IMF Working Paper

WP/98/169   INTERNATIONAL MONETARY FUND

IMF Institute

Do IMF-Supported Programs Work?
A Survey of the Cross-Country Empirical Evidence

Prepared by Nadeem Ul Haque and Mohsin S. Khan*

December 1998

Abstract

Over the last two decades a number of cross-country empirical studies have been undertaken to assess whether IMF-supported adjustment programs have led to an improved balance of payments and current account balance, lower inflation, and higher growth. These studies use a variety of methodologies and cover different country samples and time periods. This paper critically surveys the evidence yielded by the cross-country studies, paying special attention to the pros and cons of the respective empirical methodologies employed. These studies, particularly the more recent ones, conclude that IMF-supported programs have generally been successful in stabilizing the economy.

JEL Classification Numbers: E63, F34, F41, O19

Keywords: IMF programs, cross-country studies, program evaluation

Author's E-Mail Address: mkhan@imf.org
                          nhaque@imf.org

Contents   Page

Summary

An important question often raised in connection with IMF-supported adjustment programs is whether such programs "work" in terms of improving the current account balance, increasing international reserves, lowering inflation, and raising the growth rate. The answer requires evaluating the effects of past programs on the variables of interest. Over the past twenty years or so there have been a number of empirical studies on the subject, both inside and outside the IMF; this paper reviews the evidence provided by these studies.

The survey of the literature on the cross-country empirical evidence on the macroeconomic effects of IMF-supported adjustment programs points to two broad conclusions. First, the empirical analyses have been conducted using different methodologies, the relative merits of which deserve careful examination. This paper argues that the proper standard for measuring program effects is to compare the macroeconomic outcomes under a program with the outcomes that would have emerged in the absence of a program. Several recent studies attempt to apply this "counterfactual" criteria to evaluating program performance through the estimation of policy reaction functions for program and non-program countries.

Second, the results of most of the cross-country empirical studies indicate that IMF-supported programs lead to an improvement in the current account balance and the overall balance of payments. Several recent studies indicate that the rate of inflation falls, but this change is generally not found to be statistically significant. In the case of growth, the consensus seems to be that output will be depressed in the short run, but over time growth begins to rise. These newer empirical results indicate that, on average, IMF-supported adjustment programs have been more effective in achieving their objectives than earlier analyses suggest.

I. INTRODUCTION

The International Monetary Fund ("Fund") plays a central role in the adjustment efforts of its member countries by assisting in the design of appropriate adjustment programs to achieve viability of the balance of payments, accompanied by price stability and a sustained high rate of growth, and by providing financing to support these programs.[1] A Fund-supported adjustment program includes a mix of stabilization and structural reform measures that aim to restore a sustainable balance between aggregate demand and supply and to simultaneously expand the production of tradables. Such an adjustment program takes the form of a set of policy intentions by the government (a "letter of intent") that is judged by the Fund to warrant financial support. This financial support is therefore conditional on the policy measures being carried out.[2] The targets for the key macroeconomic variables, which in a typical program include the level of net international reserves, the current account balance, inflation, and the growth of real GDP, [3] the choice of policies to achieve these targets, and the amount of financing provided by the Fund all result from extensive and detailed negotiations between the country authorities and the Fund. Thus, a Fund-supported adjustment program necessarily reflects the individual economic situation of the country and the preferences of the government.

While the Fund's role is widely regarded as both necessary and useful, one important question that is often raised in connection with Fund-supported adjustment programs is whether such programs "work". In other words, are these programs effective in achieving their objectives of improving the current account balance, increasing international reserves, lowering inflation, and raising the growth rate? This is essentially an empirical question that requires evaluating the effects of past programs on the variables of interest. Such an evaluation is conducted regularly by the Fund's Policy Development and Review Department and the results reported to the Fund Executive Board. In addition, there have been a number of studies published over the last 20 years or so, both inside and outside the Fund, that examine this question using a variety of empirical methods.

Although the final verdict is probably not in, it would be fair to say that a consensus has emerged that programs have on balance worked. But, as will be seen, evaluating the

---

[1]Santaella (1995) details the macroeconomic characteristics of countries prior to their adopting a Fund-supported program. Typically, the countries are suffering from a worsening current account balance, a loss of international reserves, and an increase in inflation.

[2]Williamson (1982). See Polak (1991) for a recent description of the conditionality associated with Fund-supported programs.

[3]Other macroeconomic objectives could include the resumption or expansion of private capital inflows, exports, investment, and so on.

effects of programs is not an easy task, which is certainly one reason why there has been controversy on this issue.

The purpose of this paper is to review the evidence that is available on the effects of Fund-supported programs, paying special attention to the methodologies employed in various studies to estimate these effects. The paper is in the nature of a survey and relies exclusively on published studies. As such, it does not contain any new empirical evidence, but focuses instead on outlining the current state of thinking on the subject and assessing the existing evidence.

The remainder of the paper is organized as follows. Section 2 examines the conceptual and methodological issues related to evaluating programs. The available empirical evidence is reviewed in Section 3. The final section outlines the broad conclusions that can be drawn from the review of the evidence, as well as the implications for future work on the question.
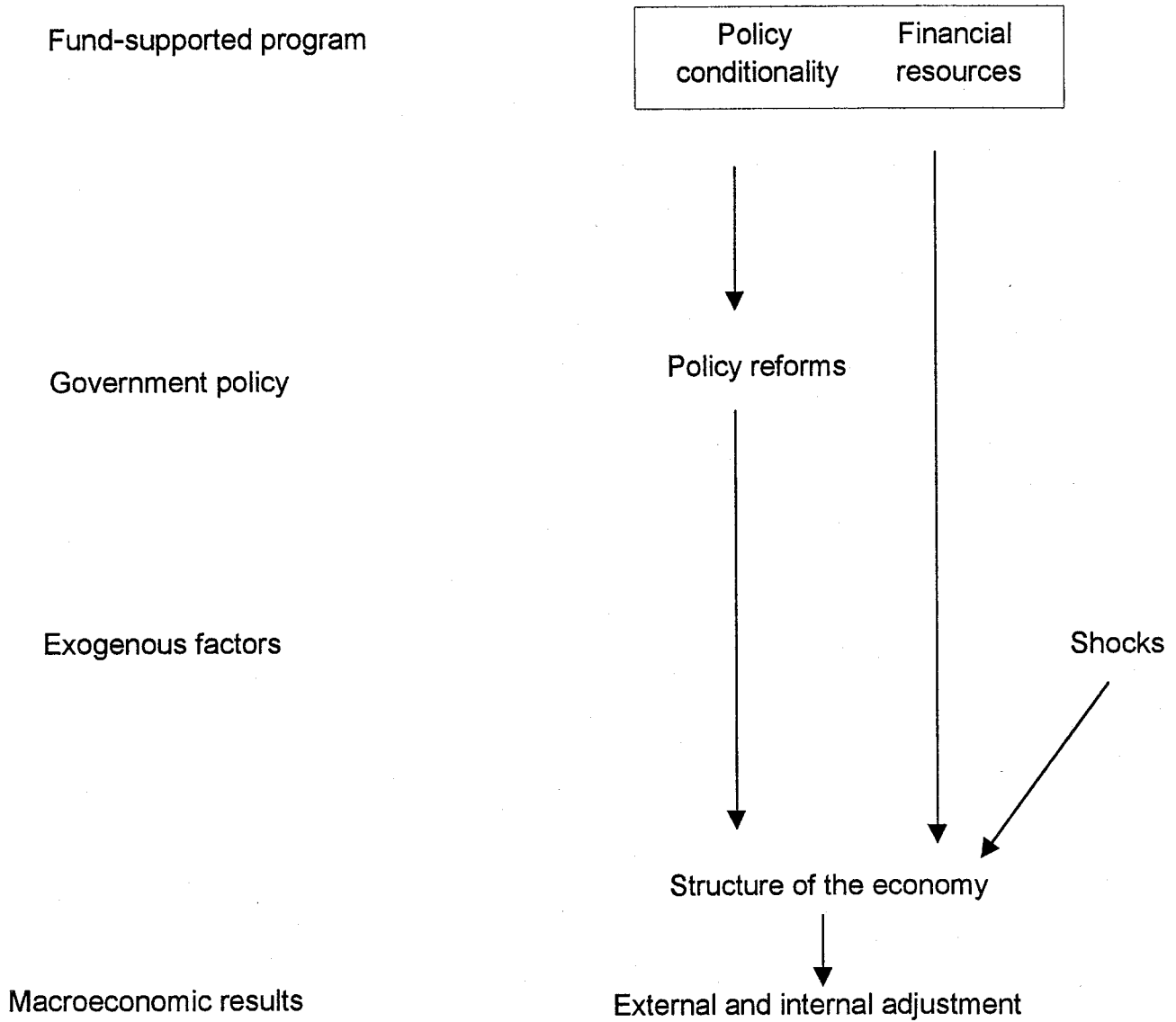
II. CONCEPTUAL AND METHODOLOGICAL ISSUES

A. Defining the Effectiveness of Programs

Defining the effectiveness of Fund-supported adjustment programs is not a straight-forward task for at least two reasons. First, although the success of a program is measured in terms of macroeconomic outcomes, e.g. an improved balance of payments, the conditions agreed upon in Fund-supported programs by the two parties relate to policy variables, e.g. the expansion of domestic credit or reduction in the fiscal deficit. While it is relatively easy to check whether the program country has implemented the agreed policy changes via the setting of performance criteria on key policy variables, it is more difficult to know if making these changes would lead to the desired macroeconomic outcomes. This is because Fund-supported programs are complex packages of policy measures that include, among other things, monetary and exchange rate policies, fiscal measures, policies to raise investment and improve its efficiency, trade liberalization, wage reforms, and financial sector reforms. The theory underlying the dynamic linkages between such a policy package—combining aggregate demand policies with supply-enhancing and relative price policies—and a set of multiple macroeconomic targets is not well established. Thus, the adjustment package is not necessarily guaranteed, at least from a theoretical viewpoint, to achieve the desired outcomes.

Second, the Fund-supported program is only one of many macroeconomic "shocks" to the country with a program. External shocks, such as changes in the terms of trade or in the cost of servicing foreign debt, will also affect the country's ability to achieve the macroeconomic objectives of the program. Measures of program effectiveness have to filter out these unanticipated exogenous shocks.

Figure 1 illustrates these concurrent and sometimes conflicting influences on effectiveness of programs. As can be seen, the linkage between the program and the eventual external and internal adjustment is not direct. The program's policy recommendations must

Figure 1: Linkages Between the Fund-Supported Program and Macroeconomic Adjustment

Fund-supported program

| Policy conditionality | Financial resources |

Government policy

Policy reforms

Exogenous factors

Shocks

Structure of the economy

Macroeconomic results

External and internal adjustment

first be translated by the country into actual policy reforms. These policy reforms and the financial resources provided to support the program must interact with the structure of the economy to bring about the desired economic adjustment. While the government is implementing the policy reforms, exogenous factors (terms of trade movements, weather fluctuations, productivity shocks) can work through the economic structure to affect the macroeconomic outcome. These exogenous factors may help the cause of macroeconomic adjustment, but they may also at times work against the policy reforms.

Since the introduction of a Fund-supported program aims to achieve an adjustment in policies and hence in macroeconomic outcomes, the proper measure of the effectiveness of the program has to isolate the impact of the program on the outcomes and compare that to the alternative of what would have happened in the absence of the program. In other words, a comparison has to be made between the actual outcome due to the program with the **counterfactual**, that is, the macroeconomic outcome that would have resulted had there been no program. The counterfactual is the right yardstick against which to assess program performance and the standard most widely employed in economics to measure the impact of government policy interventions.

However, the counterfactual is not observed. In practice, countries fall into only one of two states—program or non-program—but obviously never both. An accurate measure of the effects of programs requires the comparison of experience in these two states. Thus, the counterfactual has to be constructed, and the ideal counterfactual will be an equilibrium for the same economy facing the same exogenous factors, and differing only in its non-participation in the Fund-supported program.[4]

The informational requirements of the ideal measure of effectiveness are quite daunting. To do the comparison requires a macroeconomic model whose parameters are invariant with respect to different policy settings. If the structural parameters themselves change, the comparison would require two models of the economy—one describing the structure when the country has a program, and the other describing the structure in the absence of a program. Information is also required on the values on the structural parameters and the policy reaction function parameters. In reality, this amount of information is not available for the typical program country and thus various methods have been advanced to construct a counterfactual that can be a good approximation to the true counterfactual.

---

[4]For a simple theoretical derivation of the relevant counterfactual, see Conway (1998).

B. Alternative Approaches to Estimate the Effects of Fund-Supported Programs

In the literature there have been four main approaches suggested to measure the effects of Fund-supported programs. These approaches, each of which constructs a different type of counterfactual, are:[5]

(i)   The **before-after** approach, which compares the macroeconomic performance under a program (or after it has been initiated) and performance prior to the program;

(ii)   The **with-without** approach, which compares the macroeconomic performance in countries with programs and performance in a "control group" of non-program countries;

(iii)   The **generalized evaluation** approach, which compares performance in program and non-program countries, adjusting for differences in initial conditions among the countries and controlling for exogenous (principally external) influences; and,

(iv)   The **comparison of simulations** approach, which compares the simulated performance under Fund-supported program policies and simulated performance with some other set of policies.

The **before-after** (BA) approach carries the basic advantage of ease of calculation. One merely has to compare the macroeconomic outcomes in the pre-program period with those in the program period to estimate the program effects.[6] The problem, however—and it is an overriding one—is that the BA approach assumes that all other things remain equal, and will not yield an estimate of the independent effects of programs whenever the non-program determinants of the macroeconomic outcomes are changing between the pre-program and the program period. The fact of the matter is that these non-program determinants, involving external factors like industrial-country growth rates, terms of trade variations, and movements in international interest rates, as well as domestic factors such as productivity shocks and shifts in weather conditions, do change markedly from year to year in reality. This means that

---

[5]Another approach, referred to as the **actual versus targets approach** compares actual performance under the program with the objectives specified in the program. This particular approach, which does not involve the counterfactual, is not pursued here. For examples of this type of analysis, see Reichmann (1978), and Zulu and Nsouli (1985).

[6]The pre- and post-program periods can be of any length. As will be shown later, most studies using this approach look at the year before and the year after the program is initiated, but this is arbitrary.

the BA estimation of program effects will typically be biased and unsystematic over time.[7] They will be biased because this approach incorrectly attributes all of the change in outcomes between the pre-program and program periods to program factors. They will also vary over time, because the estimated program effects for a given year will often be dominated by specific non-program influences of that year. Thus, for example, if the terms of trade improve between year t and year t+1, all t+1 programs will look as if they performed very well, while if the terms of trade decline in some later year, all programs for that later year will look as if they performed poorly. In other words, the outcomes may have nothing to do with the program.

These shortcomings of the BA approach also make it a poor estimator of the ideal counterfactual, since it assumes a counterfactual in which policies and the external environment would have remained constant at their pre-program values. The basic reason why the BA approach is flawed as an estimator of the counterfactual is that the situation prevailing before the program is not likely to be a good predictor of what would have happened in the absence of the program.[8]

The **with-without** (WW) approach is an attempt to get closer to the ideal counterfactual. It is designed to overcome the inability of the BA approach to distinguish between program and non-program determinants of macroeconomic outcomes. The basic reasoning behind the WW approach is as follows. Assume that program countries and non-program countries are subject to the same non-program determinants, that is they face the same external environment. Then, according to the arguments underlying the WW approach, by comparing before-after changes in outcomes in program countries to those in a control group of non-program countries, the effects of external factors will cancel out, leaving the difference in group performance to reflect only the effects of Fund-supported programs. Put in terms of the counterfactual, the idea is to use the observed performance in non-program countries as an estimate of what the performance in program countries would have been in the absence of a Fund-supported program.

However, the WW approach has problems as well. The main problem is that program countries can and do differ systematically from non-program countries prior to the start of a program, and this turns out to matter for performance evaluation. Basically, the point is that program countries are not randomly selected. Instead, they are adversely selected in the sense of having relatively poor economic performance prior to the program period. This should not

---

[7]For a demonstration of the bias, see Appendix. See also the papers by Goldstein and Montiel (1986) and Conway (1998).

[8]By making a judgmental correction for the influence of non-program factors it would be possible to improve upon the estimates of the counterfactual that would emerge from a mechanistic application of the BA approach. However, such judgmental corrections are difficult to make, especially when the number of non-program factors is large.

be too surprising, because an essential requirement for Fund financial support is that the country have a balance of payments need. This alone suggests that program countries would be expected to have weaker than average external positions—namely the current account balance and level of international reserves—when the program was implemented. The implication of non-random selection of program countries is that the simple WW estimates of program effects will be biased.[9] Intuitively, the bias occurs because, under non-random selection, the WW estimator attributes differences in outcomes exclusively to program status when in fact the difference in starting or initial positions itself is a cause of differences in subsequent performance of the two groups of countries. Furthermore, the direction of the bias can go either way. If past economic difficulties signal less serious current difficulties—even in the absence of a program—then the WW approach will overstate the positive effects of a Fund-supported program. Conversely, if past difficulties signal even more serious current difficulties, then the effects of Fund-supported programs will be understated.

The recognition of the biases inherent in both the BA approach and the WW approach has led to the development of the **generalized evaluation estimator** (GEE) approach.[10] The GEE approach modifies the simple WW approach in two important ways. First, it accepts the non-random selection of program countries, identifies the specific differences between program and non-program countries in the pre-program period, and then controls for these differences in initial positions in the comparison of subsequent economic performance.[11] Second, it attempts to capture the effects of policy and other variables on the macroeconomic outcomes, taking into account how policies would have evolved in the absence of a program.

To make the GEE approach operative one needs to identify, aside from the initial conditions, the relevant reduced form relationships linking policy instruments and other exogenous variables to policy targets, and the policy reaction functions that show how policy instruments evolve as the state of the economy changes. The reduced-form equations are needed to determine the effects of alternative policies on the target variables, controlling for exogenous variables. The policy reaction functions are needed to determine what policies would have been chosen, given pre-program conditions, in the absence of a program.[12]

---

[9]See Appendix.

[10]This procedure was developed by Goldstein and Montiel (1986), and later refined by Khan (1990) and Conway (1994).

[11]By selecting the control group of non-program countries to include only those that had balance of payments problems in the pre-program period, Gylfason (1987) also attempts to adjust for the bias while applying the WW approach.

[12]See Appendix for the type of equation to be estimated.

These various empirical relationships require a substantial amount of data and are not easily implemented. There will still be some margin of error between the results of the GEE approach and the ideal counterfactual. However, this error is likely to be considerably smaller than for the BA and WW approaches. In summary, there are sound econometric reasons to prefer the GEE measure over the BA and WW approaches. First, it uses more information about the program country and other non-program countries, and can thus define a more precise and accurate counterfactual. Second, statistical techniques are available to correct for selection bias with this methodology, while such corrections cannot be easily made in the cases of the BA and WW approaches.

Finally, there is the **comparison of simulations** (SIM) approach, which unlike the other three approaches outlined above, does not determine program effects from actual macroeconomic outcomes in program countries. Instead, it relies on simulations of econometric models to infer the hypothetical performance of policies included in a Fund-supported program and an alternative policy package. If the aim of the overall exercise is to evaluate the results of specific Fund-supported adjustment programs, then the use of actual program outcomes is indispensable. However, if the purpose is to evaluate the design and effectiveness of Fund-supported adjustment programs in general, then examining the likely effects of alternative policy packages can be quite useful and revealing.

In broad terms, the SIM approach carries three advantages. First, one can draw on a wider body of adjustment experience, since the database need not be restricted to countries with Fund-supported programs. Second, since one specifies the policy simulations, one does not have to be concerned that incomplete implementation of policies, which is often a problem in Fund-supported programs, will blur the results. In contrast, the approaches that rely on actual outcomes require the untangling of the effects of the program from the degree of implementation, and this is usually not done. Third, the SIM approach, by its very nature, focuses on the relationship between policy instruments and policy targets. As such, it provides better information on how programs work than do approaches that look at the bottom line of policy targets.

There are, however, practical problems with the SIM approach. To use this approach it is necessary to have an econometric model that incorporates the relations between various policies and certain important macroeconomic variables. While there have been a number of attempts at building such models for developing countries,[13] there is as yet no single model available that covers the whole range of policy measures contained in a typical Fund-supported program. Existing econometric models are clearly unable to analyze all the questions relating to Fund-supported programs and, in particular, they do not capture the complex ways in which policy variables are related to the ultimate objectives of programs. Even if one had a suitable model to work with, one would still have to face up to the critique of ex-ante econometric policy evaluation (the so-called Lucas critique). Specifically, the

---

[13]See, for example, the studies contained in Khan, Montiel and Haque (1991).

parameters in econometric models may not remain invariant to changes in the policy regime, so that it would be incorrect to treat such parameters as fixed across alternative policy simulations. In other words, the actual effects of hypothetical policy packages on macroeconomic variables may turn out to be quite different from the simulated results and in ways that are difficult to know in advance.[14] An additional concern is that, due to credibility factors, the effect of a given policy may be different when it is implemented within the context of a Fund arrangement than when it is implemented outside it. Agents may believe, for example, that policies agreed to with the Fund are more likely to be carried through and thus would be more inclined to change their behavior.

III. EMPIRICAL APPLICATIONS

A summary of the studies evaluating the effects of Fund-supported programs is provided in Table 1. For convenience, attention is focused only on the principal macroeconomic variables—the overall balance of payments, the current account balance, inflation and growth.[15] Furthermore, comparisons of performance are made over a one-year time horizon, unless otherwise noted. While comparing performance between the year prior to the program with performance during the program year is essentially an arbitrary choice, most studies have restricted themselves to annual comparisons. In the discussion that follows, the studies are grouped according to the approach that was employed by them.

A. Before-After Approach

The most popular approach in the early literature on the effects of programs has been the BA approach. The first study to use this approach in the analysis of Fund-supported programs was by Reichmann and Stillson (1978). These authors examined a total of 79 Fund-supported programs implemented during 1963-72 and compared the behavior of the balance of payments, inflation, and growth during the two-year periods before and after the program has been initiated. Using non-parametric statistical tests, they found that programs generally succeeded in slowing down the rate of expansion of total domestic credit and credit to the government—the two key monetary policy variables. However, a significant improvement in the balance of payments was achieved in only about a quarter of the programs. Of the 29 programs involving countries with high inflation during the program period, the rate of inflation fell in six of the eleven programs for which there was a notable deceleration in the rate of domestic credit expansion; in the nine programs which included a devaluation, inflation was higher in five. Finally, growth performance was examined for 70 programs and it was

---

[14]Some of the papers in Khan, Montiel and Haque (1991) attempt to handle this problem by adjusting certain relevant parameters of the model.

[15]Some studies have considered other variables, such as the degree of income inequality. See Johnson and Salop (1980), Sisson (1980), and Pastor (1987).

Table 1. Summary of Empirical Evaluations of the Effect of Fund Programs

Study	Time period	Number of programs	Number of countries	Effects on: [a]			
				Balance of payments	Current account	Inflation	Growth
Before-after							
Reichman and Stillson (1978)	1963-72	79	...	0	..	0	+
Connors (1979)	1973-77	31	23	0	0	0	0
Killick (1984)	1974-79	38	24	0	0	-*	0
Zulu and Nsouli (1985)	1980-81	35	22	..	0	0	0
Pastor (1987)	1965-81	...	18	+*	0	0	0
Killick, Malik and Manuel (1995)	1979-85	...	16	+*	+*	-*	+*
Schadler, et al. (1993)	1983-93	55	19	+	-	-	+
With-without							
Donovan (1981)	1970-76	12	12	..	..	-	+
Donovan (1982)	1971-80	78	44	+	+	-	.
Loxley (1984)	1971-82	38	38	0	0	-*	0
Gylfason (1987)	1977-79	32	14	+*	..	..	0
Generalized Evaluation							
Goldstein and Montiel (1986)	1974-81	68	58	-	-	+	-
Khan (1990)	1973-88	259	69	+*	+*	-	-*
Conway (1994)	1976-86	217	73	..	+*	-	-,+*
Bagci and Perraudin (1997)	1973-92	...	68	+*	+*	-	+*
Dicks-Mireaux, et al. (1997)	1986-91	88	74	..	..	-	+*
Simulation							
Khan and Knight (1981)	1968-75	...	29	+	+	-	-
Khan and Knight (1985)	1968-75	...	29	+	+	-	-,+*

a/ Direction of change: (+) indicates positive effect, (-) indicates negative effects, (0) indicates no effect. An asterisk (*) indicates statistically significant at the 5 percent level.

- 14 -

concluded that on balance Fund-supported programs did not have adverse effects on growth rates. In certain cases, growth did decline after the start of the program relative to the previous year's rate of growth, but this result was matched by a number of instances where the growth rate rose.

A similar procedure was followed by Connors (1979) who examined 31 programs adopted during the 1973-77 period, comparing the behavior of the main macroeconomic variables one year before and after the inception of the program. Connors concluded that Fund-supported programs apparently had no discernible effects on growth, inflation, the current account deficit, and the overall balance of payments.

The study by Killick (1984) attempted to capture the effects of lags by comparing the pre-program behavior of the balance of payments, the current account, growth and inflation, with the behavior of these variables both one and two years after the program. Killick employed non-parametric statistics and regression analysis to gauge the effects of 38 programs during 1974-79. His results showed no statistically significant effects—positive or negative—of Fund-supported programs on the balance of payments or the current account deficit. There was a small positive effect on growth, but at the same time inflation was higher as a result of programs. Zulu and Nsouli (1985) also constructed before-after measures of program effects in their study of 35 programs for 22 African countries. They found a reduction in the growth rate for most countries, a worsening of inflation in just over half the countries, and no general pattern with respect to either the current account position or the overall balance of payments.

Also looking at regional patterns, Pastor (1987) estimated program effects for 19 Latin American countries during 1965-81. Using one-year comparisons and on the basis of alternative statistical tests, Pastor concluded that Fund-supported programs led to a significant improvement in the balance of payments, but that apparently there was no effect on the current account, inflation, or the rate of growth of nominal GDP.[16]

Killick, Malik and Manuel (1995) updated and extended the earlier Killick (1984) study. In examining the effects of programs in 16 countries over the period 1979-85, they found that in these cases the results were quite different. Fund-supported programs led to an improvement in the balance of payments, the current account balance, and the growth rate, while inflation was reduced. Also, in contrast to the previous study, the estimates of the program effects were statistically significant.

Finally, the study by Schadler et al. (1993) examined the performance of 19 countries that entered into SAF and ESAF arrangements with the Fund during 1983-93. The results, none of which were tested for statistical significance, indicate that while the overall balance of

---

[16]In contrast to other studies, Pastor (1987) focuses on nominal rather than real GDP.

payments improved, the current account balance worsened. Insofar as other targets are concerned, inflation declined and the growth rate rose as a result of programs.

B. With-Without Approach

The WW approach was first used in two studies by Donovan (1981, 1982), which analyzed a sample of programs implemented from 1970 to 1980. Donovan compared changes in target variables in program countries with contemporaneous changes in these variables for a control group of non-program countries. In both studies the control group consisted of all non-oil developing countries, and the comparisons were carried out over one-year and three-year time horizons. The results indicated relative improvements for program countries in the ratio of the current account to GDP, in the ratio of the overall balance of payments to exports, and in the rate of inflation for both the 1970-76 and 1971-80 periods. However, the growth performance of program countries differed little from that of the control group.

Loxley (1984) applied the same types of tests as Donovan (1982) to a group of 38 least-developed countries (defined as countries with per capita incomes of less than $690 in 1980) that had programs with the Fund during 1971-82. His results were less definitive, however, than those obtained by Donovan. The least-developed countries with programs did no better, on average, in terms of current account, balance of payments, and growth performance relative to various control groups (comprising other least-developed countries without programs, the 44 program countries considered by Donovan (1982), and all non-oil developing countries). The relative improvement was statistically significant only in the case of inflation, and then only in the three-year comparisons.

Gylfason (1987) also used a version of the WW approach, taking as a control group a set of non-program countries that had experienced economic difficulties in the pre-program period for program countries. The variables considered were changes in the rate of growth of domestic credit, in the ratio of the balance of payments to GDP, and in the growth of output. Tests were conducted to see whether these variables differed statistically between program countries and the control group countries. The results indicated that program countries experienced statistically significant reductions in domestic credit growth and improvements in the ratio of the balance of payments to GDP, but no significant difference between program countries and the control group was evident in the case of real output growth.

C. Generalized Evaluation Estimator

In order to correct for the biases inherent in the BA and WW approaches, Goldstein and Montiel (1986) developed the generalized evaluation estimator approach. This approach has now become the estimator of choice in evaluating the effects of Fund-supported adjustment programs.[17]

Goldstein and Montiel applied the GEE approach to a sample of 68 programs for 58 countries implemented during 1974-81. These authors found that program countries systematically demonstrated weaker performance—that is, higher inflation, slower growth, and larger current account and balance of payments deficits—than non-program countries in the pre-program period. Adjusting for these pre-program differences in performances and taking into account the effects of policy instruments on targets, Goldstein and Montiel used regression analysis to estimate the program effects. Two sets of interesting results emerged in this study. First, there were no statistically significant effects of programs on the current account and balance of payments, on the rate of inflation, or on the growth of real output. Second, the estimated program effects using the GEE approach were quite different from those obtained with the BA or WW approaches in terms of the signs of the coefficients measuring program effects.

The GEE approach was refined and applied by Khan (1990) to a much larger sample of 259 programs in 69 countries over the period 1973-88. Both one-year and two-year comparisons were used to be able to pick up the dynamics of the relationship between program and the target variables. Khan found that there was an improvement in the balance of payments, but this improvement was statistically significant only when the period of evaluation was extended to two years after the inception of the program. The current account deficit was also immediately reduced, and this effect was strengthened over time. While the inflation rate was lowered, this reduction did not show up as statistically significant in any of the tests performed. Finally, the growth rate declined in the program year, but when the time horizon of performance evaluation was extended beyond the program year, the adverse growth effects were diminished.[18]

The results of Khan (1990) for the effects of programs on inflation are supported by Dicks-Mireaux, Mecagni and Schadler (1997) for the more recent period of 1986-91 and covering a larger group of countries. Programs do reduce inflation, but the coefficient measuring the program effects is not statistically significant. On the other hand, the results for

---

[17]It has also been employed in evaluating World Bank structural adjustment programs. See Corbo and Fischer (1994) for a discussion of the merits of this approach, and Faini et al. (1991) and Corbo and Rojas (1992) for empirical applications.

[18]Khan (1990) also showed that the achievement of macroeconomic stability defined in various ways had a positive impact on long-run growth.

growth turn out quite different from what other studies found. The implementation of a program leads to an immediate improvement in growth, and this effect was found to be statistically significant.

One reason not yet considered as to why standard approaches to measuring the effects of programs, and this applies as well to the GEE approach, may not be adequate is the possibility that participation in Fund-supported programs may not be random. Indeed, countries are likely to be self-selecting themselves into Fund-supported programs. Program participation is a conclusion of successful negotiations between the Fund and the countries. The success of the negotiations depends on the existence of balance of payments need as well as a policy package that is expected to alleviate the need and improve the viability of the longer-term balance of payments. Consequently, the country must be prepared to take the policy measures that are agreed, as well as determine that the costs of alternative financing are less than making the necessary changes in policy. Thus, financing availability, implementation capacity, and the magnitude of required changes are all elements of the program participation decision. Participation takes place according to countries self-selecting themselves based on their respective economic circumstances. Looked at in this manner, the choice of participation in Fund-supported programs is itself an endogenous variable, and ignoring this characteristic may lead to biased results. The participation decision thus has to be estimated simultaneously with the performance and policy choice parameters that are part of the GEE approach to allow for the endogenous determination of the program participation variable.

There are two recent studies that take selectivity and the decision to participate into account in estimating program effects. Conway (1994) applied the modified GEE approach to a sample of 217 programs implemented during 1976-86. The results were the following: the current account improved in the context of a Fund-supported program, inflation was lower, and the rate of growth first fell and then rose. All these effects were statistically significant. The Conway results were confirmed by a similar analysis done by Bagci and Perraudin (1997) for 68 program countries over the period 1973-92. These authors used a maximum likelihood estimation method, in contrast to Conway (1994) who employed two-stage least-squares, to estimate the two-equation model. Bagci and Perraudin found that programs improved the current account balance and the overall balance of payments, reduced inflation, and raised growth in the short run, and these changes were also all statistically significant.

D. Comparison of Simulations

An alternative to examining actual program experience, as has been done in the cross-country studies described so far, is to perform policy simulations with a macroeconomic model to estimate program effects. For example, Khan and Knight (1981) constructed a small dynamic econometric model and estimated its parameters on a pooled cross-section time-series sample of 29 developing countries, most of which had programs with the Fund. They then investigated the hypothetical effects of a stabilization program whose objective was an external balance target using policies that figure prominently in Fund-supported programs. The simulation experiments showed that such a program produced a sharp price deflation in

the first year, followed by a temporary burst of inflation as prices rose back to their equilibrium level. Output, on the other hand, contracted sharply in the first year, then rose temporarily above its full-employment level, approaching that equilibrium level gradually over several years.

In a further study, Khan and Knight (1985) extended the simulation analysis to a comparison of alternative policy packages. Specifically, they compared the results for the balance of payments, inflation, and real output growth of a package of demand-management policies (that is, a once-for-all reduction in the rates of growth of domestic credit and government expenditures, plus a devaluation) with a combined package of demand-management and structural policies (that is, the above-mentioned demand-management policies plus a set of structural reforms that would gradually raise the rate of growth of capacity output). The demand-management package improves the balance of payments almost immediately, but at the cost of a temporary higher rate of inflation and a short-run reduction in growth. The simulations of the combined package showed that structural policies could help to offset, at least partially, any short-run effects that result from demand restraint and the inflationary consequences of devaluation. Furthermore, the longer-run effects on the balance of payments, inflation, and growth turn out to be much more favorable than the short-run effects.

IV. Conclusions and Implications for Future Work

Over the past two decades or so, there have been a number of empirical studies examining the effects of Fund-supported adjustment programs on key macroeconomic variables, such as the current account and the overall balance of payments, inflation, and the rate of economic growth. Such evaluations play an important role in the design of programs as the lessons they yield, both positive and negative, can be incorporated into the operations of the Fund. The objective of this paper was to provide an overview of the methodologies adopted and results obtained by the studies on the subject, with a view to assessing where we stand at present with respect to our knowledge about the effectiveness of past programs, and where we go from here in terms of future evaluation of Fund-supported adjustment programs.

The survey of the literature on the cross-country empirical evidence on the macroeconomic effects of Fund-supported programs points to two broad conclusions.

First, the empirical analysis that is available has been conducted using different methodologies, the relative merits of which deserve careful examination. Many of the earlier studies attempted to gauge program effectiveness by comparing macroeconomic outcomes in program countries with the performance prior to the implementation of the program, or with the observed performance of countries without programs. If the proper standard for measuring program effects is to compare the macroeconomic outcomes under a program with the outcomes that would have emerged in the absence of a program, or under a different set of policies, then none of the earlier approaches—labeled as the before-after and with-without approaches—is fully satisfactory. More recent studies attempt to apply the counterfactual

criteria to evaluating program performance through estimation of policy reaction functions for program and non-program countries and through simulation experiments with macroeconomic models. One can place more confidence in the results yielded by these later studies.

Second, it is now becoming increasingly well-accepted that Fund-supported programs lead to an improvement in the current account balance and the overall balance of payments. The results for inflation are less clear cut. Most of the recent studies indicate that the rate of inflation falls, but this change is generally not found to be statistically significant. In the case of growth, the consensus seems to be that output will be depressed in the short run as the demand-reducing elements of the policy package dominate. Over time the structural reform elements of the program start to take effect and growth begins to rise. These newer empirical results indicate that, on average, Fund-supported adjustment programs have been more effective in achieving their objectives than earlier analyses would suggest.

What needs to be done in terms of future evaluation work? The principal message that emerges from this survey is that comparing the macroeconomic outcomes of a program with the corresponding outcomes obtained under an alternative set of feasible policies is the most appropriate way of judging the effects of programs. However, the difficulties involved in using this criteria should not be underestimated. Criteria based on the determination of the counterfactual involve a degree of subjectivity by definition and are thus hard to employ in practice. Until we are able to develop suitable econometric techniques to estimate the counterfactual, any study of the quantitative effects of Fund-supported programs will not yield absolutely correct answers. In this context, the GEE approach, modified to allow for the endogeneity of the decision to enter into a program, appears most promising and has to be considered seriously in future evaluation exercises. By using this approach, one will go a long way towards estimating the "ideal" counterfactual.

The estimates from the GEE approach can also be improved upon by taking explicit account of the degree of implementation of the policies agreed to between the Fund and the program country. In the studies to date all programs have been treated alike, whether or not the policy intentions specified in the program were carried out. Mixing countries with varying implementation records in the sample may bias the judgement about effectiveness. Had the tests been restricted to only those countries that successfully implemented the recommended policies, it is conceivable that an even more positive picture would emerge. Future studies should attempt to separate successful implementers from other program countries, although one has to acknowledge that this is not always easy to do when there are waivers and modifications of programs that alter program targets and policies.

Case studies, as opposed to large multi-country studies, permit one to delve more deeply into program implementation.[19] However, case studies are an extremely time-consuming and expensive way to obtain evidence on program effects, and it is difficult to generalize from the findings of a few case studies. Large cross-country samples, in contrast, are more amenable to the application of standard statistical techniques, since there are a sufficient number of data points. But in the process, one loses some of the country-specific aspects of programs, including importantly the degree to which the policies in the program were implemented. Based on this reasoning, supplementing the results from cross-country studies with information from case studies highlighting some specific issues may well be a good compromise.

In conclusion, as long as Fund-supported programs are to be an integral part of the adjustment strategies of developing countries, the search for the most appropriate way to conduct evaluations of the effects of past programs must obviously continue. Much progress has been made to date, but more refinements of the empirical methodology employed are still necessary to increase the confidence with which one can answer the question of whether Fund-supported programs "work" in the context of cross-country empirical analyses.

---

[19]For an example of the case study approach to program evaluation, see Gomulka (1995) and Goldsbrough et al. (1996).

Methodology of Evaluating Fund-Supported Programs

This Appendix considers formally some of the methodological issues associated with estimating program effects.

Suppose that the j'th target variable in country i is determined according to:

$$y_{ij} = \beta_{oj} + x_i'\beta_j + w'\alpha_j + \beta_j^{IMF} d_i + \epsilon_{ij} \tag{1}$$

where y_{ij} is the target variable (e.g., the balance of payments, the current account, the inflation rate, and the growth rate), x_i is a vector of policy instruments (e.g., domestic credit, the exchange rate, the fiscal deficit), w is a vector of foreign exogenous variables (e.g., the terms of trade, international interest rates), d_i is a dummy variable, and ϵ_{ij} is a random shock. The parameter vectors β_{oj}, β_j and α_j are assumed for simplicity to be uniform across countries. The dummy variable d_i takes on the value 1 if a Fund-supported program is in effect during the period in question and zero otherwise, and the parameter β_j^{IMF} measures the effects of the program during this period on the variable y_{ij}.

It is important to note that this definition of β_j^{IMF} means that x_i refers to the policies that would be adopted in the absence of a program. The vector x_i is therefore directly observable only for nonprogram countries; for program countries the x_i must be estimated. One way in which x_i can be estimated is via the following simple reaction function:

$$\Delta x_i = \gamma'[y_i^d - (y_i)_{-1}] + \eta_i \tag{2}$$

where y_i is a vector of target variables, y_i^d is a vector of their desired values, γ is an adjustment parameter, η_i is a vector of random shocks, and Δ is a first-difference operator. Equation (2) basically says that the change in country i's macroeconomic policy instruments between the current and previous period will be a function of the difference between the desired values of the target variables this period and their actual values last period.

The model can be employed to examine the statistical properties of the before-after and with-without approaches to estimation of program effects. The *before-after* estimator β_j^{BA} is:

$$\beta_j^{BA} = \Delta y_{ij} \text{ for } i \in P \tag{3}$$

where P denotes the set of program countries during the current period. The expected value of this estimator, conditional on observed values of the foreign exogenous variables, is:

$$E\left(\beta_j^{BA} \mid i \in P, \Delta w'\right) = \overline{\beta_j^{IMF}} + \overline{\Delta w'}\alpha_j + E(\Delta x_i'\beta_j + \Delta \epsilon_{ij} \mid i \in P, \ \Delta w), \tag{4}$$

which is equal to the true value β_j^{IMF} only if:

$$E(\Delta x_i'\beta_i + \Delta \epsilon_{ij} \mid i \in P, \Delta w) = \overline{-\Delta w'}\alpha_j \tag{5}$$

That is, the before-after estimator is unbiased if one expects that the nonprogram determinants of y_{ij} would have behaved in such a way as to leave y_{ij} unchanged, on average, between the program and nonprogram periods. In other words, if any developments in the external markets, innovations in policies, and other unobserved variables cancel each other out.

The *with-without* estimator β_j^{WW} is given by:

$$\beta_j^{WW} = \Delta y_{ij} - \overline{\Delta y_{Nj}}, \tag{6}$$

where $\overline{\Delta y_{Nj}}$ is the average value of Δy_{ij} over some set N of nonprogram countries. Since we can observe Δx and $\Delta \epsilon_{ij}$ for all $i \in N$, the information set, defined as Ω, now consists of:

$$\Omega = \{(\Delta x_i, \Delta \epsilon_{ij} \text{ for } i \in N), \Delta w\}$$

Taking expectations of β_j^{WW} conditional on $i \in P$ and Ω we have:

$$E(\beta_j^{WW} \mid i \in P, \Omega) = \beta_j^{IMF} + E(\Delta x_i'\beta_j + \Delta \epsilon_{ij} \mid i \in P, \Omega) - (\overline{\Delta x_N'}\beta_j + \overline{\Delta \epsilon_{Nj}}) \tag{7}$$

The with-without estimator will be unbiased if:

$$E(\Delta x_i'\beta_j + \Delta \epsilon_{ij} \mid i \in P, \Omega) = \overline{\Delta x_N'}\beta_j + \overline{\Delta \epsilon_{Nj}}. \tag{8}$$

In other words, if it can be expected that, in the absence of the program, the country would have behaved just like the average member of the nonprogram reference group, then the estimator will be unbiased. Although the with-without estimator (unlike the before-after estimator) controls for the effects of changes in the global economic environment on target variables—because such global factors are assumed to affect program and nonprogram countries equally— it introduces a new source of bias, namely the characteristics of

nonprogram countries. If nonprogram countries differ systematically from program countries prior to the program period, then β_j^{ww} will be biased.

An alternative to the before-after and with-without estimators can be derived by using equation (2) to substitute out the unobservable policy changes that would occur in the absence of a Fund program (i.e., for x_i) from equation (1). The *generalized evaluation* estimator is:

$$\Delta y_{ij} = \beta_{oij} - (y_{ij})_{-1}(\gamma'\beta_j + 1) + (x_i')_{-1}\beta_j + w'\alpha_j + \beta_j^{IMF}d_i + (\epsilon_{ij} + \eta_i\beta_j), \qquad (9)$$

where the desired values y_{ij}^d have been subsumed into the constant.

Econometric estimation of equation (9) produces an estimate of β_j^{IMF} which is not subject to the criticisms leveled at the before-after and with-without estimators above. This equation takes care of the estimation of the counterfactual by controlling for the factors that are systematically related to the policies that would have been followed in the country without the program, that is by including the lagged values of target variables and policy instruments in the specification.

Equation (9) can be re-written as:

$$y_{ij} = Z'\Theta + \beta_j^{IMF}d_i + \xi_{ij} \qquad (10)$$

Where the predetermined, policy and other exogenous variables have been combined into a single vector Z. Θ defines the vector of parameters that are to be estimated.

If countries are randomly selected to enter Fund-supported programs then equation (10) can be estimated by OLS. However, if countries self-selected themselves into programs, then OLS will yield biased estimates.

All countries are assumed to have a continuous and measurable desire to participate in Fund-supported programs, F_i, that unfortunately is unobservable. However, it is possible to identify a program participation function which relates this F_i to observable variables as follows

$$F_i = W_{ij}'\varphi + \xi_{ij} \qquad (11)$$

Where W_{ij} is the vector of country characteristics that determine a country's decision to participate in a program. Although F_i is not observable we are able to only observe the discrete event of participation which occurs when F_i crosses some threshold.

$$d = 1 \qquad \text{if} \quad F_i > 0$$

$$= 0 \qquad \text{otherwise} \tag{12}$$

The disturbance terms in the participation and the target variables are assumed to be correlated.

$$E(Y_{ij} \mid d_i = 1)Z'\Theta + \beta_j^{IMF} + E(\xi \mid d_i = 1) \tag{13}$$

Using the conditional distribution of the normal density we can rewrite equation (13), conditional on participation in Fund-supported programs, as follows:

$$E(Y_{ij} \mid d_i) = Z'\Theta + \beta_j^{IMF} + \rho\,\sigma_\xi \frac{\phi(w'\varphi)}{\Phi(w'\varphi)} \tag{14}$$

Where Φ and ϕ are the standard normal distribution and density function.

Similarly, conditional on nonparticipation the expectation for a target variable is

$$E(Y_{ij} \mid d_i = 0) = Z'\Theta + \rho\,\sigma_\xi \frac{-\phi(w'\varphi)}{1 - \Phi(w'\varphi)} \tag{15}$$

The difference in expectations of the participants and the nonparticipants is:

$$E(Y_{ij} \mid d_i = 1) - E(Y_{ij} \mid d_i = 0) = \beta_i^{IMF} + \rho\,\sigma_\xi \frac{\phi(w'\varphi)}{1 - \Phi(w'\varphi)} \tag{16}$$

The above suggests the following three-step approach to estimation:

1.   Estimate equation (17) below by probit. This has additional advantage of allowing the participation in Fund-supported programs to be estimated.

$$d_i = W_{ij}'\varphi + \xi_{ij} \tag{17}$$

2.   Calculate the selectivity variable from the probit (λ) and add that as an additional variable in the generalized evaluation method estimations, i.e.,

$$y_{ij} = Z'\Theta + \beta_j^{IMF} + \lambda_{ij} + u_i \qquad (18)$$

3.   Calculate the impact of Fund-supported programs using equation (18) above and test for its significance. Heteroscedasticity corrections have to be made to obtain the correct variance matrix.

References

Bagci, P., and W. Perraudin, "Do IMF Programs Work?", Global Economic Institutions Working Paper, 1997.

Connors, T., "The Apparent Effects of Recent IMF Stabilization Programs", International Finance Discussion Paper 135, Board of Governors of Federal Reserve System, April 1979.

Conway, P., "IMF Lending Programs: Participation and Impact", *Journal of Development Economics* 45, December 1994, pp. 365-391.

————, "Evaluating Fund Programs: Methodology and Empirical Estimates", unpublished, May 1998.

Corbo, V., and P. Rojas, "World Bank-Supported Adjustment Programs: Country Performance and Effectiveness", in V. Corbo, S. Fischer and S. Webb, (eds.), *Adjustment Lending Revisited: Policies to Restore Growth*, World Bank: Washington, D.C., 1992.

————, and S. Fischer, "Structural Adjustment, Stabilization and Policy Reforms: Domestic and International Finance," in J. Behrman and T. N. Srinivasan (eds.), *Handbook of Development Economics* Vol. III, North-Holland, 1994.

Dicks-Mireaux, L., M. Mecagni and S. Schadler, "Evaluating the Effect of IMF Lending to Low-Income Countries", IMF unpublished, 1997.

Donovan, D., "Real Responses Associated with Exchange-Rate Action in Selected Upper Credit Tranche Stabilization Programs", *IMF Staff Papers* 28, December 1981, pp. 698-727.

————, "Macroeconomic Performance and Adjustment Under Fund-Supported Programs: The Experience of the Seventies", *IMF Staff Papers* 29, June 1982, pp. 171-203.

Faini, R., J. De Melo, A. Senhadji, and J. Stanton, "Growth-Oriented Adjustment Programs," *World Development* 19, August 1991, pp. 957-67.

Goldsbrough, D., S. Coorey, L. Dicks-Mireaux, B. Horvath, K. Kochhar, M. Mecagni, E. Offerdal, and J. Zhou, *Reinvigorating Growth in Developing Countries: Lessons from Adjustment Policies in Eight Economies*, International Monetary Fund Occasional Paper No. 139, 1996.

Goldstein, M., and P. Montiel, "Evaluating Fund Stabilization Programs with Multicountry Data: Some Methodological Pitfalls", *IMF Staff Papers* 33, June 1986, pp. 304-344.

Gomulka, S., "The Fund-Supported Programs of Poland and Russia, 1990-94: Principles, Errors and Results", *Journal of Comparative Economics* 20, June 1995, pp. 316-46.

Gylfason, T., "Credit Policy and Economic Activity in Developing Countries with IMF Stabilization Programs", *Princeton Studies in International Finance* 60, August 1987.

Johnson, O., and J. Salop, "Distributional Aspects of Stabilization Programs in Developing Countries", *IMF Staff Papers* 27, March 1980, pp. 1-23.

Khan, M. S., "The Macroeconomic Effects of Fund-Supported Adjustment Programs", *IMF Staff Papers* 37, June 1990, pp. 195-231.

Khan, M. S., and M. Knight, "Stabilization Programs in Developing Countries: A Formal Framework", *IMF Staff Papers* 28, March 1981, pp. 1-53.

————, *Fund-Supported Programs and Economic Growth*, International Monetary Fund Occasional Paper 41, 1985.

Khan, M. S., P. Montiel and N. Haque, (eds.), *Macroeconomic Models for Adjustment in Developing Countries*, International Monetary Fund: Washington, D.C., 1991.

Killick, T., *The Quest for Economic Stabilization: The IMF and the Third World*, New York, NY: St. Martin's Press, 1984.

Killick, T., and M. Malik, "Country Experiences with IMF Programmes in the 1980s", *World Economy 15*, September 1995, pp. 599-632.

Killick, T., M. Malik and M. Manuel, "What Can We Know About the Effects of IMF Programmes?", *World Economy* 15, September 1995, pp. 575-597.

Loxley, J., *The IMF and the Poorest Countries*, Ottawa, Canada: North-South Institute, 1984.

Pastor, M., "The Effects of IMF Programs in the Third World: Debate and Evidence from Latin America", *World Development* 15, February 1987, pp. 249-262.

Polak, J., "The Changing Nature of IMF Conditionality", *Princeton Essays in International Finance* 184, 1991.

Reichmann, T., and R. Stillson, "Experience with Programs of Balance of Payments Adjustment: Stand-by Arrangements in the Higher Credit Tranches, 1963-72", *IMF Staff Papers* 25, June 1978, pp. 119-139.

Santaella, J., "Four Decades of Fund Arrangements: Macroeconomic Stylized Facts before the Adjustment Programs", IMF Working Paper WP/95/74, 1995.

Schadler, S., F. Rozwadowski, S. Tiwari and D. Robinson, *Economic Adjustment in Low-Income Countries--Experience Under the Enhanced Structural Adjustment Facility*, International Monetary Fund Occasional Paper 106, 1993.

Sisson, C., "Fund-Supported Programs and Income Distribution in LDCs", *Finance and Development* 23, March 1986, pp. 33-36.

Williamson, J., (ed.), *Fund Conditionality*, Washington, D.C.: Institute for International Economics, 1982.

Zulu, J. and S. Nsouli, *Adjustment Programs in Africa: The Recent Experience*, International Monetary Fund Occasional Paper 34, 1985.