do depend on an individual's contributions, the incentive to cheat is probably greater when no adequate records of an individual's contributions are maintained. Keeping such records is imperative, not just to reduce evasion but also to ensure that benefits are fairly evaluated.

Substitution of the VAT for the payroll tax has been seen as a remedy for the problem of compliance because with the credit-invoice system of the VAT there is no incentive to underinvoice at either the wholesale or manufacturing stage of production. The VAT is considerably more difficult to administer than the payroll tax, however, and if it does not fully replace the payroll tax there is no saving from the dismantling of the administrative apparatus of the payroll tax (McClure (1981)).

[23] This relationship assumes that contributors would not discount future income excessively.

Tardy remittances have also been a problem. Because of inadequate penalties for late remittances, in some countries business enterprises have enjoyed a loan from the social security regime at a negative rate of interest. Either periodic adjustments of the penalty for late payment or a constant, very high, penalty rate is necessary to resolve this problem when the rate of inflation is high and variable. Finally, late remittances are tacitly encouraged if the government is late in making its own contributions to the social security system, or in establishing and settling claims to pensions.

The stratification of social security schemes in the countries that first introduced social security must have contributed to substantial administrative inefficiency in the past. Because of the large number of plans, administrative resources were duplicated. Moreover, the lack of coordinated administration was responsible for serious inequities, including the nontransferability of years of participation under one scheme to another and the possibility of an individual's receiving two pensions at excessively generous terms because benefits were not coordinated when contributory service in one plan was transferred from another. The reform measures implemented recently in Chile and Uruguay have substantially increased the unification of those systems. Nonetheless, the process is not yet complete, and there is evidence that some advantages of a unified system have yet to be exploited fully (Mesa-Lago (1985, p. 199)).

Administrative expenditures in Latin American countries are high in relation to those of OECD member countries if the ratio of administrative expenditures to total expenditures is taken as the index of relative cost (Table 5). In Bolivia and Ecuador, this index is high by any standard. Economies of scale in program administration could explain the lower cost of the OECD programs if these programs were in general larger than programs in Latin America. However, Denmark and Norway, both with relatively small programs, have cost indices lower than any Latin American country.[24] Mesa-Lago (1985) has drawn attention to the lavishness of social security administrations' headquarters and has argued that payrolls and benefits of the administrative staff have typically been excessive.

The health expenditure component of social security systems, particularly the medical services component, presents special administrative

[24] A further possible explanation is that some of the capital equipment used by administrative agencies in Latin America may be as expensive as that used by OECD countries, whereas the value of benefits per capita is much lower, given the much lower per capita income levels in Latin America to which benefits are scaled.

problems. In many countries, parallel hospital systems exist: one for social security contributors, another—operated by the ministry of health—for the rest of the population not relying on private clinics. The social security medical program typically covers a minority of the population—mainly blue- and white-collar workers in the cities. The rural population and indigent city dwellers in large part are covered by the system operated by the ministry of health. Despite its smaller coverage, the social security medical care system usually has a larger budget than the medical care component of the ministry of health, and its per capita expenditure is much larger (Ugalde (1985, p. 111)).

In some countries the services of the two institutions have not been coordinated, leading to a considerable duplication of services; an example is two hospitals serving the same community where one would suffice, so that in many countries the occupancy rates of hospital beds have been as low as 50 percent (Mesa-Lago (1985, p. 26)). Efforts to improve the coordination of services have in general met with little success. One source of difficulty has been the preference of contributors to social security for a separate hospital system because of the poorer quality of service in the public system;[25] however, not all countries face this type of problem. In Costa Rica, for example, the Ministry of Health is responsible for public health programs and for medical care in rural areas not covered by the social security system, and the social security system is responsible for the bulk of medical care services. The two services are being gradually integrated.

The concentration of social security's resources in the cities and the financial constraints imposed on the budgets of health ministries imply that a significant proportion of the rural population has limited or no access to modern medical care. These inequalities cannot be eliminated by an extension of the coverage of the social security system at its present level of benefits per capita. Such an extension would require an enormous increase in the financial resources mobilized by the system.

III. Conclusions

The social security systems of some Latin American countries are large enough that their financial imbalances can have serious macroeconomic consequences. These imbalances reflect the influence of the erosion of the payroll tax base because of falling employment, as well as

[25] This is discussed in Chapter 3 of International Social Security Association (1982).

the secular influence of an aging population on the finances of the system's pension component. To forestall the emergence of more serious financial disequilibria, several countries have had to allow a marked erosion in the level of social security expenditure in real terms. This reduction has been accomplished by various means, including less than full and timely indexation of pension benefits and increases in minimum retirement ages.

The social security systems of most Latin American countries now absorb only a small share of GDP, essentially because the coverage of the systems is low, even though expenditure per participant can be quite high. However, as the pension component of these systems matures—as a greater share of the elderly population, having contributed for a longer time, becomes entitled to more benefits—the share of expenditure in GDP can be expected to grow, even without a marked expansion in the systems' coverage. The reported accounting surplus in the pension system in some of these countries most likely disguises an actuarial deficit. Future benefits promised under current legislation cannot be financed out of the system's income from reserves and contributions at current rates of contribution by the system's participants. An additional problem confronts these countries: any significant expansion of the system will not be feasible at the current level of expenditure per participant, so there will be a trade-off between the generosity of benefits and the degree of coverage of the system.

Although demographic forces have contributed to the upward pressure on social security expenditures in the last twenty-five years, their influence should abate in most countries in the near future because no marked increase in the ratio of elderly to working-age population is projected to take place in the next twenty years. In the more distant future, however, a further decline in the birthrate and some further increase in life expectancy could push the dependency ratio up again, putting pressure on pension expenditure. At the same time, the increase in life expectancy and demands for high-technology medical care will create pressure on medical expenses.

Funding Versus Pay-As-You-Go

No country in the region has a fully funded pension system except Chile. Pay-as-you-go systems may reduce national saving and capital formation, but evidence to support this view is not conclusive. Funded systems can be successful in a stable financial environment, but they are vulnerable to both inflationary shocks and misguided investment poli-

cies. Their reserves should not normally be used to finance increases in government expenditure or tax cuts, so that the introduction or expansion of a funded system should reduce the deficit of the consolidated central government. An exception to this rule can be made if the private sector views its contributions to the system as a tax, rather than as a kind of forced saving.

Revenue Sources and Distributional Considerations

The main but not exclusive source of finance for Latin American social security systems is the payroll tax. The base of this levy is considerably less broad than it is in most OECD countries, so that exclusive reliance on it would impose a stringent limit on the size of the system. The payroll tax is a relatively simple levy to administer, and if its effective burden rests on participants in the social security system it has the additional advantage that those who benefit from the system pay for it.

When coverage is less than universal, as is the case in virtually all Latin American countries, and the system is partially financed out of general revenues, the incidence of the system is probably regressive because persons excluded from coverage will bear part of the burden of the taxes that finance it. This conclusion is strengthened if the burden of the payroll tax is shifted forward to the consumer.

The narrow base of the tax means that its rates must be high to finance an extensive system; if the tax is not borne by labor but results in higher labor costs, these rates can encourage capital-intensive production methods and reduce employment. If the base of the tax cannot be broadened to finance the expansion of the system, then both equity and efficiency considerations suggest that additional revenue must be sought from a tax or combination of taxes whose incidence is less regressive and more neutral in its impact on factor costs.

Alternatives to Conventional Social Security Programs

Alternatives to the traditional model of social security are possible in Latin America, as the reforms in Chile demonstrate. The new Chilean system preserves some basic features of the old public system but reduces the direct role of the government. Thus, the new pension system obliges the current generation of young adults to make contributions to a kind of individual retirement account managed by a private financial institution under public supervision. If its participants completely trust the new system, it will have the merits of lessening the incentive to evade

the making of contributions and of creating a more solid and transparent link between the value of contributions and the value of benefits that contributors will ultimately receive. The successful operation of a Chilean-style plan, however, requires both a well-developed system of financial intermediation and a stable financial environment.

Impact of Inflation

The inflationary environment of Latin America complicates the management of social security systems, particularly of their pension component. Accelerating inflation tends to reduce both revenue and expenditure relative to GDP. A high and volatile rate of inflation can produce sizable swings in the real value of benefits and their share of GDP even if the frequency of adjustment is fairly high and the lag in adjustment short. With a significant lag in revenue collection, and prompt and timely indexation, the deficit could grow as a share of GDP.

Administrative Issues

Finally, administrative expenditures in Latin American social security systems are high by comparison with OECD countries because of excessive staffing levels and bureaucratic competition over the supply of medical services between social security institutions and ministries of health. In some Latin American countries the systems still carry the vestiges of numerous benefit plans and institutions that cover different occupational and social groups, with resultant administrative duplication and inefficiency. In consequence, the potential for cost savings is great.

REFERENCES

Aaron, Henry J., *Economic Effects of Social Security* (Washington: The Brookings Institution, 1982).

Arellano, José Pablo, *Políticas Sociales y Desarrollo—Chile 1924–84* (Santiago: CIEPLAN, 1985).

Boskin, Michael, *Too Many Promises: The Uncertain Future of Social Security* (Homewood, Illinois: Dow Jones-Irwin, 1986).

———, Lawrence J. Kotlikoff, and John B. Shoven, "Personal Security Accounts: A Proposal for Fundamental Social Security Reform," Publication No. 63 (Stanford, California: Center for Economic Policy Research, 1985).

Break, George F., "The Economic Effects of the OASI Program," in *Social Security Financing*, ed. by Felicity Skidmore (Cambridge, Massachusetts: MIT Press, 1981), pp. 45–81.

Foxley, Alejandro, Eduardo Aninat, and José Pablo Arellano, *Redistributive Effects of Government Programmes—The Chilean Case* (Elmsford, New York: Pergamon Press, 1979).

Halter, William A., and Richard Hemming, "The Impact of Demographic Change on Social Security Financing," *Staff Papers,* International Monetary Fund (Washington), Vol. 34 (September 1987), pp. 471–502.

International Labor Office, *The Cost of Social Security, Eleventh International Inquiry, 1978–80* (Geneva, 1985).

———, *The Cost of Social Security, Twelfth International Inquiry, 1981–83* (Geneva, 1988).

International Monetary Fund, *International Financial Statistics* (Washington, various issues).

International Social Security Association, Chapter 3 in *Medical Care Under Social Security in Developing Countries,* Studies and Research No. 18 (Geneva, 1982).

McClure, Charles E., Jr., "VAT Versus the Payroll Tax," in *Social Security Financing,* ed. by Felicity Skidmore (Cambridge, Massachusetts: MIT Press, 1981), pp. 129–64.

Mesa-Lago, Carmelo, "Social Security and Extreme Poverty in Latin America," *Journal of Development Economics* (Amsterdam), Vol. 12 (February–April 1983), pp. 83–110.

———, *El Desarrollo de la Seguridad Social en América Latina* (Santiago: CEPAL, 1985).

——— (1986a), "Comparative Study of the Development of Social Security in Latin America," *International Social Security Review* (Geneva), Vol. 39 (1986), pp. 127–52.

——— (1986b), "Social Security and Development in Latin America," *CEPAL Review* (Santiago), No. 28 (April 1986), pp. 135–50.

Shome, Parthasarathi, and Lyn Squire, "Alternative Mechanisms for Financing Social Security," World Bank Staff Working Paper 625 (Washington, 1983).

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

Ugalde, Antonio, "The Integration of Health Care Programs into a National Health Service," in *The Crisis of Social Security and Health Care—Latin American Experiences and Lessons,* ed. by Carmelo Mesa-Lago (Pittsburgh: University of Pittsburgh Press, 1985), pp. 109–46.

Vu, My T., *World Population Projections 1985—Short- and Long-Term Estimates by Age and Sex with Related Demographic Statistics* (Baltimore: Johns Hopkins University Press, 1985).

Wallich, Christine, "Savings Mobilization Through Social Security—The Experience of Chile During 1916–77," World Bank Staff Working Paper 553 (Washington, 1983).

Wilner-Green, Dianne, "Some Effects of Social Security Programs on the Distribution of Income in Costa Rica" (Ph.D. dissertation; Pittsburgh: University of Pittsburgh, 1977).

Wilson, Richard R., "The Impact of Social Security on Employment," in *The Crisis of Social Security and Health Care—Latin American Experiences and Lessons*, ed. by Carmelo Mesa-Lago (Pittsburgh: University of Pittsburgh Press, 1985), pp. 245–84.

Yáñez, José H., "Impacto de las Políticas Previsionales y del Trabajo en el Presupuesto Público: 1979–83," *Estudios de Economía* (Santiago), Vol. 12 (August 1985), pp. 143–67.

Shorter Papers and Comments

The Case for Earmarked Taxes

RANJIT S. TEJA*

The earmarking (or setting aside) of revenues from various taxes for specific types of expenditure is a much-maligned fiscal practice. This paper surveys theoretical arguments and institutional circumstances under which earmarking (even widespread earmarking) may enhance welfare. The paper also questions the criticism that earmarking seriously erodes budgetary efficiency.

THE EARMARKING of taxes refers to the designation of funds—either from a single tax base or from a wider pool of revenues—to a particular end use. This practice may be contrasted with general fund financing, whereby expenditures are met from consolidated receipts. Earmarking provisions are a pervasive fiscal phenomenon in both developed and developing countries and are even written into several constitutions. Typical examples include the earmarking of revenues from property taxes for education, gasoline taxes for highway construction, and payroll taxes for social security payments.

For various reasons the practice has been condemned by economists as wasteful and inefficient. At the bottom of much of this criticism is the homely analogy to households forced to spend receipts from each source of income for specified items of consumption. But what economist has not experienced a twinge of shame when earmarking income from a spouse's new job toward the downpayment on a house, or converting windfall profits into a vacation budget? The source of embarrassment, of course, is the knowledge that an earmarking provision is—from the individual's perspective—an unnecessary constraint in the utility-maximization problem of allocating the last dollar to yield equal marginal utility in every direction. To appreciate the case for earmarked

*Mr. Teja is an economist in the Asian Department. He holds degrees from Delhi University, Jawaharlal Nehru University, and Columbia University. This paper was written while he was with the Fiscal Affairs Department of the Fund.

taxes, a necessary first step is to get away from the tendency to view the fiscal problem from the perspective of an individual called the benevolent social planner.

I. The Rationale for Earmarked Taxes

Decisions on the provision of public goods may be taken either in the context of a vigorous democracy with an active legislature or, alternatively, by executive decree. The former perspective is adopted here, with the legislature viewed as a forum wherein *conflicting* preferences are expressed and resolved.

Lindahl's Theory

This essential feature, which distinguishes fiscal decisionmaking from that of the individual household, was elegantly captured in the writings of Eric Lindahl (1958) over sixty years ago. A by-product of that analysis was a welfare-theoretical case for earmarked taxation based on the Pareto criterion. The main point is that the alternative to earmarked taxes—general fund financing—will result in non-Pareto-efficient outcomes.

Following Johansen's (1963) exposition, consider an economy with two pure public goods, G and S, and a composite private good X. Let Y_A and Y_B be the incomes of two types of individuals—A and B—and X_A and X_B their consumption of the private good X. Let h denote A's "tax share" in the financing of total expenditures $(G + S)$, so that B's share is $(1 - h)$. Under general fund financing, both public goods are financed by the revenues derived from these lump-sum taxes. Given utility functions $U_A(X_A, G, S)$ and $U_B(X_B, G, S)$, each agent maximizes utility subject to the corresponding budget constraint,

$$Y_A = X_A + h(G + S) \tag{1}$$

$$Y_B = X_B + (1 - h)(G + S), \tag{2}$$

yielding the following first-order conditions:

$$dU_A/dG = dU_A/dS \tag{3}$$

$$dU_A/dS = h dU_A/dX_A \tag{4}$$

$$dU_B/dG = dU_B/dS \tag{5}$$

$$dU_B/dS = (1 - h)dU_B/dX_B. \tag{6}$$

The system of equations consists of six equations but only five unknowns (G, S, X_A, X_B, and h) and hence is overdetermined. In other words, no tax share h is capable of meeting all the requirements for efficiency.

One way to arrive at a determinate solution is to assume that preferences are identical. Then equations (3) and (5) are duplicated, leaving five equations in five unknowns. But this trivializes the whole problem, robbing fiscal analysis of one of its distinguishing features.

Suppose instead that expenditures on G and S are met from two separate "earmarked" funds. Thus, expenditure on G is met by payment of gG by A and $(1 - g)G$ by B; likewise, for S, A pays sS and B pays $(1 - s)S$. Agents now maximize utility functions subject to the new budget constraints,

$$Y_A = X_A + gG + sS \tag{1a}$$

$$Y_B = X_B + (1 - g)G + (1 - s)S, \tag{2a}$$

which yield a new set of first-order equations:

$$dU_A/dG = gdU_A/dX_A \tag{3a}$$

$$dU_A dS = sdU_A/dX_A \tag{4a}$$

$$dU_B/dG = (1 - g)dU_B/dX_B \tag{5a}$$

$$dU_B/dG = (1 - s)dU_B/dX_B. \tag{6a}$$

The system now consists of six equations and six variables and hence can be solved for the optimal values of G, S, X_A, X_B, g, and s.

Note that rearranging equations (3a)–(6a) yields

$$(dU_A/di)/(dU_A/dX_A) + (dU_B/di)/(dU_B/dX_B) = 1, \qquad i = G, S. \tag{7}$$

The terms on the left are the sums of A and B's marginal rates of substitution for each public good. The marginal rates of transformation of G and S for X are, by construction, unity. The earmarking solution, then, is equivalent to Samuelson's (1969) well-known condition for the Pareto-efficient allocation of public goods and, indeed, is *necessary* for Pareto efficiency: any move from earmarking to general fund financing would leave both individuals worse off.

This case for earmarking has been couched in terms of numbers of equations and unknowns. However, if one recalls that the problem is resolved when preferences are identical, it should be clear that the role of earmarking lies in facilitating the mutual accommodation of differing preferences in the economy. Note also that, because tax shares are derived from individuals' marginal rates of substitution in utility, the tax

shares may be viewed as a form of "benefits taxation." Although everyone consumes the same level of public goods, agents pay for this in accordance with their marginal utility (benefit).

A well-known shortcoming of Lindahl's (1958) analysis is the assumption that agents in the economy will truthfully reveal their preferences. Owing to nonexcludability in consumption inherent in pure public goods, an agent has every incentive to understate the marginal benefit derived from an additional unit of the public good. Although economists have devised clever (sometimes bizarre) mechanisms for the truthful revelation of preferences that approximate the Lindahl equilibrium, the implementation of Lindahl's solution remains a complex and difficult issue.

The Public Choice Perspective

Whereas a mainstream theorist might view earmarking as a constraint on expenditure, the public choice school stands this proposition on its head and argues instead that, in the context of majority voting, it is general fund financing that imposes constraints on voters' choices. An insightful distinction between earmarked taxation and general fund financing can be illustrated as follows.

Under earmarking, the equilibrium quantity of each public good is determined by a separate vote on expenditure together with a specified tax, or set of taxes, to finance that expenditure. The opportunity cost to the voter of an additional battleship, then, is higher taxes rather than reduced expenditures on other public goods. General fund financing, on the other hand, is characterized by separate voting on the *size* of the budget (tax bills) and the *composition* of expenditures (expenditure bills). Given the government's separately determined budget constraint, the opportunity cost of an extra battleship is no longer higher taxes but instead reduced expenditure on other public goods. This separation of tax decisions and expenditure decisions lies at the heart of the public choice school's distaste for general fund financing.

One strand of this approach—initiated in Buchanan's (1963) seminal paper—argues that the voter who might have approved a tax increase if it were earmarked for, say, environmental protection would oppose it under general fund financing because he or she may expect the increment to be allocated to an unfavored expenditure such as defense. Earmarked taxation then permits a more satisfactory expression of individual preferences. Although Buchanan's analysis is basically an exercise in positive economics, it has distinct normative overtones.

A second strand—developed in Browning (1975)—draws attention to the possibility of perverse outcomes under general fund financing. The constraints imposed by the separation of tax decisions and expenditure decisions under majority voting may be such that a lower quantity of a public good will be supplied even if every voter's preference for that good has increased.

The public choice approach is frequently received with courteous cynicism; Goode (1984), for example, has drawn attention to its irrelevance in countries that lack provision for citizen participation. A stronger criticism (see, for example, Deran (1965)) is the observed tendency for earmarking provisions to become embedded in the state's financial structure and not to be re-evaluated as conditions change (as is implicitly assumed in the public choice model). In practice, the level of public goods supplied may depend entirely on the amount of earmarked revenues and costs, regardless of whether that level has become excessive or deficient. Whereas the usual example cited in this regard is something of a caricature (the continued presence in two U.S. states during the mid-1960s of taxes earmarked for Confederate pensions), most would agree that rigidity in earmarking provisions is the ultimate flaw in the concept. The moral, then, is that such provisions should be reviewed regularly whenever their introduction becomes inevitable.

The Benefits Principle of Taxation and the Economics of User Charges

Frequently, the use of earmarked taxation is justified by invoking the "benefits principle of taxation," which argues that taxes should be borne by those who most benefit from the associated expenditure. The notion is appealing to economists because it parallels the market mechanism for private goods. The analogy makes most sense when an impure public good is characterized by excludability in consumption.[1] Then it becomes possible to finance the activity with a user charge. Although user charges are in a sense "earmarked" for their associated activity, the implementation of a user charge is not equivalent to an earmarked tax in terms of efficiency and equity. When the implementation of user charges is judged to be administratively infeasible or too costly, an earmarked tax can be used as a *second-best* instrument of finance (for example, a gasoline tax as a *proxy* for charges on highway users).

[1]An example of a pure public good that is not excludable in consumption is street lighting.

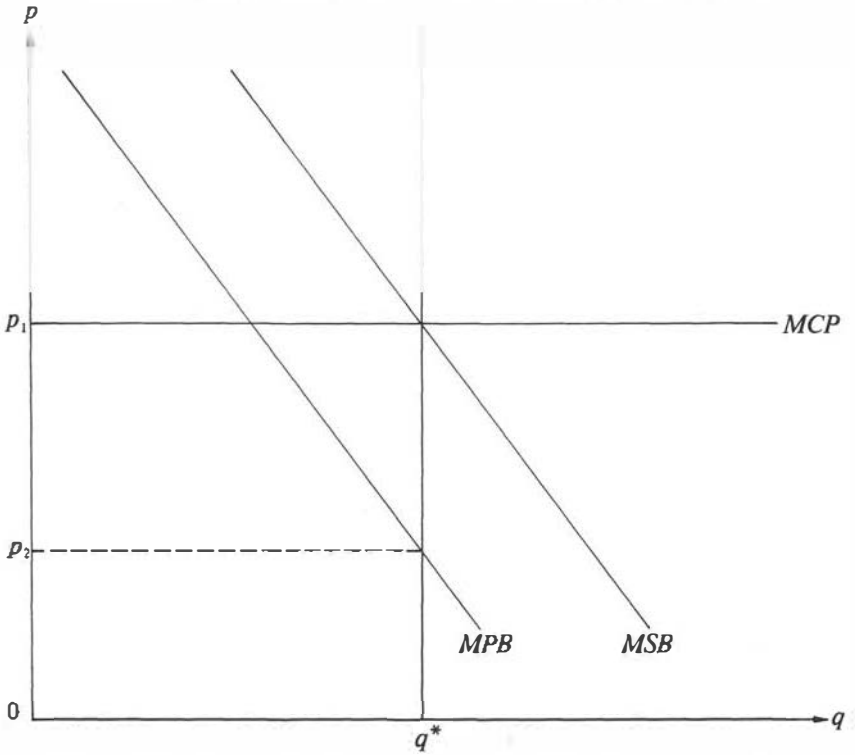
The simple economics of user charges sheds some light on the principles that determine the extent of revenues to be earmarked. Consider the provision of an excludable public good characterized by positive externalities. In Figure 1, the marginal social benefit (MSB) from a unit of consumption exceeds the corresponding private benefit (MPB). The marginal cost of production (MCP) is, for simplicity, assumed to be constant at p_1. The optimal level of consumption for society is q^*, for which consumers are willing to pay only p_2. Then q^* can be publicly provided by a user charge of p_2, together with a per unit subsidy of $p_1 - p_2$. Ideally, earmarked revenues should total p_2q^*, with the remainder, $(p_1 - p_2)q^*$, being financed by general revenues unrelated to the level of consumption of the good. This is precisely how expenditure on road construction and maintenance is financed in a number of countries. Eklund (1972) reports that, for a cross section of countries with earmarking provisions, an average of 63 percent of all road expenditure was financed out of the general fund.

Bureaucracy Versus the Legislature

A somewhat more institutionalist perspective on earmarking may be gained by pursuing in greater detail—as Niskanen (1971) has done—the question of precisely who supplies public goods and services. Although demand for public goods is expressed and resolved in the legislature, public goods are in general supplied by bureaucracies. According to this view, the two distinguishing characteristics of a bureaucracy are the absence of external control on efficiency and weak internal incentives. Weak external control results from the ambiguous nature of a bureaucracy's output and the dependence of the legislature on the bureaucracy for information. Weak internal incentives to produce efficiently are a consequence of an absence of financial incentives for managers and a lack of competition in the market for their final output. The effect of this constellation of circumstances is to make bureaucrats pursue goals other than efficiency, such as larger staff establishments, prestige, and patronage.

This approach to bureaucratic behavior can be used to rationalize the legislative tactic of earmarking revenues for specific end uses. There are a great many expenditures that are clearly in the public interest, but bureaucracies may have little interest in pursuing them adequately. A good example is expenditure on the operations and maintenance of public structures and capital. The prestige-maximizing bureaucrat in general prefers to see his department's allocation go to new and high-

Figure 1. *Provision of a Public Good with Positive Externalities*



profile investment and construction projects rather than to something as mundane as maintaining an old road or irrigation ditch. Although it is true that bureaucratic reputation may depreciate along with public structures, the process is long, and managers are rarely stationary targets. Since proper maintenance of the capital stock is an urgent issue in many countries, the legislature (or the executive branch) may force the bureaucracy's hand on the matter by earmarking funds for such use.

Earmarking and Intertemporal Horsetrading

Earmarked revenues cau also serve to enforce long-term deals between legislators—and in the process be welfare enhancing. Consider the case of the U.S. Hazardous Substance Response Fund (or "Superfund"), which obtains revenue from excise taxes on petroleum and chemicals and uses the proceeds to help clean up environmental damage caused by production and disposal of these same goods. Even though

taxing the chemical industry as a whole can be rationalized as a crude approximation to the "polluter-must-pay" principle, one can still question the validity of earmarking the proceeds to a special fund. Why should such revenues not be assigned to--and expenditures be met from—the general fund?

Consider now the following explanation. The ex ante probability of discovering a toxic waste dump is roughly the same across the country. Ex post, however, a legislator from Nevada has no incentive to approve cleanup expenditures in New York (because it is a pure income transfer) unless the legislator from New York can assure reciprocal support in the event of a similar disaster in Nevada at some future date. But legislators from New York cannot bind their successors to such promises. A way out of this impasse is to earmark revenues for a special fund whose expenditures do not require legislative approval. Such a one-time agreement enforces mutual promises through time and, in this instance, permits the introduction of a welfare-enhancing public activity.

It is interesting to note that the Superfund may not be used to clean up oil spills (even in the unlikely event that the perpetrator cannot be identified). This feature of the Superfund can be explained by the fact that the ex ante probability of an oil spill is not uniform across states. Only coastal states are concerned, and legislators from the interior of the country will naturally oppose a long-term commitment (that is, an earmarked fund) for such purposes.

The role of earmarking in enforcing commitments among legislators can be extended to several situations. Eklund (1969), for instance, has pointed out that, in fractionalistic societies with unstable majority coalitions, earmarking may be the only way in which new expenditure decisions can be agreed on. In the absence of such a provision, society will tend toward the status quo, in the process forgoing public projects with a high return to society.[2]

Although the rule-enforcing aspect of earmarking enhances welfare in these cases, the opposite may also occur in practice. Interest groups may want to earmark specific taxes as a means of limiting their tax burden. The petroleum industry, for example, may lobby for earmarking any gasoline tax for highway maintenance, if they suspect that there is limited scope for expansion of this activity. Similarly, firearms manufac-

[2] Legislators are not the only class that may resort to earmarking in such a circumstance. Wilkie (1974) and Premchand (1983) have suggested that, in several Latin American countries, earmarking was motivated by the executive branch of government, which wanted to bypass the problem posed by legislative logrolling and unstable coalitions.

turers would attempt to earmark taxes levied on their industry for a complementary activity such as maintenance of wildlife.

Other Arguments

One justification for earmarking taxes concerns the effects of erratic financing that results in costly idling of manpower and machinery over extended periods. Earmarking receipts from a stable revenue base is a means of protecting socially important projects from the exigencies of a budgetary crunch and can, over time, be an important cost-saving device for the public sector.

A second justification is that in some developing countries, where democratic institutions are weak and mistrust of the government is high, earmarked taxation can improve tax compliance. If the public can be assured that taxes will be spent in their locality rather than used, say, to indulge a political strongman's penchant for military expansion, they will be more willing to comply with existing taxes. In this case earmarking substitutes for lack of representative power in the executive and legislative branches of government.

Third, many countries have in recent years been forced to make painful (but necessary) fiscal adjustments to serious external shocks. In some cases the brunt of fiscal adjustment has been borne entirely by capital and infrastructural expenditure, with minimal reduction of current expenditures. Although the distinction between current and development expenditure should not be exaggerated—underfunding of the recurrent costs of development expenditure is a long-standing problem in developing countries—there is a tendency to cut back on vital infrastructural and maintenance expenditures rather than on public sector employment and wages. An earmarking clause, designed to protect vital projects and expenditures, would strengthen the government's hand in reducing less socially productive components of public expenditure.

II. Conclusions

The foregoing has examined several theoretical reasons and practical circumstances under which earmarking revenues might be socially beneficial. There is one practical criticism of earmarking that has not been addressed so far: the proposition that earmarking induces budgetary rigidity and inefficiency.

One can respond to this proposition in at least two ways. First, is the

alternative any better? The arguments against earmarking implicitly assume that general fund financing will eliminate the problems associated with earmarking. For instance, it is implicitly assumed that expenditures under general fund financing are indeed periodically reviewed and adjusted to ensure that no program is under- or overfunded. The empirical basis for such an assertion is highly questionable. Most budgeting under general fund financing occurs incrementally rather than by a procedure that evaluates each tax and expenditure afresh from year to year. Indeed, Bird (1982) has argued that the growth of earmarked taxes and the decentralized sector in Colombia should be viewed as a flight from the exceedingly cumbersome and inefficient budgetary and expenditure systems in that country.

Second, it should be recognized that, although widespread earmarking may induce rigidities in the budget, there are ways to reduce this effect. For example, adjustments can be made to the base or the tax rate of an earmarked pool of revenues. Alternatively, an activity can be jointly financed by both earmarked and general fund revenues, the latter providing the latitude necessary to make discretionary changes at the margin.

More fundamentally, one can question whether it is really worth making such a fuss about budgetary rigidity. Earmarked taxes are an application of the benefits principle of taxation and may be associated with higher revenues. As Bird (1984, p. 110) has pointed out, "it makes little sense to criticize earmarking for budgetary rigidity, since without earmarking, there would be less of a budget to be rigid about."

REFERENCES

Bird, Richard M., "Budgeting and Expenditure Control in Colombia," *Public Budgeting and Finance* (New Brunswick, New Jersey), Vol. 2 (Autumn 1982), pp. 87–99.

———, "Earmarked Revenues," in *Intergovernmental Finance in Colombia: Final Report of the Mission on Intergovernmental Finance* (Cambridge, Massachusetts: Law School of Harvard University, 1984), pp. 87–115.

Browning, Edgar K., "Collective Choice and General Fund Financing," *Journal of Political Economy* (Chicago), Vol. 83 (April 1975), pp. 377–90.

Buchanan, James M., "The Economics of Earmarked Taxes," *Journal of Political Economy* (Chicago), Vol. 71 (October 1963), pp. 457–69.

Deran, Elizabeth, "Earmarking and Expenditures: A Survey and a New Test," *National Tax Journal* (Columbus, Ohio), Vol. 18 (December 1965), pp. 354–61.

Eklund, Per, "Taxation and Earmarking in Developing Countries," World Bank Staff Working Paper 43 (Washington, 1969).

———, "A Theory of Earmarking Appraised," *National Tax Journal* (Columbus, Ohio), Vol. 25 (June 1972), pp. 223–28.

Goode, Richard, *Government Finance in Developing Countries* (Washington: The Brookings Insitution, 1984), pp. 11–13.

Johansen, Leif, "Some Notes on the Lindahl Theory of Determination of Public Expenditures," *International Economic Review* (Philadelphia), Vol. 4 (September 1963), pp. 346–58.

Lindahl, Erik, "Just Taxation: A Positive Solution," in *Classics in the Theory of Public Finance,* ed. by Richard A. Musgrave and Alan Peacock (London: Macmillan, 1958), pp. 168–77.

Niskanen, William A., Jr., *Bureaucracy and Representative Government* (Chicago: Aldine-Atherton, 1971).

Premchand, A., *Government Budgeting and Expenditure Controls: Theory and Practice* (Washington: International Monetary Fund, 1983)

Samuelson, Paul A., "Pure Theory of Public Expenditures and Taxation," in *Public Economics,* ed. by J. Margolis and H. Guitton (New York: St. Martin's, 1969).

Wilkie, J.W., "The Budgetary Dilemma in the Economic Development of Mexico, Bolivia, and Costa Rica," in *Fiscal Policy for Industrialization and Development in Latin America,* ed. by David T. Geithman (Gainesville: University of Florida Press, 1974).

Islamic Interest-Free Banking and 100 Percent Money

Comment on Khan

ERVIN JOHN DOAK*

MOHSIN KHAN'S recent article (1986) on Islamic banking notes various trends in this respect and provides a valuable list of references. His main purpose, however, was to provide a "theoretical analysis." This is indeed a much-welcomed contribution and should appeal to anyone researching the possibilities for similar nontraditional financial systems, such as the 100 percent money proposal of Simons (1934) and Fisher (1945). Khan alludes to this application when he notes that "Islamic banking is really a particular variant of equity-participation systems" (Khan (1986, p. 7)). The fact that some countries are actually using the Islamic system does provide us all with valuable experimental information.

There are two major questions I would like to raise about Khan's article. Should he have chosen a different model to contrast the Islamic and traditional banking systems? Should he have tried to contrast these alternative systems in more ways that just that of economic stability?

Khan assumes that the banking system is solely carrying on a saving-investment operation and that all money is "outside money" (pp. 10–11). Although this does help with simplifying his analysis, there are certain difficulties. Would it not have been better to have assumed much the opposite, that the private banking system is responsible for creating most, if not all, the money in the system, and that it holds a mixture of demand and savings deposits? Surely this is more the situation we have in North America. A model that incorporated such assumptions and that could also be reformulated along the lines advocated by the 100 percent money school would probably have been more appropriate.

* Mr. Doak is an Associate Professor of Economics at Saint Mary's University in Halifax, Nova Scotia, Canada.

It certainly would have helped the reader when it comes to understanding Khan's discussion about two windows for deposit transactions (pp. 20–21). It appears that the model used is too rigid and does not really permit as full an analysis of the matter as one would like.

The use of a more appropriate model may also have permitted at least some discussion of other issues of concern to economists such as Simons and Fisher. It would help to show how the move toward a 100 percent reserve plan could help to reduce government debt and taxes. It also might explain how industrial concentration might be reduced, especially if this is considered a problem. Finally, it might throw some light on questions of a political nature.

In a recent paper (Doak (1988)) I have constructed a simple model that attempts to deal with some of the concerns mentioned above. The model assumes that private bank deposits, D, constitute the only type of money in circulation, that the central bank can control the growth of money supply through its purchases of government debt, C, and that the simple money-multiplier rule applies; that is,

$$D = C/r,$$

where r is the private banks' required and actual cash reserve ratio.

A number of theorems can be derived from this model. The most significant finding, however, is that of an objective measure of usury, U:

$$U = [(i)(1 - X)(G - T)] + \left\{\frac{(i)[X - rX - r(1 - X)](G - T)}{r}\right\},$$

where i is the loan rate of interest, X is the percent of public debt purchased annually by the central bank, and $G - T$ is the amount of annual borrowing by government to cover its deficit. The usury is of two major types: the usury on public debt, which can be eliminated by making $X = 1 = 100$ percent, and the usury on private debts, which can be eliminated by making $r = 1 = 100$ percent. The possibilities for increased transfer payments and lower taxes can also be demonstrated.

The principles of Islamic banking and the Simons-Fisher 100 percent money school seem to share the view that private banks should not have the power to create money, that money creation should be a power reserved for the government or its central bank. North American principles are obviously in conflict with these ideas, which may suggest or explain why the relative burdens of interest payments, taxes, and debts are excessive. To the extent that any government can wisely use the power of money creation to pay for its expenses, taxes could be reduced and the private sector freed from having to incur unnecessary debts.

Khan begins his article by referring to the "general resurgence of

fundamental Islamic values in many parts of the world" and by describing how some countries are transforming their economic systems "to accord more closely with the precepts and conditions of Islam" (p. 1). It would have been valuable had Khan's model been such as to demonstrate more of the benefits of such a transformation of the banking system.

These comments notwithstanding, Khan's article was valuable, and it is to be hoped that he and other Fund staff will continue to pursue the subject, including the questions raised here.

REFERENCES

Doak, Ervin John, "Interest and Usury," paper presented at the Twenty-Fifth International Atlantic Economic Conference, London, April 20, 1988 (unpublished; Halifax, Nova Scotia, Canada: St. Mary's University, 1988).

Fisher, Irving, *100% Money* (New Haven, Connecticut: City Printing, 1945).

Khan, Mohsin S., "Islamic Interest-Free Banking," *Staff Papers,* International Monetary Fund (Washington), Vol. 33 (March 1986), pp. 1–27.

Simons, Henry C., *A Positive Program for Laissez-Faire* (Chicago: University of Chicago Press, 1934).

Islamic Interest-Free Banking

Reply to Doak

MOHSIN S. KHAN*

IN HIS COMMENT on my paper (*Staff Papers*, Vol. 33, March 1986, pp. 1–27) Doak contends that I used the "wrong" model to characterize the Islamic banking system, and that I should have employed criteria other than dynamic stability to contrast the Islamic and traditional banking systems.

The objective of my paper was clearly stated: to examine the dynamic response of an Islamic bank to a sudden decline in income (fall in asset values). The model specified, which is simply a dynamic version of the standard IS-LM framework, seems perfectly sensible for the purpose at hand. The assumption of only outside money is perhaps unrealistic, but relaxing this assumption to allow banks to issue deposits does not change the analysis in any way. There is only one bank in the system (with no distinction made between the central bank and commercial banks), and it issues a liability. I called that liability money but could easily have called it deposits.

I agree with Doak that my model does not explain the effects of moving to a 100 percent reserve requirement for banks. But the model was not designed to do so. There are a whole host of important issues in Islamic banking and finance other than stability, and, as shown in a recent collection of papers on the subject,[1] one needs different models to analyze them. Doak's criticism of my model—that it is not sufficiently general to be able to handle all the relevant issues, including "questions of a political nature"—is certainly correct. But how many models in economics would be immune to such criticism?

*Mr. Khan is Assistant Director in the Research Department. He is a graduate of Columbia University and the London School of Economics and Political Science.

[1] Khan, Mohsin S., and Abbas Mirakhor, eds., *Theoretical Studies in Islamic Banking and Finance* (Houston: Institute for Research and Islamic Studies, 1987).

THE WORLD BANK ECONOMIC REVIEW

Volume 2 September 1988 Number 3

Subscriptions to *The World Bank Economic Review* for the first two volumes are available without charge upon written request to World Bank Publications, Sales Unit, Washington, D.C. 20433, U.S.A.

JOURNAL OF MONEY, CREDIT, AND BANKING

RECENT DEVELOPMENTS IN MACROECONOMICS

Papers and comments from the conference sponsored by the Federal Reserve Bank of Cleveland, held in October 1987. Mailed to subscribers as a supplement to the August 1988 issue.

CONTENTS

The regular August 1988 issue contains articles by the following authors: Steven R. Beckman, Joshua N. Foreman, Walter Wasserfallen, W. Douglas McMillin, C. Richard Long, Mark L. Gardner, Ali F. Darrat, Paul R. Allen, William J. Wilhelm, and Michael Devereux.

Published quarterly by **Ohio State University Press**
1050 Carmack Road, Columbus, Ohio 43210-1002 (614) 292-6930

One-Year Subscriptions: $25 Individual $40 Library $19.50 Student

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, 1981–83 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, 1981/82) to indicate a fiscal year or a crop year;

components of tables may not add to totals shown because of rounding.