

Hutchison, Michael M.

Working Paper

Selection Bias and the Output Costs of IMF Programs

EPRU Working Paper Series, No. 2004-15

Provided in Cooperation with:

Economic Policy Research Unit (EPRU), University of Copenhagen

Suggested Citation: Hutchison, Michael M. (2004) : Selection Bias and the Output Costs of IMF Programs, EPRU Working Paper Series, No. 2004-15, University of Copenhagen, Economic Policy Research Unit (EPRU), Copenhagen

This Version is available at:

<https://hdl.handle.net/10419/82025>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

EPRU Working Paper Series

**Economic Policy Research Unit
Institute of Economics
University of Copenhagen
Studiestræde 6
DK-1455 Copenhagen K
DENMARK
Tel: (+45) 3532 4411
Fax: (+45) 3532 4444
Web: <http://www.econ.ku.dk/epru/>**

Selection Bias and the Output Costs of IMF Programs

Michael M. Hutchison

2004-15

ISSN 0908-7745

**The activities of EPRU are financed by a grant from
The National Research Foundation**

Selection Bias and the Output Costs of IMF Programs

June 2004

Michael M. Hutchison

Department of Economics
Social Sciences 1
University of California, Santa Cruz
Santa Cruz, CA 95064
Email: hutch@ucsc.edu

Abstract

Questions over the role of the IMF in the economic development and adjustment in developing countries have been the topic of intensive research and debate in recent years. Although most studies find that participation in an IMF program helps facilitate balance of payments adjustment, research in this area almost uniformly finds that growth is reduced at the same time (e.g. Bordo and Schwartz, 2000; Przeworski and Vreeland, 2000). In this paper we emphasize that the evaluation of the benefits and costs of participating in IMF-sponsored stabilization programs is complicated by the fact that countries typically enter into an agreement with the IMF only when facing dire economic problems. We argue that the sample selection bias is mainly responsible for the common perception that real output growth declines *because* countries choose to participate in IMF programs. This article uses four recently developed “matching” statistical methods (e.g. Heckman et al., 1997 and 1998; Rubin and Thomas, 1992; and others), based on the “selection on observables” bias, to estimate the growth effects of IMF program participation. In contrast with the extant literature, none of the matching method results (nearest neighbor, strata, radius and regression-adjusted) find an adverse growth effect. Rather, there is some evidence of a positive impulse to economic growth when countries entering IMF programs are compared to the appropriate counter-factual (i.e. non-participating countries with similar characteristics).

JEL: E63, F34, F41, O19

Research Assistance from Ilan Noy is gratefully acknowledged. Timothy Lane and Patricia Gillett graciously helped in obtaining data on IMF programs.

Section 1. Introduction

The relative value and effectiveness of participation in IMF-sponsored stabilization programs is the topic of intense debate in the academic literature and amongst policymakers (see Willett, forthcoming, for a review). This controversy seems especially heated for countries facing acute balance of payments problems and currency crises, as witnessed in 1997 in Korea, Indonesia and Thailand and elsewhere. Stiglitz (2000), for example, supports critics of the IMF who argue "...the IMF's economic 'remedies' often make things worse—turning slowdowns into recessions and recessions into depressions."

This article focuses on a key aspect in evaluating IMF programs that has not been adequately addressed-- the decision of a country to apply for participation in an IMF-sponsored program and the decision of the IMF to offer a loan are not random events. And if the participation decision is correlated with macroeconomic variables that are in turn correlated with output growth, then standard estimation techniques will yield a biased measure of the participation effect. Sample selection bias of this nature-- the measured effect reflecting systematic, unaccounted-for differences between countries that enter into and do not enter into IMF programs—is a serious problem since countries entering into IMF programs typically do so under very adverse economic conditions.

Most studies find sizeable declines in output growth arising from participation in IMF programs (e.g. Bordo and Schwartz, 2000; Hutchison, 2002; Hutchison and Noy, forthcoming; Przeworski and Vreeland, 2000; Vreeland, 2003). This article investigates whether sample selection bias into IMF programs is an important factor behind the common perception and dominant research finding that entering into an IMF program generally lowers economic growth. To date, the literature on IMF program evaluation uses the Inverse Mills Ratio (IMR) as an additional explanatory variable in the program

evaluation equation to explicitly account for sample selection bias. This methodology, developed by Heckman (1979) is designed to account for the “selection on unobservables” bias, but the results are typically weak in the IMF evaluation literature – not much change is observed in the participation coefficient and the IMR is generally insignificantly different from zero (e.g., Hutchison, 2002; Przeworski and Vreeland, 2000).

Selection bias, however, may be separated into a “selection on observables” as well as “selection on unobservables,” and it is likely in some cases—such as evaluating the outcomes of IMF programs—that the bias arising from the former is much more severe. Heckman, Ichimura and Todd (1997), for example, find that an unbalanced distribution of observables among participants and non-participants in a major US job training program was a much more important source of bias than the conventional selection problem. The method of matching was developed to help account for the estimation bias arising from the ‘selection on observables’ problem, and to date has mainly been applied in the medical and labor economics literature. (Persson, 2001, is an exception in applying matching methods to a macroeconomic data set investigating currency unions and trade growth). The basic idea is straightforward. Each participation observation is matched to a non-participation observation that has the same observed values of a vector of other characteristics that determine participation (X). Under certain standard assumptions, the difference in the observed outcome between the two matched observations is thus the program’s effect. As Heckman et al. (1997) state: “...simple balancing of observables in the participant and comparison group samples goes a long way toward producing a more effective evaluation strategy” (p. 607).

We evaluate the growth effects of participation in IMF-sponsored stabilization programs using several recently developed matching methods designed to deal with sample selection bias based on observables. In particular, we use “nearest neighbor”, “radius” and “stratification” matching methods, as well as a “regression-adjusted matching” estimator suggested by Heckman et al. (1997)—all methods designed to account for the “selection on observables” approach that has heretofore been neglected in the literature on IMF program evaluation. We find that the nature of non-random participation is a central problem in the evaluation of IMF programs and, after rigorously control for the selection on observables bias (and obtaining unbiased estimates), we find that IMF-program participation is *not* associated with adverse real growth effects. Rather, there is some evidence- albeit not strong—that participation in an IMF program generally leads to a positive growth impulse when compared to the appropriate counter-factual (i.e. the non-participation group of countries with similar characteristics). This result is completely at odds with the conventional view of the effects of IMF programs and may explain why countries routinely choose to participate in an IMF stabilization program if it were not in their best interests to do so.

Section 2 discusses the matching methodology in more detail and its application to the problem at hand. Section 3 discusses the data—selection of the appropriate IMF programs, sample of countries, and variables used in the propensity score (likelihood of participating in an IMF program) and treatment effect (growth) equations. Section 4 presents the preliminary empirical results—calculation of the propensity scores and estimation of the growth equations used for the regression-adjusted matching methods.

Section 5 presents the main empirical results. Section 6 concludes the paper and compares the results with the existing literature.

Section 2. Matching Methodology

The advantage of matching methods is that they address non-random sample selection and, as a non-parametric statistical method, avoid strong assumptions about functional form.¹ To examine the average effect of IMF program participation on output growth we employ three matching algorithms--nearest neighbor, stratification and radius matching—as well as two regression-adjusted matching estimates. These are different approaches to matching observations with *similar characteristics* except that one group participates in an IMF program and the other does not. Following the matching of observations, the difference in output growth between the two groups is measured. This is termed the “treatment effect”.

In order to assess similarity among countries, we investigate a list of observable characteristics. One approach would be to match each participation observation to a non-participation observation that has exactly the same observed values of a vector of other characteristics that determine participation (X). In macroeconomic studies, where the size of the sample is typically limited, this matching method is difficult or impossible to implement. Fortunately, Rosenbaum and Rubin (1983 and 1985) have shown that, if the probability of participation – $P(X)$ – is known then matching by $P(X)$ instead of X is sufficient. This collapses the multidimensional problem of matching to one dimension based on the estimated probabilities or *propensity scores* and greatly simplifies the

¹ See Persson (2001) for an excellent review of the methodology and an application with macroeconomic data. .

procedure. Rubin and Thomas (1992) show that using an estimated probability of participation $\tilde{P}(X)$, based on the list of observable characteristics, instead of $P(X)$ still reduces selection-on-observables bias. When countries have a similar propensity score, they are paired according to one of the three matching criteria, and the treatment effect is measured. The two steps to the matching method is to first estimate the propensity score and then to estimate the treatment effect.

The *nearest neighbor* approach matches each participation observation to the non-participation observation that has the nearest propensity score. After each non-participation observation is used it is ‘returned’ to the pool of non-participation observations. A simple average of the paired matches is then computed. The *radius* approach matches each participation observation to the average of all the non-participation observations with propensity scores falling within a pre-specified radius from the propensity score of the participation observation.² An average of the difference is obtained with weighting according to the number of non-participation observations used in the construction of each matched pair. The *stratification* approach divides the sample into five groups (strata) based on their propensity scores. Within each strata, the average of the participation observations is matched with the average of the non-participation observations. An average of the five strata, weighted by the number of participation observations in each one, is then calculated.

The first two algorithms are implemented in Dehejia and Wahba (2002), while a version of the third one is used by them to develop an algorithm for estimating the

² In our case, and following Persson (2001), we chose a 0.05 radius so that each participation observation with an estimated propensity score ρ is matched with all the non-participation observations whose propensity scores (q) satisfies the condition $\rho-0.05 < q < \rho+0.05$.

propensity scores. All three are implemented in Persson (2001). In all three cases, weighted standard errors are constructed using weighted averages as is given in the appendix of Persson (2001).

Rubin (1979) and Heckman et al. (1997, 1998) also implement a *regression-adjusted matching estimator*. While the previous matching methodologies do not impose any structure on the output growth equation, biases could result from omitted variables that are correlated with both the outcome (output growth) and treatment (participating in an IMF program). Both consistency and efficiency may be improved by implementing a regression-adjusted estimator. Rubin (1979) suggests that the output regression should contain all observations while Heckman et al. (1998) conclude that estimation using only the non-participation observations is preferable. We employ both methods. From these regressions we obtain the residuals for each country-year observation. These residuals are then used for our three matching algorithms.

Section 3. Selection of IMF Programs and Data Description

The objective of this study is to evaluate the impact on output growth arising from participating in an IMF-sponsored stabilization program. To this end, we focus on the two IMF facilities designed to meet short-run balance of payments stabilization-- Standby Arrangements (SBA) and the enhanced fund facility (EFF).³ In general, Fund members can access credit tranches from the General Resources Account (GRA) either by means of IMF program arrangements or by means of “outright purchases.” Outright purchases are limited, typically, for the first 25% of the member’s quota and do not involve any

³ See Hutchison (2003) for a complete discussion of IMF programs.

phasing or conditionality. Stand-by arrangements have been the main instrument through which members gain access to further credit tranches.⁴ Stand By Arrangements (SBA) typically last for 12-18 months (the legal maximum is 3 years) and first tranche drawings do not require strict conditionality. Any drawings beyond the first tranche require both phasing out and stricter conditionality and are limited to 100% of quota annually (300% cumulatively together with the Extended Fund Facility, EFF, as discussed below). Repurchase obligations last about 3 ¼ - 5 years from the date of purchase.

The Extended Fund Facility, established in 1974, provides somewhat longer-term financing to countries in need of structural economic reforms. EFF arrangements typically last for 3 years; phasing and conditionality are similar to the SBAs with an emphasis on longer-term structural reforms. Quota limits are identical to the SBAs while repurchases last much longer (4½ - 10 years). Both facilities are subject to the same rate of interest for repayments.⁵ The Supplemental Reserve Facility (SRF), introduced in 1997 in the Korean stabilization program, aims to supplement resources made available under SBAs and the EFF in order to provide financial assistance for exceptional balance of payments difficulties. Penalty interest rates (increasing over time) and short repayment periods (1–1½ years) insure that these are taken only in exceptional circumstances.⁶

We use the SBA and EFF programs (and, for Korea in 1997, the new SRF program) as our definition of “IMF-supported stabilization programs” since these are the only programs clearly linked to short-term balance of payments adjustment. (There are no cases of SBA and EFF programs being approved in the same year in this data sample).

⁴ As the Articles of Agreement state, they were defined as “a decision by the Fund by which a member is assured that it will be able to make purchases from the General Resources Account in accordance with the terms of the decision during a specified period and up to a specified amount” (Article XXX (b)).

⁵ Starting in 1989, the rate of charge was linked directly to the SDR interest rate, and adjusted weekly.

By contrast with these programs, some Fund facilities are directed primarily toward structural reform and poverty reduction and are not included in our study.

Middle-to-high income developing countries over 1975-97 are investigated since this group (twenty-five emerging market countries) is only eligible for short-run stabilization funds from the IMF and a broad set of macroeconomic data is available. This set of countries are not eligible for the IMF concessionary loan programs. A broad set of macroeconomic data availability is necessary to estimate the IMF-program participation and output growth equations with a sufficient number of control variables.

⁶ In our sample, the only such case is the agreement with Korea in 1997.

Section 4. Preliminaries: Estimating Propensity Score and Output Growth Equations

A. Propensity Scores

In controlling for sample selection bias, a probit equation explaining the likelihood of IMF-program participation is estimated to calculate propensity scores. Our selection of potential variables is guided by previous literature in this area, especially Knight and Santaella (1997) who test a number of supply side (e.g. willingness of the IMF to approve programs) and demand side (e.g. demand of a particular country for IMF credits) determinants. This literature demonstrates that entering into an IMF agreement is not random, but guided by "...a clear set of observable economic factors that are strongly correlated with the event of approval of a financial arrangement." (p. 431). In particular, Knight and Santaella find that a low level of international reserves, low per capita GDP, high ratio of external debt service (to export earnings), movements in the real exchange rate, weak GDP growth and a low rate of domestic investment induce countries to seek an IMF-supported program. Policy measures to enhance fiscal revenues, reduce government expenditures, to tighten domestic credit, and to adjust the exchange rate are significant factors likely to win IMF approval of programs.

We are interested in a reliable prediction equation, and therefore estimate a reduced form equation rather than identify separately the determinants of the supply and demand of IMF programs. In addition to the variables noted above, we investigate whether (lagged) foreign exchange reserves to imports ratio, the change in the current account to GDP ratio, and real *per capita* GDP growth are reliable predictors of IMF-program

participation. These macroeconomic data series are taken from the International Monetary Fund's IFS CD-ROM.

The results are shown in table 1. Countries with high foreign debt and facing a currency or balance-of-payments crisis are more likely to enter into an IMF program. Countries with improving fiscal positions and higher foreign exchange reserves (relative to imports) are less likely to enter into IMF programs. The model correctly predicts over 80 percent of the 412 observations, where a correct prediction is defined as a propensity score of over (less than) 50 percent corresponding with IMF-program participation (non-participation). Other lagged values were investigated but did not add explanatory power to the model.

B. Output Growth Equations

In order to implement Rubin (1979) and Heckman et al.'s (1997, 1998) regression-adjusted matching estimators, it is necessary to first specify an equation that controls for the factors—other than IMF program participation—that may influence the evolution of GDP growth. To this end we employ a benchmark model based on a reduced form equation, derived in turn from an output equation and a policy reaction function, termed the General Evaluator Estimator by Goldstein and Montiel (1986). Recent applications of this framework include Dicks-Mireaux et al. (2000), Hutchison (2003) and Hutchison and Noy (2003). A complete description of the basic model, model derivation, and definitions of the explanatory variables may be found in Hutchison (2003).

The dependent variable estimated is the difference in the real GDP growth rate. The explanatory variables in the model are contemporaneous external output growth rates,

contemporaneous and lagged currency crises; lagged values for the change in the budget surplus to GDP ratio, inflation rates, credit growth, real exchange rate overvaluation, real GDP growth; and dummy variables for Asia and Latin America.

The indicator of currency and balance of payments crises is constructed from “large” changes in an index of currency pressure, defined as a weighted average of monthly real exchange rate changes and monthly (percent) reserve losses.⁷ Following convention (e.g. Kaminsky and Reinhart, 1999) the weights are inversely related to the variance of changes of each component over the sample for each country. An episode of severe exchange rate pressure is defined as a value in the index—a threshold point-- that exceeds the mean plus 2 times the country-specific standard deviation, provided that it also exceeds 5 percent.⁸ The first condition insures that any large (real) depreciation is counted as a currency crisis, while the second condition attempts to screen out changes that are insufficiently large in an economic sense relative to the country-specific monthly change of the exchange rate. For each country-year in our sample, we construct binary measures of currency crises, as defined above (1 = crisis, 0 = no crisis).

The external growth rate measure is the trade-weighted average growth rate of the country’s major trading partners. Real exchange rate overvaluation is defined as deviations from a fitted trend in the real trade weighted exchange rate. The real trade-weighted exchange rate is the trade-weighted sum of the bilateral real exchange rates (defined in terms of CPI indices) against the U.S. dollar, the German mark, and the Japanese yen. The trade-weights are based on the average bilateral trade with the United

⁷ Our currency pressure measure of crises does not include episodes of defense involving sharp rises in interest rates. Data for market-determined interest rates are not available for much of the sample period in many of the developing countries in our dataset.

States, the European Union, and Japan in 1980 and 1990. Definitions of the other variables are conventional.

We use annual observations in our analysis. While we employ monthly data for our (real) exchange rate pressure index to identify currency crises and date each by the year in which it occurs, using annual data enables inclusion of a relatively large number of countries.

Section 5. Real Output Effects of IMF Programs using Matching Methods

A. Trimmed Regression Estimates

As a preliminary analysis, we compare the standard GEE approach to estimating the effects of IMF-program participation (using an unrestricted sample) to that with a sample “trimmed” by discarding those observations that are outside the common support of the participating and non-participating observations. The basic idea is to exclude those observations where a country, in a particular year, is either very likely or very unlikely to participate in an IMF program. Using the propensity scores, matched against actual program participation, the common support criteria involves excluding 55 non-participation observations that have very low estimated propensity scores (<0.079) and 4 participation observations that have the highest estimated propensity scores (>0.82).

The first column of Table 3 reports the standard GEE regression using the unrestricted sample. The baseline model follows Hutchison (2002), and explains output

⁸Other studies also define the threshold of large changes in terms of country-specific moments. While the choice of cut-off point is somewhat arbitrary, Frankel and Rose (1996) suggest that the results are not very

growth as a function of a number of control variables and an IMF participation dummy variable (coded one if the country in question was in an IMF program at a given year). The estimate on the IMF dummy indicates that output growth is reduced by 68 percentage points for each year of program participation. This compares with a reduction of 78 percentage points, reported in Hutchison (2003) and Hutchison and Noy (2003), for a broader sample of developing countries. This estimate is similar in magnitude to Conway (1994), and substantially less than the output loss estimated by Bordo and Schwartz (2000) and Przeworski and Vreeland (2000).

The second column of Table 3 reports estimates from the trimmed sample, based on the propensity scores as an exclusionary criteria. The coefficient estimate on the IMF dummy drops in absolute value to -0.38 and is no longer statistically significant at conventional levels. This result suggests that the previously estimated negative effect of IMF programs may be the result of selection-on-observables bias and not from adverse effects from IMF-program participation.

B. Unconditional Matching Results

The second part of the analysis utilizes nearest neighbor, radius and strata matching methods to evaluate the effects of IMF programs. These results are shown in table 4. The propensity scores are derived from the probit equations. We term these results “unconditional matching” since the mean values of output growth are compared for the participating and non-participating observations, i.e. differences in output growth are not conditioned on any explanatory factors. Panel A presents the results of differences in output growth one year following the approval of the IMF program. Panel B presents the

sensitive to the precise cut-off chosen in selecting crisis episodes.

results comparing the difference in average output growth for three years following the approval of the IMF program.

None of the results point to any significant differences in output growth performance between IMF program participation and non-participation. Output growth is higher for IMF-participation observations using the nearest neighbor (1.31-1.85 percent) and radius methods (0.77 percent), though not statistically significant at conventional levels, and virtually identical using the strata matching method. Though not statistically significant at conventional levels, it is interesting that most of the results show “large” and positive growth effects from participating in IMF programs—with the largest positive one-year effect being an increase of output by 1.3 percentage points. This is weak statistical evidence of a positive effect of IMF programs, but is noteworthy given the contrast with the extant literature, and lack of statistical significance is not unexpected since the matching procedure discards many observations (see Persson, 2001, for a discussion).

C. Regression-adjusted Matching Estimators

The next matching method employed compares differences in “conditional” output growth, i.e. the residuals from output growth equations shown in table 2 using all the observations (Rubin method) and using only non-program observations (Heckman method). These regressions explain output growth using the standard control variables, and the residuals (“unexplained output growth”) is compared for program participation and non-participation observations using the three matching methods. These results are shown in table 5 for output growth based on the Rubin method and in Table 6 for the Heckman method. Panel A (in both tables) presents the difference in output growth for

the year following the IMF program, Panel B presents the difference in *growth rate differences* between program and non-program observations, and Panel C presents the difference in average growth rates for the three years following the IMF program.

In no case are the differences (or difference in differences) statistically significant, but most are positive. There is no evidence that the impact of an IMF program is negative.

Section 6. Conclusion

Most studies find that the estimated cost of an IMF stabilization program, in terms of foregone output growth, is about 0.7-2.5 percentage points during each year of program participation. This paper argues that the issue of sample selection bias has not been adequately addressed in this literature and proposes several trimming and matching methods using propensity scores to investigate whether there is a significant difference in growth performance between years of IMF program participation and non-participation.

The first part of the analysis considers a standard output regression used in the literature (“baseline model”) to analyze the effects of IMF program participation. The standard result is obtained, suggesting that IMF program participation is associated with low output growth. The baseline model is compared with an identical model other than the sample of observations was “trimmed” to exclude those observations where a country was either very likely or very unlikely to participate in an IMF program. These results do not suggest any reduction in output growth for those countries participating in an IMF program. The second and main part of the analysis uses alternative forms of matching methods to investigate the difference in output performance between participation and

non-participation in IMF programs. There is no evidence that IMF program participation imposes output costs on countries. Rather, most estimates indicate a positive though statistically insignificant output growth effect from participation in IMF programs.

References

- Bird, G., 2002. The completion rate of imf programmes: what we know, don't know and need to know. *World Economy* 25(6), 833-47.
- Bird, G., Hussain, M. and Joyce J., forthcoming. Many happy returns? Recidivism and the IMF. *Journal of International Money and Finance*.
- Blundell, R. and Costa Días M., 2000. Evaluation methods for non experimental data. *Fiscal Studies* 21(4), 427-468.
- Bordo, M., Schwartz, A., 2000. Measuring real economic effects of bailouts: Historical perspectives on how countries in financial distress have fared with and without bailouts. NBER Working Paper 7701.
- Boughton, J. M., 2000. From Suez to Tequila: The IMF as a crisis manager. *Economic Journal* 110(460), 273-291.
- Conway, P. 1994. IMF lending programs: Participation and impact. *Journal of Development Economics* 45(2), 365-391.
- Dehejia, R. H., Wahba S., 2002. Propensity score matching methods for non-experimental causal studies. *Review of Economics and Statistics* 84(1), 151-161.
- Dicks-Mireaux, L., Mecagni M., Schadler S., 2000. Evaluating the effect of IMF lending to low-income countries. *Journal of Development Economics* 61(2), 495-526.
- Glick, R., Hutchison, M., 2001. Banking and currency crises: how common are twins? In: R. Glick, R. Moreno, Spiegel, M., (Eds.) *Financial Crises in Emerging Markets*. Cambridge, UK: Cambridge University Press.
- Goldstein, M., Montiel P., 1986. Evaluating Fund stabilization programs with multi-country data: Some methodological pitfalls. *IMF Staff Papers* 33(2), 304-44.
- Gylafson, T., 1987. Credit policy and economic activity in developing countries with IMF stabilization programs. *Princeton Essays in International Finance* 60.
- Heckman, J., 1979. Sample selection bias as a specification error. *Econometrica* 47(1), 153-161.
- Heckman, J., Ichimura, H., Todd, P., 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies* 64(4), 605-654.
- Heckman, J., Ichimura, H., Todd, P., 1998. Matching as an econometric evaluation estimator. *Review of Economic Studies* 65(2), 261-294.

Hutchison, M., 2003. A cure worse than the disease? Currency crises and the output costs of IMF-supported stabilization programs. In: Dooley, M. P., Frankel, J. A., (Eds.) *Managing Currency Crises in Emerging Markets*. Chicago, University of Chicago Press.

Hutchison, M., Noy, I. 2003. Macroeconomic effects of IMF-sponsored programs in Latin America: Output costs, program recidivism and the vicious cycle of failed stabilizations. *Journal of International Money and Finance* (December) 22, 991-1014.

Joyce, J., 2001. Time present and time past: A duration analysis of IMF programs spells. Federal Reserve Bank of Boston Working Paper 01-02. Forthcoming in *Review of International Economics*.

Khan, M. 1990. The macroeconomic effects of Fund-supported adjustment programs. *IMF Staff Papers* 37(2), 195-231.

Pastor, M. 1987. The effects of IMF programs in the Third World: Debate and evidence from Latin America. *World Development* 15(2), 249-262.

Persson, T. 2001. Currency Unions and Trade: How Large is the Treatment Effect? *Economic Policy* (October), 435-448.

Przeworski, A., Vreeland, J., 2000. The effect of IMF programs on economic growth. *Journal of Development Economics* 62(2), 385-421.

Rosenbaum, P.R., Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41-50.

Rosenbaum, P.R., Rubin, D. B., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39(1), 33-38.

Rubin, D. B., Thomas, N., 1992. Characterizing the effect of matching using linear propensity score methods with normal distributions. *Biometrika* 79(4), 797-809.

Ul Haque, N., Khan, M. S., 1998. Do IMF-supported programs work? A survey of the cross-country empirical evidence. *IMF Working Paper* WP/98/169.

Willett, T. (forthcoming). Understanding the IMF Debate. *The Independent Review: A Journal of Political Economy*.

Table 1 - Probit Equation for Estimating Propensity Scores
 Dependent Variable: onset of short-term IMF programs

Capital Formation	-0.016 (0.87)
Debt to GDP Ratio	1.162*** (3.10)
Debt Service to Imports Ratio	-0.021* (1.92)
Change in budget surplus to GDP ratio (t-1)	-5.863* (1.78)
Foreign Exchange to Imports Ratio (t-1)	-0.640* (1.73)
Other IMF programs (t)	-0.025 (1.42)
Currency crises dummy (t)	0.750*** (3.20)
Asia Dummy	-0.121 (0.17)
Latin America dummy	0.124 (0.20)
Observations	413
Chi Squared	110.28
Percent of Correctly Predicted Observations	81

Table 2 - Growth Equations for Estimating Residuals
 Dependent Variable: difference in real GDP growth rate

	Heckman Residuals (no-program observations only)	Rubin Residuals (all observations)
Change in budget surplus to GDP ratio (t-1)	10.420 (1.49)	5.146 (0.75)
Inflation Rate (t-1)	-0.020 (-1.61)	-0.031*** (-2.74)
Credit growth (t-1)	0.017 (1.53)	0.025** (2.36)
External growth rates (t)	0.232* (1.89)	0.264** (2.14)
Real exchange rate overvaluation (t-1)	-0.058*** (-5.00)	-0.045*** (-3.91)
Currency crises dummy (t-1)	-1.491** (-2.33)	-2.351*** (-4.15)
Currency crises dummy (t)	-2.039*** (-3.34)	-2.860*** (-5.27)
Real GDP growth (t-1)	-0.791*** (-19.45)	-0.816*** (-19.98)
Asia dummy	1.012 (1.61)	1.141** (1.98)
Latin America dummy	-1.187** (-2.04)	-0.960* (-1.81)
Observations	371	455
Adjusted R ²	0.54	0.48
Estimated correlation of error terms	0.05	0.02

Table 3 - Growth Equations with Trimmed Sample

Dependent Variable: difference in real GDP growth rate

	(1) All Sample	(2) Trimmed Sample
Change in budget surplus to GDP ratio (t-1)	5.252	7.653
Inflation Rate (t-1)	-0.028***	-0.039**
Credit growth (t-1)	0.024***	0.038***
External growth rates (t)	0.261**	0.229
Real exchange rate overvaluation (t-1)	-0.047***	-0.033**
Short-term IMF program dummy (t)	-0.680*	-0.383
Currency crises dummy (t-1)	-2.253***	-2.060***
Currency crises dummy (t)	-2.887***	-2.702***
Interactive dummy (CC t or t-1 and IMF at t)		
Real GDP growth (t-1)	-0.819***	-0.812***
Latin America dummy	-0.814	-1.111
Asia dummy	1.192**	0.945
Observations	455	236
Adjusted R ²	0.48	0.48
Estimated correlation of error terms	0.02	0.01

Table 4

Unconditional Matching

A. Unconditional Output Growth in Program Year

	Nearest-Neighbor	Radius (<0.05)	By Strata
Estimated effect of Short-term IMF programs	1.31	0.44	-0.14
t-statistics	1.46	0.72	0.20
Number of observations	154	254	254

B. Unconditional Output Growth in 3 Years following Program

	Nearest-Neighbor	Radius (<0.05)	By Strata
Estimated effect of Short-term IMF programs	1.85	0.86	-0.06
t-statistics	1.05	0.77	0.04
Number of observations	140	216	216

Table 5

Matching Rubin Residuals

A. Conditional Output Growth in Program Year

(Rubin Regression Residuals)

	Nearest-Neighbor	Radius (<0.05)	By Strata
Estimated effect of Short-term IMF programs	0.29	-0.01	0.09
t-statistics	1.25	0.04	0.50
Number of observations	142	236	236

B. Conditional Output Growth in Program Year

(Diff-Diff Rubin Regression Residuals)

	Nearest-Neighbor	Radius (<0.05)	By Strata
Estimated effect of Short-term IMF programs	0.13	-0.13	-0.11
t-statistics	0.43	0.64	0.47
Number of observations	134	229	229

C. Conditional Output Growth in 3 Years following Program

(Rubin Regression Residuals)

	Nearest-Neighbor	Radius (<0.05)	By Strata
Estimated effect of Short-term IMF programs	0.31	0.36	0.17
t-statistics	0.80	1.58	0.57
Number of observations	116	188	188

Table 6

Matching Heckman Residuals

A. Conditional Output Growth in Program Year

(Heckman Regression Residuals)

	Nearest-Neighbor	Radius (<0.05)	By Strata
Estimated effect of Short-term IMF programs	-0.53	0.31	0.01
t-statistics	0.62	0.57	0.01
Number of observations	138	236	236

B. Conditional Output Growth in Program Year

(Diff-Diff Heckman Regression Residuals)

	Nearest-Neighbor	Radius (<0.05)	By Strata
Estimated effect of Short-term IMF programs	-0.19	0.32	0.69
t-statistics	0.14	0.44	0.77
Number of observations	138	229	229

C. Conditional Output Growth in 3 Years following Program

(Heckman Regression Residuals)

	Nearest-Neighbor	Radius (<0.05)	By Strata
Estimated effect of Short-term IMF programs	-1.20	=1.36	-0.57
t-statistics	0.84	-1.60	0.52
Number of observations	116	188	188