

IMF Working Paper

Finance Department

**Repeated Use of IMF-Supported Programs:
Determinants and Forecasting**

Prepared by Martin Iseringhausen, Mwanza Nkusu, and Wellian Wiranto

Authorized for distribution by Olaf Unteroberdoerster

November 2019

***IMF Working Papers* describe research in progress by the author(s) and are published to elicit comments and to encourage debate.** The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Abstract

This paper studies the determinants of repeated use of Fund-supported programs in a large sample covering virtually all General Resources Account (GRA) arrangements that were approved between 1952 and 2012. Generally, the revolving nature of the IMF's resources calls for the temporary support of member countries to address balance of payments problems while repeated use has often been viewed as program failure. First, using probit models we show that a small number of country-specific variables such as growth, the current account balance, the international reserves position, and the institutional framework play a significant role in explaining repeated use. Second, we discuss the role of IMF-specific and program-specific variables and find evidence that a country's track record with the Fund is a good predictor of repeated use. Finally, we conduct an out-of-sample forecasting exercise. While our approach has predictive power for repeated use, exact forecasting remains challenging. From a policy perspective, the results could prove useful to assess the risk IMF programs pose to the revolving nature of the Fund's financial resources.

JEL Classification Numbers: F33, F34, F42

Keywords: IMF lending, Fund resources, GRA arrangements

Authors' E-mail Addresses: martin.iseringhausen@ugent.be; mnkusu@imf.org;
wellianwiranto@gmail.com

When this paper was written, Martin Iseringhausen and Wellian Wiranto were in the IMF's Finance Department as summer intern and economist, respectively. The authors would like to thank Thomas Krueger, David Moore, and Olaf Unteroberdoerster for very useful discussions and suggestions. They also thank Ling Zhu, Anta Ndoye, Bikas Joshi, Ursula Wiriadinata, Seyed Reza Yousefi, and the participants of an IMF Finance Department Seminar and a Ghent University Macroeconomics, Policy, and Econometrics Seminar for their helpful comments. All remaining errors are the authors' responsibility.

Contents

I	Introduction	3
II	Related literature	5
III	Data	7
	A Variables	8
	B Stylized facts	10
IV	A probit model with missing observations	12
	A Overview of the model	12
	B Goodness of fit and performance evaluation	14
V	Estimation results	15
	A Baseline regressions	15
	B Expanded regressions	18
	C Robustness checks	20
VI	Out-of-sample forecasting	23
	A Design and limitations	23
	B Results	24
VII	Conclusion	25
	References	27
	Annex 1: Figures and Tables	28
	Annex 2: Other Dataset Figures and Tables	39

I Introduction

The Articles of Agreement of the International Monetary Fund (IMF) provide that Fund resources should be made available to member states temporarily and under adequate safeguards to help them address balance of payments (BoP) problems. The temporary provision of IMF lending is meant to ensure the revolving nature of Fund resources (IMF, 2018).¹ However, over the years numerous members have repeatedly requested the IMF's support and, in most cases, successor programs started shortly after the initial programs ended. Several countries have resorted to multiple successor arrangements. While policy adjustments and reforms needed in order to solve deep-rooted structural and macroeconomic problems can require more than one arrangement, extensive repeated use of Fund resources can indicate program failure (see e.g. Easterly, 2005). With the member's BoP problem remaining unsolved, it would also be an indicator of higher risks to the revolving nature of the Fund's financial resources.²

What factors lead countries to become repeat users of Fund resources? From a policy perspective, it is important to understand the drivers of programs that end up helping member countries resolve their balance of payments difficulties in a way that avoids the need for a successor arrangement. Therefore, this paper asks the following questions: What is the difference between programs that have not been followed by a successor arrangement compared to those that have? Can we identify country-specific and/or global variables that relate to the probability of (no) repeated use of Fund resources? Which role do program-specific characteristics as well as a country's history with the IMF play for the risk of repeated use? Finally, provided that relevant drivers can be identified, can (no) repeated use be predicted?

This study attempts to answer these questions focusing on non-concessional arrangements, which are those financed with resources from the IMF's General Resources Account (GRA). Concessional arrangements under the Fund's Poverty Reduction and Growth Trust (PRGT) are excluded from the analysis. The characteristics of low-income countries eligible for assistance under the PRGT tend to differ from other countries in terms of their underlying balance of payments problems and resulting financing needs. Focusing on strong and durable poverty reduction and growth, PRGT-supported programs are generally expected to foster progress addressing (as opposed to resolving) a member's BoP problems, making repeated use a generally available option (IMF, 2018).³ However, some of the results likely also have implications for PRGT-supported programs.

¹The paper uses the terms *program* and *arrangement* interchangeably.

²Throughout the paper, we use *Fund* and *IMF* interchangeably.

³Also, PRGT-supported programs are meant to remove certain liquidity and credit risks from the GRA's balance sheet (IMF, 2018).

Our econometric analysis is based on a binary response model. We use a sample of almost 900 GRA arrangements from 1952 to 2012. First, we define the dependent variable as the absence of a successor program within a certain period following the original program. Second, using probit models, we link the probability of ‘no successor arrangement’ to a large set of variables characterizing the domestic economy, the global economic environment, a country’s history with the Fund, and program-specific features. Third, we assess the out-of-sample forecasting performance of selected model specifications.

Our contribution to the existing literature is twofold. First, to the best of our knowledge this is the first study that looks at repeated use of IMF lending using a large sample that includes virtually all disbursing GRA-supported programs since 1952. The large sample allows us to identify changes in the relative importance of different explanatory variables over time. However, it requires an econometric method that achieves efficient estimation in the presence of missing observations and avoids dropping programs from the sample. Second, while previous studies have focused on in-sample analysis, we also study the forecastability of repeated use and highlight its importance for policy applications.

Our findings can be summarized as follows: First, a small number of country-specific variables such as growth, the current account balance, the international reserves position, trade openness, and the quality of institutions are significant determinants of the probability of repeated use. Second, results on the impact of the global economic environment on the probability of repeated use are ambiguous. Third, some IMF-specific and program-specific variables can help explain the occurrence of successor arrangements. In particular, a country’s track record with the Fund appears to be a relevant factor for the risk of repeated use. Finally, parsimonious model specifications including only the most important determinants of repeated use, have reasonable out-of-sample predictive power. However, while exact forecasts remain inherently challenging, due to unforeseen shocks and uncertainty surrounding macro-forecasts and assumptions about policy implementation, a carefully selected model specification could serve as a valuable tool to gauge risks of repeated use in general and to the revolving nature of the Fund’s resources in particular.

The remainder of the paper is structured as follows: Section II reviews the related literature. Section III discusses the sample of GRA programs, defines the dependent variable and the explanatory variables, and presents some stylized facts. Section IV introduces a probit approach for dealing with missing observations. Section V presents the estimation results and Section VI presents an out-of-sample forecasting exercise. Section VII concludes.

II Related literature

This paper relates to at least four strands of the literature on IMF programs. First, there exist numerous papers that have looked at repeated and prolonged use of IMF resources and examined whether repeated and temporary users have distinct differences. Bird (2004) reviews the earlier literature and summarizes quantitative evidence on the determinants of prolonged use. Likely drivers include low real growth, large current account deficits, a high debt burden, low levels of international reserves, and a weak institutional framework. Bird et al. (2004) model the total number of programs requested by 90 developing countries over the period 1980 to 1996. They show that, on average, frequent borrowers have weaker macroeconomic characteristics compared to temporary users. Marchesi and Sabani (2007) interpret prolonged and repeated recourse to IMF resources as arising from an alleged distortion of the IMF's lending practice towards laxity in addressing non-compliance with conditionality to protect its own reputation because, besides its role as creditor, the IMF is also an advisor on reforms. Easterly (2005) finds that repeated use of IMF programs is an indication of program failure as even multiple programs do not succeed in solving macroeconomic problems in a particular country.

Second, this paper complements studies that analyze the determinants of program success. In a study closely related to ours, Larch et al. (2017) try to identify the underlying drivers of program success in a sample of 176 GRA-supported programs from 1993 to 2011. The authors define a program as successful if post-program levels of economic activity and the general government debt are at least comparable to pre-program levels. When regressing their binary success indicator on a wide range of domestic and global variables, they find a country's initial growth, the initial fiscal situation as well as its adjustment over the program period, trade openness, the exchange rate regime and the global economic environment to be important drivers of program success. They also conclude that some program-specific characteristics such as the agreed number of structural adjustments targeting supply side disruptions somewhat increase the chance of program success, while others such as program size or the number of so-called quantitative performance criteria (QPC) do not have significant explanatory power. Ivanova et al. (2001) apply a more direct definition of program success. They measure success by the fraction of implemented economic policy and structural conditions under a program and whether there have been program interruptions. They focus on the role of institutions on program success in a sample of 170 GRA- and PRGT-supported programs between 1992 and 1998. In line with previous research on World Bank programs (Dollar and Svensson, 2000), the authors find a few institutional variables such as bureaucracy, political cohesion, and government stability to be major drivers of program success.

Beyond the papers discussed above which look at specific definitions of program success, this paper relates to a large body of literature evaluating more generally whether IMF-supported programs have been successful in promoting sustained long-term growth. Examples of contributions to this literature include Przeworski and Vreeland (2000), Barro and Lee (2005), Easterly (2005), and Dreher (2006). Instead of reviewing this research in detail, we refer the reader to Przeworski and Vreeland (2000) for an overview of the literature. In summary, while the authors are critical as to the growth enhancing effects of IMF-supported programs, they note that: ‘statistical findings range all over the spectrum of possible conclusions’ (Przeworski and Vreeland, 2000, p. 386).

Third, the paper also relates to studies analyzing the determinants of the demand for IMF resources. Research at the IMF has produced several papers in this strand of the literature (Knight and Santaella, 1997; Cerutti, 2007; Elekdağ, 2008; Ghosh et al., 2008; Poulain and Reynaud, 2017). Based on a quarterly dataset of 59 non-PRGT-eligible developing countries over the period from 1982 to 2005, Cerutti (2007) investigates the factors that lead countries to request IMF support. The results suggest that a small number of country-specific macroeconomic variables such as growth, international reserves and the current account balance significantly influence a country’s probability to request an IMF-supported program. On the other hand, global economic activity is found to be relevant only during the period of debt crises in the 1980s. In contrast, Elekdağ (2008) emphasizes the role of the global economic environment in the demand for IMF resources. Specifically, by looking at the demand for IMF resources in a dataset of 412 GRA-supported programs from 1970 to 2004, oil prices, world interest rates, and global growth are shown to be important determinants. Finally, Bird and Rowlands (2001) evaluate the impact of political and institutional factors on IMF lending. While the authors find some political variables as well as a country’s past program history to be significant factors, they note that their contribution to successfully predicting arrangements is negligible.

To the extent that repeated use of IMF-supported adjustment programs can be seen as a risk indicator for countries’ capacity to repay the Fund, this study also indirectly relates to a strand of the literature that has investigated the determinants of arrears to the IMF. Though this literature faces the problem that arrears to the IMF are very rare events which can distort the results obtained from standard binary response models, its findings are nonetheless quite intuitive. Aylward and Thorne (1998) use logit analysis for a panel of 138 developing countries over the period 1976 to 1993 to shed light on the factors explaining the incidence of arrears to the IMF. They find that a small number of financial and macroeconomic factors such as low economic growth and the amount of IMF credit relative to quota drives the

likelihood of arrears. Moreover, the strong explanatory power of past arrears for current arrears indicates strong time-dependence in the occurrence of arrears to the IMF. Oka (2003) focuses on whether arrears are predictable by applying probit-based models and a so-called signals approach. Specifically, the risk of arrears relates positively to the outstanding debt relative to quota, the occurrence of arrears to external creditors and negatively to reserves relative to imports and the growth rate of the economy. Finally, Oeking and Sumlinski (2016) also analyze what explains the incidence as well as the duration of arrears to the Fund. They find that arrears to the IMF in the previous five years, reserves coverage of imports, and institutional quality are among the most important factors that correlate with the occurrence of arrears. Regarding the duration of protracted arrears, they conclude that IMF credit outstanding, real GDP growth, the share of exports to advanced economies, and episodes of civil unrest are correlated with how long a country remains in arrears.

III Data

Our sample of IMF-supported programs consists of virtually all GRA arrangements that started between January 1952 and January 2012.⁴ Specifically, the dataset contains 890 GRA-supported programs: 700 Stand-By Arrangements (SBA), 81 Extended Fund Facilities (EFF) and 109 First-Credit Tranche Arrangements (FCTA).⁵ The programs in the sample were requested by 138 different IMF member countries and the number of programs per country ranges from 1 to 23 with an average of 6.4 programs per country. We exclude precautionary arrangements that ended without an actual BoP need associated with disbursements by the IMF as these arrangements do not involve subsequent repayment obligations. We also exclude some arrangements such as those requested by Cuba (1), former Czechoslovakia (2), former Serbia & Montenegro (2) and former Yugoslavia (12) due to very limited data availability and/or changes in the country's existence as an IMF member.

⁴The latest possible starting date is chosen such that our definition of 'no successor arrangement' is applicable to all programs in the sample.

⁵The FCTAs included in the sample are First Credit Tranche Stand-by Arrangements. They differ from outright purchases within the first credit tranche in that they were generally not different from other types of arrangements and most even had ex-post conditionality designed as indicative targets. An IMF paper from 1991 on operational issues associated with the use of Fund resources indicates that 'all of the 17 requests for first credit tranche transactions in the 1980s were supported by programs with objectives similar to those of programs in the upper credit tranches — the correction of economic imbalances with a view to achieving medium-term balance of payments viability with price stability and sustainable economic growth. All included undertakings for substantive policy actions, and most were based in part on important policy measures — in particular with respect to fiscal, monetary and exchange rate actions — that had already been implemented by the time the requests were presented to the Executive Board.'

A Variables

The dependent variable in our analysis is binary. It takes the value one if an arrangement is not succeeded by another within a certain period and zero otherwise. Given the risks associated with repeated use, we will thus evaluate what makes a program "successful" in the narrow sense that there has been no successor arrangement.⁶ While our sample only contains GRA arrangements, for the assessment of repeated use we also take into account concessional successor arrangements approved after concessional trust-lending was introduced in the mid-1980s. Ignoring these arrangements when defining the dependent variable would introduce an upward bias in the share of programs without a successor. Our baseline analysis considers three years after expiration or cancellation as a threshold period to categorize programs as "successful" or not as the first repayment obligations to the IMF would typically fall due by this time. Figure A-1 plots the corresponding periods of each country's programs over time and visualizes the occurrence of repeated use.

We consider three subsets of explanatory variables that could drive the probability of (avoiding) repeated use: (i) domestic conditions, (ii) the global economic environment, and (iii) variables reflecting a country's relations with the IMF, as well as program-specific characteristics. Table A-1 contains all variables used in either the baseline specifications or any of the robustness checks along with their description, details on their measurement, data sources, and the share of missing values across programs. Moreover, Figure A-4 visualizes the time dimension of the variable definitions and their measurement. It also shows that the explanatory variables are measured before the dependent variable materializes which limits the possibility of reverse causality in our empirical analysis.

The selection of domestic and global variables tries to cover important dimensions of a country's economy and relies heavily on the existing literature (see e.g. Aylward and Thorne, 1998; Cerutti, 2007; Ivanova et al., 2001; Larch et al., 2017; Poulain and Reynaud, 2017). We include key variables that are typically at the center of IMF-supported programs such as GDP growth, the primary balance, public debt, the current account balance, international reserves, the exchange rate, and inflation. For some of these variables (primary balance, public debt, current account balance and international reserves), we account for both a country's initial condition at the year of program start and the respective change over the program period. Moreover, we consider the role of trade openness as well as the exchange rate regime. Finally, we look at the impact of institutional quality on program 'success'. As our baseline indicator

⁶We emphasize that 'no successor program' is only a proxy for program success. While this definition correlates with other definitions of program success (see e.g. Larch et al., 2017) such as favorable macroeconomic outcomes or implementation of conditionality, it does not capture every single case that evidently has been a success or failure. This caveat should be kept in mind for the discussion in the paper.

of the institutional environment, we use a bureaucracy quality index, which provides some indication of a country's capacity to effectively implement IMF programs.

We consider several proxies for the global economic environment. These include global economic growth, world real interest rates defined as the U.S. federal funds rate adjusted by CPI inflation and market volatility as measured by the VIX index. In a related study, Elekdağ (2008) finds improving global economic activity to have a negative impact on the demand for IMF resources while interest rate increases lead to a higher probability of an arrangement being approved within a certain year. Regarding volatility, Larch et al. (2017) finds that rising market uncertainty decreases the chance of program success. All global determinants enter the analysis with their average value over the relevant period prior to a (potential) successor program. For programs with a successor this is the period between the start of the original program and the start of the successor program whereas for programs without a successor this is the period between program start and the end of the threshold period, i.e., three years after program end.

Finally, we evaluate determinants that relate to a country's relations with the Fund as well as characteristics of a particular program. First, for any given program, we include the fraction of the previous five years that the corresponding country has spent under an IMF-supported program.⁷ Second, in order to measure a country's long-run track record with the IMF, we include successful predecessor programs as a share of total predecessors.⁸ A high value of this variable could potentially capture countries' relatively positive experience in collaborating with the IMF. In contrast, a low value could reflect a dependence on Fund resources that would increase the likelihood of future arrangements. Third, as in Cerutti (2007) we include a country's outstanding loans to the IMF as a share of the country's quota. In order to reflect countries' relative power within the Fund and to account for political economy considerations along the lines discussed in Bird and Rowlands (2001), we also consider a country's quota share. Moreover, to assess the impact of program size on the probability of no successor arrangement, we include a dummy variable that takes the value one for arrangements whose approved sizes exceed normal access limits set under the IMF's lending policies and zero otherwise.⁹ Since the possibility of exceptional access has only been introduced relatively

⁷The definition of this variable is backward-looking as opposed to the dependent variable of (no) repeated use which is forward-looking. Hence, we can exclude the possibility of reverse causality.

⁸To allow the computation of this variable for each country-program in the sample, specifications including this variable are estimated by dropping the first program of each country.

⁹The IMF sets annual and cumulative limits in percent of a country's quota. These limits can be exceeded in exceptional circumstances subject to higher scrutiny, including the requirement that the requesting country meets a predetermined set of criteria related to debt sustainability, prospects for program success and (re-)gaining private capital market access (IMF, 2018).

recently, the number of programs belonging to this category is limited.¹⁰ Finally, to somewhat take into account the duration of a program, we define two dummy variables that measure whether a program had ended earlier or later than originally planned. The ‘short dummy’ will capture canceled arrangements that may or may not have been replaced by new ones. While the ‘long dummy’ is meant to measure the potential impact of, for example, review delays on the probability of repeated use, it also captures program extensions arising from changed circumstances.

B Stylized facts

Summary statistics can provide preliminary ideas on (no) repeated use and its potential drivers, as defined in the previous section, and serve as a prelude to a more in-depth analysis. Figure A-2 distinguishes groups of countries according to their number of GRA-supported programs. Over our sample period, extensive use of GRA resources is a phenomenon that is prevalent in a rather small number of countries. For instance, around 30 out of 138 countries had more than 10 GRA arrangements since 1952. Complementary to this, Figure A-3 illustrates the concentration of arrangements across countries by contrasting the share of countries that have had the most arrangements with the corresponding share in the total number of programs. For example, the top 20 percent of countries in terms of number of programs approved per country account for approximately 50 percent of all GRA programs, pointing towards significant concentration. The unconditional probability of ‘no successor arrangement’, which equals the sample average of the binary dependent variable, is approximately 22 percent.¹¹ Figure 1 plots the number of programs with and without successors per year along with a five-year moving average of the share of programs without successor arrangements. This share features sizable time variation over the sample period and has increased significantly since the year 2000.

The sample averages of the explanatory variables provide intuitive insights when comparing arrangements with successors to those without (Table A-2). Around half of the variables collected at the program level have unconditional means that differ significantly between programs with and without successors.¹² Moreover, the mean differences are broadly in line with what has been documented in the previous literature as well as what one would expect based on economic theory and intuition.

¹⁰Starting from Mexico’s SBA in 1995, there have been 36 exceptional access cases in our sample.

¹¹The unconditional probabilities for the different types of GRA programs are 22 percent (SBA), 25 percent (EFF), and 17 percent (FCTA).

¹²The last column of Table A-2 contains the p-values of a statistical test for significance of the difference in the unconditional means of the two groups of arrangements.

First, programs that did not have a successor are characterized by an average real growth rate that is one percentage point higher than the one of programs with successors. Second, the initial conditions of countries implementing programs without successors are, on average, more favorable, including: a lower public debt-to-GDP ratio, a larger current account surplus (or a smaller deficit) and a larger stock of reserve assets in the year of program start. Third, countries implementing programs that did not require a successor have on average stabilized debt, improved both their current account balance and their primary balance and increased their reserves. Fourth, trade openness is, on average, higher and the average quality of the institutional framework better in countries implementing such programs. By contrast, inflation, the exchange rate adjustment and the existing exchange rate regime do on average not differ significantly between both program groups.

The variables describing the global economic environment are not associated with a clear-cut differentiation between arrangements with and without successors. In particular, average world growth and market uncertainty have not been significantly different for the arrangements in the two groups. In contrast, and to some extent intuitive, global interest rates have been somewhat lower during and after programs that did not have a successor.

The IMF-related and program-specific variables generally reveal some significant and intuitive differences in the unconditional means across programs with and without successors. First, countries whose GRA arrangements did not have successors tend, on average, to have a larger share of the IMF quota, spent a smaller fraction of the previous five years under a program, and have been more successful in implementing IMF-supported programs in the past. Second, regarding program size, exceptional access cases have an unconditional probability of 'no successor' that is six percentage points higher compared to their normal access counterparts. Finally, the sample averages of the duration dummies and the stock of credit a country owes to the IMF do not differ significantly between both groups.

Looking at selected key macroeconomic variables from a more dynamic perspective also points towards intuitive differences between programs with and without successors. Figure 2 shows the evolution of four key macroeconomic variables during the pre-program, program, and post-program periods for the two subgroups of programs. Specifically, the plots present the median values across programs along with the first and third quartiles around the year of program start T . Country programs without a successor show on average some improvement in the primary balance, the current account balance, and the reserves position in the five-year period subsequent to program approval. While average growth declines initially in year T , it picks up from $T + 1$ onwards and returns on average to pre-program levels in the medium run. In contrast, programs that have a successor are associated with virtually no movement

in the selected macroeconomic variables. Moreover, the dispersion across programs, reflected in the interquartile range, is much larger in this group of programs.

Overall, the descriptive statistics suggest distinct differences between programs that have been followed by successor arrangements and those that have not. Determining whether these differences in the unconditional means matter even after controlling for other factors will be analyzed in the following sections.

IV A probit model with missing observations

A Overview of the model

This section introduces a method to analyze the determinants of (no) repeated use in greater detail. As defined earlier, our dependent variable y_i is binary and takes the value one if a program is not succeeded by another arrangement within three years. Our starting point is a standard probit model where the observed $i = 1, \dots, n$ realizations of the binary variable y_i relate to an underlying latent variable y_i^* as follows,

$$\begin{aligned} y_i &= 1 && \text{if } y_i^* > 0, \\ y_i &= 0 && \text{if } y_i^* \leq 0. \end{aligned} \tag{1}$$

The unobserved variable y_i^* is assumed to follow a linear model,

$$y_i^* = x_i\beta + \varepsilon_i, \quad \text{with } \varepsilon_i \sim \mathcal{N}(0, 1), \tag{2}$$

where x_i is a $(1 \times k)$ vector of explanatory variables, β is a $(k \times 1)$ vector of regression coefficients and ε_i is an error term with unit variance which is the conventional assumption required for identification of the model. The conditional (fitted) probability of $y_i = 1$ can then be written as,

$$\Pr(y_i = 1|x_i) = \Phi(x_i\beta), \tag{3}$$

where Φ is the cumulative distribution function of the standard normal distribution. Given data on y and x , the coefficient vector β of this non-linear model can be estimated using the maximum likelihood (ML) approach.

However, in our application, not all of the explanatory variables in x are observed for every program in the sample. While one could continue the analysis only with the complete data programs, this would significantly reduce our sample size and at the same time throw away

potentially important information in the data. Alternatively, in order to allow for missing observations, we apply the approach of Conniffe and O’Neill (2011) and split up the vector x_i as follows,

$$y_i^* = x_{1i}\beta_1 + x_{2i}\beta_2 + \varepsilon_i, \quad (4)$$

where x_{1i} and x_{2i} have dimensions $(1 \times k_1)$ and $(1 \times k_2)$, respectively. Conniffe and O’Neill (2011) assume that data is available on y_i, x_{1i}, x_{2i} for $i = 1, \dots, r$ (complete observation sample) and on y_i, x_{1i} alone for the remaining $(n-r)$ observations. The incomplete sample observations are modeled as,

$$x_{2i} = x_{1i}C + u_i, \quad \text{with} \quad u_i \sim \mathcal{MVN}(0, \Sigma), \quad (5)$$

where C is a $(k_1 \times k_2)$ matrix of parameters. Taking together Equations (4) and (5) yields,

$$y_i = x_i(\beta_1 + C\beta_2) + \epsilon_{y_i}, \quad (6)$$

where (ϵ_{y_i}, x_{2i}) are multivariate normally distributed, conditional on x_{1i} . While these parametric assumptions are convenient to derive the analytical form of the log-likelihood over the entire sample and hence to obtain an efficient estimator by ML, Conniffe and O’Neill (2011) also present evidence pointing towards robustness in case of various potential departures from the assumption in Equation (5). Instead of presenting the detailed derivations of the efficient estimator and its asymptotic variance, we only report the building blocks of the Conniffe and O’Neill (2011) procedure (see also Liu and Moench, 2016):

1. Estimate a linear regression of x_2 on x_1 by OLS for the r complete observations.
2. Estimate a probit model for y on x_1 and x_2 using the r complete observations.
3. Estimate a probit model for y on x_1 using the $n - r$ incomplete observations.
4. Combine the estimation outputs of steps 1 to 3 to obtain the so-called ‘probit-miss’ estimator and its corresponding covariance matrix.

While the approach of Conniffe and O’Neill (2011) is a convenient way to deal with missing observations in a binary response framework, it has a significant limitation. Conniffe and O’Neill (2011) assume that the number of missing values is constant across the incomplete sample observations and that they occur for the same variables. This assumption is violated in our dataset where missing values across variables are somewhat concentrated in earlier years and certain program countries but do generally not occur regularly (see also Table A-1,

last column). Nonetheless, the estimator is still applicable and can yield significant efficiency gains for estimating the coefficients on the variables without missing data, but given our data structure it does not fully explore the information in the variables with missing values. In particular, with our sample of programs, we cannot use for estimation those values that variables with missing data have for any of the $(n - r)$ programs with incomplete data. Again, this is because the approach is based on estimating separate probit models on the complete and incomplete parts of the sample, i.e. it rules out the possibility that the number of missing values can vary across variables. Consequently, only the coefficients of the variables without missing data are expected to be estimated more precisely whereas the coefficients on the ones with missing values will have estimation uncertainty comparable to the complete sample analysis.¹³ However, since most of our variables are available for all programs (see Table A-1) and missing values often occur for different variables in the same programs, using the Conniffe and O’Neill (2011) approach still yields significant efficiency gains.

The Conniffe and O’Neill (2011) approach assumes the missing observations to be ‘missing at random’ (MAR). This means, that the probability of data missing on x_{2i} is unrelated to the value of x_{2i} conditional on the other variables in the model. Following Liu and Moench (2016), this implies that the reason for missing values to occur should not be linked to an omitted variable that is correlated with (no) repeated use. Conveniently, consistency of the efficient estimator can be tested by comparing it with the complete case estimator using a Hausman-type test (see Conniffe and O’Neill, 2011, for the details). The results suggest that the assumption is valid for the majority of our model specifications.

B Goodness of fit and performance evaluation

In order to assess the suitability of the different specifications in terms of both goodness of fit as well as the model’s ability to correctly classify programs, we report two different measures. The first is the pseudo R-squared of Efron (1978) which attempts to mirror the regular R-squared of an OLS regression. However, since the residuals of a probit regression are not comparable to those obtained from a linear regression, the difference between predicted probabilities and actual outcomes is not easily interpretable. Nevertheless, the measure provides an initial idea of how much variation in the outcomes can be explained through the model.

Second, we report a measure known as the area under the receiver operating characteristic curve (AUROC), which was previously used by Liu and Moench (2016) to assess the accuracy

¹³Using Monte Carlo simulations and an empirical application, Conniffe and O’Neill (2011) illustrate that estimation is indeed more efficient when using their estimator compared to the complete case analysis.

of different binary response models in predicting U.S. recessions. If the estimated probit model is used for forecasting purposes (both in-sample and out-of-sample), it produces a predicted probability of 'no successor' conditional on the realization of a set of explanatory variables. However, in order to judge whether a program outcome was correctly anticipated, these predicted probabilities need to be complemented by a classification scheme, i.e. a probability threshold beyond which a program is predicted to not have a successor. In principle, one could use a simple classification threshold (e.g. 0.5 or the unconditional success probability). However, in this case choosing a threshold appears rather arbitrary and at the same time heavily affects the number of correctly predicted outcomes. Alternatively, one could evaluate the performance of a particular model for numerous different thresholds and summarize the respective outcomes using a single measure. The AUROC measure serves exactly this purpose and lies between 0 and 1. A value of 0.5 corresponds to a forecaster guessing outcomes of programs according to an arbitrary probability whereas 1 reflects a model with perfect forecasting accuracy.¹⁴ For details regarding the receiver operating characteristics (ROC) curve and the AUROC measure, the reader is referred to Liu and Moench (2016).

V Estimation results

This section presents the estimation results of different probit specifications. Part A elaborates on the baseline results while Part B discusses augmented specifications including additional explanatory variables. Part C briefly discusses additional robustness checks that have been conducted. In the regression tables, for each variable, we report the estimated coefficient from the probit model as well as its standard error. The estimated coefficient can be used to assess the sign and the significance of the corresponding regressor. However, they are difficult to interpret in a non-linear model, such as probit. The effect of a change in one of the explanatory variables on the dependent variable depends on the actual values of the remaining variables. In order to gauge the impact of a one unit change in a particular variable, we report the average marginal effect (AME), i.e. the average effect on the probability of 'no successor' across programs.¹⁵

A Baseline regressions

The results of the first baseline specification are generally intuitive (Table 1). The first column contains results of a model that only includes country-specific variables. First, a

¹⁴Thus, for forecasting purposes, only models with AUROC > 0.5 bear relevance.

¹⁵For details on binary choice models and the derivations of the (average) marginal effects, see, for example, Greene (2000).

country's average growth during and after the program has the expected positive sign and enters the specification significantly. The average marginal effect of around 0.7 indicates that an increase in the average growth rate by one percentage point corresponds to a 0.7 percentage points higher probability of avoiding repeated use.¹⁶ Second, the current account balance at the time of program start as well as its adjustment during the program have the expected positive sign whereas only the former features statistical significance. Third, both a country's stock of international reserves as well as its adjustment over the program period have a strong impact on the probability of repeated use. While both reserve variables have an economically meaningful positive sign, the initial condition features a higher level of statistical significance. Fourth, there is some evidence that the greater a country's trade openness the larger the probability that the country would implement a program without a successor. Similar results have been obtained by Larch et al. (2017) who argue that open countries are more likely to offset negative domestic demand effects, e.g. due to fiscal restrictions, by external demand for their goods. Finally, neither the debt level nor its adjustment over the program period have statistical significance. Previous studies have included the public debt-to-GDP ratio as a potential explanatory variable either for the demand for IMF resources (Knight and Santaella, 1997; Bird and Rowlands, 2001), the occurrence of arrears to the IMF (Aylward and Thorne, 1998), or program success (Larch et al., 2017). The results from these studies are generally mixed but public debt is predominantly found to play a minor role (Knight and Santaella, 1997; Cerutti, 2007).¹⁷

Global and IMF or program-specific variables do generally not emerge as significant in the baseline specification. The second column contains a model specification that additionally includes the global economic environment as measured by world GDP growth and the real U.S. interest rate. However, neither one appears to have explanatory power for the occurrence of repeated use. Again, this is in line with Larch et al. (2017) who find domestic variables to be more relevant for explaining the probability of a successor program. Finally, the last column also considers IMF-specific and program-specific characteristics. Only one of the four variables included shows up significantly. However, the magnitude of the marginal effect of loans outstanding is very small. While not being significant, the negative sign of the share of the previous five years spent under a program points towards some persistence of program failure in the short-run. A stronger result is obtained by Bird and Rowlands (2001) who

¹⁶While this estimated effect appears very small, it is economically quite meaningful given the relatively low unconditional probability of 'no successor', which is only 22 percent. Moreover, as discussed later in this section, the estimated effect increases significantly when the analysis is based on a smaller sample starting in 1990, which includes a more homogenous group of programs.

¹⁷The coefficients of the baseline regressions remain very similar when excluding the regressors with a significant amount of missing data, i.e. the primary balance and reserves variables. We take this as additional evidence that the reported results are not biased by the way missing observations are dealt with.

find the number of months under a high conditionality program in the past three years to be a significant determinant of a new IMF arrangement. The authors attribute this to links that build up between the Fund and the respective country and the fact that renewing a program may be attractive since the authorities have already paid the ‘political costs’ of IMF involvement.

Excluding public debt and its adjustment from the baseline specifications somewhat improves the goodness of fit measures (Table 2). Based on previous studies’ findings of a minor role for public debt as a determinant of the demand for IMF resources, the incidence of arrears and our own finding, that neither the debt level nor its adjustment are significant determinants of the probability of repeated use, we drop both debt variables from the initial regressions to derive baseline regressions for the subsequent analysis. With a few exceptions, the variables’ coefficients remain broadly unchanged and the fit of the regressions improves slightly. A notable change is the reduced importance of the adjustment in reserves, both in terms of magnitude and statistical significance.

Baseline regression results based on a smaller sample starting in 1990 (Table 3) differ from those based on the full sample and show improved goodness of fit measures. This suggests that the relative importance of certain determinants has changed over time. Other studies in the literature have also considered split-sample exercises. For instance, Bird and Rowlands (2001) also chose to split their sample starting in 1990 and motivate this choice by the end of the cold war on the one hand and an increasing focus on governance indicators by the World Bank on the other hand. We will come back to the importance of governance indicators later in this section.

Differences between results from the short sample regressions relative to the full sample are reflected in several variables’ coefficients. It can be seen that the AME of domestic growth has more than doubled and is therefore emphasizing the decisive role that economic performance plays in preventing repeated use. The stock of international reserves and its adjustment during the program have lost some significance when considering the subsample since 1990 but remain economically meaningful. In contrast, the adjustment of the current account balance over the program period has gained importance and contributes to a larger extent to achieving program success.¹⁸ The global interest rate shows up with the anticipated negative sign as for example in Elekdağ (2008), possibly reflecting greater financial and economic integration during the era of rapidly advancing globalization. However, it should be noted that this result could also be a time effect since the success share has steadily risen since 1990 (see Figure 1) while the U.S. interest rate can be characterized by a decreasing trend over

¹⁸Also in the smaller sample, the public debt variables did not enter significantly.

the same period. Moreover, global economic growth now emerges as contributing positively to the probability of no repeated use. The two program duration dummies have much larger coefficients though remaining insignificant as in the full sample regressions. The AMEs of both variables suggest that a program that lasts longer (shorter) than planned has on average a 10 (12) percentage points higher probability of repeated use. On the one hand, these results suggest that reasons leading to extensions such as delayed reviews can be seen as a potential risk to program success. On the other hand, a shorter than planned program period likely relates to cases where a program has ended and immediately been replaced by a successor (back-to-back) arrangement.

B Expanded regressions

We expand the baseline regressions with additional explanatory variables, drawing from the existing literature. First, we explore the role of exchange rate adjustment and the exchange rate regime following studies such as Cerutti (2007), Elekdağ (2008), and Larch et al. (2017) while also considering an alternative measure of a country’s track record with the IMF. The second set of variables we add controls for the role of governance following Bird et al. (2004) and Bird (2004) and the role of macroeconomic stability following Knight and Santaella (1997) and Cerutti (2007) as well as the possibility of exceptional access cases.

We do not find evidence that a country’s exchange rate affects the probability of avoiding repeated use. The role of exchange rate adjustments and the exchange rate regime that is in place has been emphasized in several studies. Ex ante, the impact of exchange rate movements is not clear-cut. While a depreciation can be beneficial for exporting goods, it might also increase a country’s debt servicing costs if debt is denominated in foreign currency (Larch et al., 2017). However, when adding the adjustment of the nominal effective exchange rate (NEER) over the program period and a measure of the exchange rate regime to our baseline specifications (Tables 4 and 5), we do not find any of these two variables to have a significant impact on the probability of (no) repeated use.¹⁹ Moreover, their inclusion affects the estimates and significance of the baseline variables only marginally. Previous studies have found significant exchange rate effects on both the probability of program success (Larch et al., 2017) and the demand for IMF resources (Elekdağ, 2008). Cerutti (2007) reports mixed results regarding the role of the exchange rate in determining the demand for Fund programs. Larch et al. (2017) also report a positive effect of exchange rate flexibility on program success while Elekdağ (2008) does not find the exchange rate regime to be important in explaining the demand for IMF resources.

¹⁹The exchange rate regime measure is obtained from Ilzetzi et al. (2019) and defined on a scale from 1 to 14 where a higher value implies a more flexible exchange rate.

When analyzing the full sample of programs, we obtain intuitive insights when replacing the original short-run measure of a country’s program history by a long-run alternative (see last column of Table 4). First, while we have previously considered a measure of a country’s history with the IMF that only accounts for the past five years, we now include the share of programs with no successor meeting our definition of repeated use over the entire past. The corresponding coefficient has the expected positive sign and is statistically significant at the 1 percent level. This gives support to our earlier hypothesis that a country’s history of implementing programs has explanatory power for the risk of repeated use.²⁰

Our findings highlight the importance of well-functioning institutions though macroeconomic stability does not emerge as statistically significant, in line with existing studies. We include inflation and an index measuring bureaucracy quality as indicators of macroeconomic stability and governance, respectively.²¹ Since governance indicators have only received attention and been systematically measured during recent decades, we consider the institutions-augmented specifications only for the subsample since 1990. The results are shown in Table 6. We do not find price stability to have any significant impact on the probability of avoiding a successor arrangement, which confirms findings in earlier work (Knight and Santaella, 1997; Cerutti, 2007). In contrast, we find that a higher quality of bureaucracy within a country contributes significantly towards the probability of not requesting a successor Fund program.²² This is in line with previous results in the literature indicating that the lower the quality of institutions the higher the risk of ‘program recidivism’ and prolonged use (see e.g. Bird et al., 2004; Bird, 2004). In related research, Ivanova et al. (2001) find a broad range of governance and institutional factors significant for explaining program implementation measured as the degree of conditionality that has been met. Moreover, the specification including the institutions variable features the highest values for the goodness of fit measures highlighting the importance of institutions for understanding repeated use. A small example can illustrate well how meaningful the corresponding average marginal effect of around 12 is. For a typical emerging market country requesting IMF support, an improvement of its bureaucracy quality score from 2 (the average rating of all emerging markets countries with Fund programs approved during 2000-11) to 4 (the level of countries such as Iceland and Ireland) would increase the

²⁰However, we note that exactly disentangling the short-run and long-run effects is challenging as both measures are significantly correlated (see Table A-3). Thus, we prefer the more general interpretation that countries’ histories with the Fund matter for the risk of repeated use.

²¹This index is published by the International Country Risk Guide (ICRG) and measured on a scale from 0 to 4 where a higher value corresponds to a higher quality of bureaucracy. Section C discusses alternative dimensions of the *institutions* variable.

²²This result is intuitive as bureaucracy directly relates to a country’s capacity to implement the agreed measures under an arrangement. To the extent that more developed countries tend to have better institutions, the result could suggest that these countries are less likely to request successor arrangements.

probability of no repeated use, *ceteris paribus*, by around 24 percentage points.²³

The last specification of Table 6 also evaluates whether program size has a significant effect on the occurrence of successor arrangements. Specifically, a binary variable measuring whether an arrangement was an exceptional access case, is included. The insignificant coefficient suggests that, after controlling for a large number of country-specific, global, and IMF-specific and program-specific variables, the probability of repeated use is not significantly different between exceptional access cases and their normal access counterparts. However, our sample of 36 exceptional access cases is rather small. Greater experience with this program type would be helpful to draw more definite conclusions.

To summarize, the estimations of the various specifications yield coefficients with generally expected and intuitive signs. Moreover, the pseudo R-squared as a measure of goodness of fit varies within ranges comparable to earlier studies. We find that strong economic growth seems to be essential to prevent repeated use. Besides, several other country-specific variables such as the current account, international reserves, trade openness, and especially institutional quality play a significant role to various degrees. Finally, a country's track record of implementing Fund-supported programs partly explains the probability that it would request a successor program. In particular, countries that avoided repeated use of IMF programs in the past, are more likely to do so again.²⁴

C Robustness checks

This section discusses several additional robustness checks. These are: (i) including a country's share of total IMF quota, (ii) including market volatility as measured by the VIX index, (iii) including a program type dummy, (iv) replacing the institutions variable quality of bureaucracy by alternative measures, (v) including the external debt-to-GDP ratio, (vi) including interaction terms, (vii) including a time trend, (viii) excluding FCTA programs, (ix) including a dummy that takes the value one for 'blended' arrangements and zero otherwise, and (x) varying the threshold period used to define (no) repeated use.²⁵ Overall, all of these tests leave the main conclusions of the previous sections broadly unchanged.

The inclusion of a country's share of total IMF quota or market volatility among the explanatory variables does not alter our findings. The inclusion of a country's share of the IMF quota, to gauge the potential impact of a country's voting power within the IMF on the probability

²³As noted in the previous footnote, bureaucracy quality is rated on a scale from 0 to 4. For EMs, the ratings used in the calculation of the average are as of the year of program approval.

²⁴While institutions and this track record variable of implementing Fund programs positively affect the probability of repeated use, the two are not significantly correlated (see Table A-3).

²⁵Blended arrangements are those that combine financing from GRA and PRGT resources.

of no repeated use, leaves our findings broadly unchanged and reveals that this variable does not have a significant impact (see Column 1 of Table 7). The finding of non-significance of this variable is in line with previous studies (Barro and Lee, 2005; Cerutti, 2007). Following Larch et al. (2017), we also control for the role of market volatility by including the VIX index both in the full sample as well as in the reduced sample regressions. We did not find a significant effect of the VIX on the probability of repeated use (see Column 2 of Table 7).²⁶ While Larch et al. (2017) find volatility to have a negative impact on program success when analyzing a smaller sample of GRA programs starting from 1993 onwards, the effect vanishes when the authors replace their original success definition with a one similar to ours.

Including a program type dummy variable in the set of explanatory variables leaves our findings broadly unchanged. Since SBAs target different BOP problems compared to EFFs, the probability of repeated use might differ between these two types of programs. To investigate the potential difference arising from program types, we include a dummy explanatory variable that takes the value one for EFF arrangements and zero otherwise. The coefficient of the EFF dummy does neither show significance nor does it affect the remaining coefficients, both in the full sample regression (see Column 3 of Table 7) and the regression based on the shorter sample (see Column 11 of Table 7).

Our finding on the relevance of institutional quality is generally robust to the use of alternative indicators of governance. We replaced our initial definition of bureaucracy quality with a measure of government stability and corruption, respectively (see Columns 4 and 5 of Table 7).²⁷ While government stability is significant, the corruption index is not. Overall, the results remain broadly unchanged regardless of which exact definition is applied. In addition, when including an alternative measure of a country's debt burden (total external debt-to-GDP), the corresponding coefficients do not play a significant role (see Column 6 of Table 7).

The inclusion of interaction terms does not alter our findings. Specifically, we include an interaction of the initial condition and the adjustment of a particular variable (e.g. international reserves) over the program period (see Column 7 of Table 7). One would expect that the extent to which adjustment affects program success strongly depends on the country's initial situation. For example, if a country has a lot of reserve assets, the marginal effect of increasing its stock of reserves even further should be lower compared to a country that started off with a lower stock of reserves. However, when including interaction terms for the primary balance, the current account balance, reserves, and public debt, we do not find supporting

²⁶For the sake of brevity, Table 7 only includes the full sample results.

²⁷All of these have been obtained from the International Country Risk Guide. The higher the index, the better the institutions.

evidence. Moreover, the results are not sensitive to the inclusion of deterministic time effects. To account for unobserved time effects we include a linear trend, i.e. the year of program start, as an additional explanatory variable (see Column 8 of Table 7). Both the estimates of the baseline variables as well as the goodness of fit remain essentially unchanged.

Additional robustness checks, which also do not significantly affect our main results consist of excluding FCTAs and taking into account the distinctive nature of blended arrangements. The need to check the robustness of our results after the exclusion of FCTAs is motivated by the fact that FCTAs entail relatively limited access to Fund resources, i.e. only up to 25 percent of a country's quota and have had a less stringent degree of conditionality. Moreover, while most FCTAs were requested by countries that at some point also had been under SBAs or EFF arrangements, some involved industrialized countries such as the United States, Japan, Spain, and the United Kingdom. When excluding FCTAs from the analysis, the results change marginally (see Column 9 of Table 7). Our dataset contains a small number (23) of blended arrangements. These arrangements differ from most in that they reflect a transition from the use of concessional finance under the PRGT to the use of 'GRA only' resources when a member graduates from the PRGT. The transition may matter for the likelihood of repeated use. However, after including a dummy for 'blended arrangements', the baseline results do not change significantly (see Column 10 of Table 7). While the corresponding coefficient is sizable and the sign confirms our hypothesis, the number of blended arrangements is too small to obtain statistical significance.

Finally, changing the initial threshold period to define repeated use only affects the relative importance of some variables but not the general findings (see Column 12 of Table 7). In particular, we change the threshold period from three to five years, which results in a reduction of the unconditional success rate in the sample of GRA programs from 22 percent to 18 percent. The results on the explanatory variables remain broadly similar.²⁸ However, the current account adjustment over the program period gains importance whereas a country's reserves position loses significance. This may indicate that in a longer time span, adjustment of the domestic economy, as reflected in the current account balance, plays an important role in reducing the probability of requesting a successor arrangement. Moreover, the coefficient of the world real interest rate now features significance and shows the expected sign.

²⁸The measurement of the explanatory variables (see Figure A-4) has been appropriately adjusted to account for the change in the threshold period. Moreover, we also experimented with a threshold of one year. While most results remain qualitatively unchanged, the 'short dummy' coefficient increases significantly indicating that often canceled arrangements were immediately followed by a new one.

VI Out-of-sample forecasting

This section presents an out-of-sample forecasting exercise aiming at predicting (no) repeated use, i.e. whether a country is likely to avoid another IMF-supported adjustment program within a defined period after a previous one. A model with significant predictive power could serve as a valuable tool to assess risks as it provides insights into the probability of IMF resources being tied up longer through repeated recourse to IMF-supported programs.

A Design and limitations

For the out-of-sample forecasting exercise, we split up the entire sample of programs into two parts. The observations of the first part are used to estimate different probit specifications. The coefficients obtained from these estimations, together with the predictor variables of the second part's programs, are then used to compute the predicted probabilities of 'no successor arrangement' for the programs in the second part. In addition, to mimic an out-of-sample forecasting situation, we drop programs from the estimation sample that overlap with the prediction period. As an example, if one wants to predict the probability of successors of programs starting after the year 2000, only programs ending by 1997 can be used to estimate the model specification. Otherwise, one would use information on program outcomes for estimation that would not yet be available to a forecaster in the year 2000. While we take this point into account, in other respects our exercise is not 'truly' out-of-sample. To predict program success, we use the realized values of the explanatory variables during the prediction period. Obviously, a forecaster would need to rely on predictions and previous research has indicated that in case of the IMF these predictions tend to be over-optimistic (Bird, 2005; Baqir et al., 2005). For example, given the results of the previous section, a too optimistic growth forecast would result in underestimating the risk of repeated use. Thus, realistic forecasts of the predictor variables are crucial to assess risks to program success and to Fund resources. Moreover, to compute the relevant averages of some explanatory variables, the forecaster would need to know *ex ante* whether and when a program is succeeded by a new arrangement.²⁹ As this is infeasible, the averages would have to be computed over the initial program period and the following threshold period. Being aware of all these limitations that hamper the direct applicability, we still consider our stylized out-of-sample exercise useful as to which variables are valuable predictors of repeated use.

²⁹This is because for the explanatory variables that enter the model specifications as average values, periods over which these averages are calculated are not the same for programs with and without successors. For programs without successors, averages are calculated over the program period and the subsequent three years whereas for programs with successors, averages are calculated over the period until the start of the successor arrangement (see Table A-1 and Figure A-4 for details).

In order to assess the predictive accuracy, we again report the area under the receiving operator characteristic curve (AUROC). We also present the p-value corresponding to a statistical test of the null hypothesis $H_0 : \text{AUROC} = 0.5$ versus the alternative $H_1 : \text{AUROC} > 0.5$. Since 0.5 is the AUROC value for a naive forecast, i.e., random guessing with an arbitrary probability, rejecting H_0 indicates that a model performs better than the naive forecast. While the AUROC is a measure of a model’s capability to identify repeated use of programs, we also report the results of the out-of-sample exercise in a more illustrative manner. Specifically, we tabulate the true versus the predicted outcomes and compute the shares of correctly predicted cases of (no) repeated use as well as their weighted average. However, to do so a single cut-off probability needs to be selected. The cut-off probability depends on the underlying loss function that the forecaster applies. In this case the loss function is rather simple and the cut-off probability is chosen such that it minimizes the number of falsely predicted outcomes (see also Cerutti, 2007). Even though we minimize a simple unweighted sum, the choice of the weights is in principal up to the forecaster and reflects preferences for either correctly predicting program successes or failures.

B Results

For the first out-of-sample forecasting exercise, we estimate a parsimonious probit specification using all programs that started since 1980 and ended before 1997. The starting date of the estimation sample was chosen to at least partly account for the time variation in the relative importance of previous determinants that has been documented in Section V. The included variables have been selected based on the results of the previous section. Moreover, experimenting with different specifications suggested that small parsimonious specifications tend to have a superior out-of-sample forecasting performance. The chosen specification represents a simplified version of our baseline model (see Column 2 of Table 2), including only essential country-specific characteristics and the global variables. Moreover, we include the long-run track record variable (‘share of past programs without successors’), which was found to have a significant impact on the probability of repeated use (see Table 4). Next, we use the estimated model to predict program outcomes of all programs from 2000 onwards. The results are shown in Table 8. Out of the total 52 programs to be predicted, the model correctly calls the outcome of 35. Specifically, the model is able to correctly predict the majority of both ‘no successor’ cases (79 percent) and repeated use cases (57 percent) but it is more successful in predicting the former. The AUROC value of 0.6994 also indicates a reasonable forecasting performance which is clearly superior to the naive forecast.

Table 9 presents the results of a forecasting exercise focusing on more recent programs and

accounting for institutional quality. The model is estimated using programs between 1990 and 2002 and applied to predict program outcomes from 2005 onwards. The set of predictors has been changed to account for the importance of institutional factors. The general forecasting performance increases significantly with a share of 77 percent of correctly predicted outcomes. This indicates that for this period the institutional quality of a country can serve as a valuable predictor to assess the potential risk of repeated use. Again, the model performs significantly better in predicting no repeated use cases than programs with a successor.

Overall, the results of this section suggest that selected parsimonious specifications can help to successfully predict repeated use of IMF-supported programs. However, the forecasting performance is certainly far from being perfect and subject to limitations as mentioned before. It needs to be kept in mind though that the set-up of a successor arrangement is a notoriously difficult variable to predict which is influenced by factors outside the empirical model, not least political economy considerations. Therefore, from a policy perspective the outlined approach could serve as one tool within a larger toolbox to assess risks to program success in general and the financial resources of the Fund in particular.

VII Conclusion

Over the years, repeated use of IMF-supported programs has characterized numerous countries' relations with the IMF. To the extent that IMF support is meant to temporarily help member countries address BoP problems, repeated use is often viewed as a signal of program failure. Understanding the determinants of repeated use is thus crucial from a policy perspective and can help to gauge the risk IMF programs pose to the revolving nature of the Fund's financial resources. To that end, this paper relies on a large sample of GRA arrangements approved between 1952 and 2012 to analyze these determinants. To take into account the fact that the type of countries that have requested Fund-supported programs has changed over decades and that the world itself has evolved over the years, a sample restricted to a shorter time period starting in 1990 is also considered.

Using probit models and accounting for missing observations of explanatory variables, we find that a small number of country-specific variables have significant impacts on the likelihood of repeated use. Specifically, countries with higher economic growth, a larger current account surplus (or a smaller deficit), a larger stock of international reserves, more trade openness, and higher institutional quality are more likely to implement a Fund-supported program that does not require a successor. In addition, program-specific variables, and variables of a country's history with the IMF can partly explain repeated use. In particular, we find some

evidence that a country's long-run track record of implementing programs without successors is a relevant factor for the risk of repeated use. For policy applications, such as ex ante risk assessment of programs during the stage of program design, it would be useful to have reliable forecasts of potential risks of repeated use of IMF programs. This paper shows that, while precisely forecasting repeated use remains challenging, parsimonious model specifications have reasonable predictive power, and can thus be a useful gauge of program risks.

This paper identifies several promising avenues for future research that could help to further inform the Fund's risk assessment framework. First, the analysis could be extended to concessional lending programs under the PRGT. However, in this case repeated use would need to be redefined to account for the longer-term nature of PRGT arrangements. Second, future work could evaluate in more detail the role of program-specific characteristics. One example would be to define a measure of realism of program targets and investigate the impact of over-ambitious targets on the probability of repeated use. This could potentially be achieved by comparing program targets to benchmarks based on country-specific or cross-country experiences and could help strengthen risk assessments as well as inform program design. Finally, the stylized forecasting exercise could be improved by working with data vintages to properly mimic a true out-of-sample forecasting situation and by including measures of forecasting uncertainty surrounding the point estimates.

References

- Aylward, L. and Thorne, R. (1998). Countries' Repayment Performance Vis-a-Vis the IMF: An Empirical Analysis. *IMF Staff Papers*, 45(4):595–618.
- Baqir, R., Ramcharan, R., and Sahay, R. (2005). IMF Programs and Growth: Is Optimism Defensible? *IMF Staff Papers*, 52(2):260–286.
- Barro, R. J. and Lee, J.-W. (2005). IMF programs: Who is chosen and what are the effects? *Journal of Monetary Economics*, 52(7):1245–1269.
- Bird, G. (2004). The IMF Forever: An Analysis of the Prolonged Use of Fund Resources. *Journal of Development Studies*, 40(6):30–58.
- Bird, G. (2005). Over-optimism and the IMF. *World Economy*, 28(9):1355–1373.
- Bird, G., Hussain, M., and Joyce, J. P. (2004). Many happy returns? Recidivism and the IMF. *Journal of International Money and Finance*, 23(2):231–251.
- Bird, G. and Rowlands, D. (2001). IMF lending: how is it affected by economic, political and institutional factors? *The Journal of Policy Reform*, 4(3):243–270.
- Cerutti, E. (2007). IMF Drawing Programs: Participation Determinants and Forecasting. *IMF Working Paper. International Monetary Fund*.
- Conniffe, D. and O'Neill, D. (2011). Efficient Probit Estimation With Partially Missing Covariates. In *Missing Data Methods: Cross-sectional Methods and Applications*, pages 209–245. Emerald Group Publishing Limited.
- Darvas, Z. (2012). Real Effective Exchange Rates for 178 Countries: A New Database. *Bruegel Working Paper*.
- Dollar, D. and Svensson, J. (2000). What Explains the Success or Failure of Structural Adjustment Programmes? *The Economic Journal*, 110(466):894–917.
- Dreher, A. (2006). IMF and economic growth: The effects of programs, loans, and compliance with conditionality. *World Development*, 34(5):769–788.
- Easterly, W. (2005). What did structural adjustment adjust?: The association of policies and growth with repeated IMF and World Bank adjustment loans. *Journal of Development Economics*, 76(1):1–22.
- Efron, B. (1978). Regression and ANOVA with Zero-One Data: Measures of Residual Variation. *Journal of the American Statistical Association*, 73(361):113–121.

- Elekdağ, S. (2008). How Does the Global Economic Environment Influence the Demand for IMF Resources? *IMF Staff Papers*, 55(4):624–653.
- Ghosh, A., Goretti, M., Joshi, B., Thomas, A., and Zaldueño, J. (2008). Modeling Aggregate Use of IMF Resources—Analytical Approaches and Medium-Term Projections. *IMF Staff Papers*, 55(1):1–49.
- Greene, W. H. (2000). *Econometric Analysis*. Pearson.
- Ilzetzki, E., Reinhart, C. M., and Rogoff, K. S. (2019). Exchange Arrangements Entering the Twenty-First Century: Which Anchor will Hold? *The Quarterly Journal of Economics*, 134(2):599–646.
- IMF (2018). *IMF Financial Operations 2018*. International Monetary Fund. Finance Department.
- Ivanova, A., Mayer, W., Mourmouras, A., and Anayiotos, G. (2001). What Determines the Success or Failure of Fund-Supported Programs? In *Second Annual IMF Research Conference*, pages 29–30.
- Knight, M. and Santaella, J. A. (1997). Economic determinants of IMF financial arrangements. *Journal of Development Economics*, 54(2):405–436.
- Larch, M., Bernard, K. M., and McQuade, P. (2017). Fortune or fortitude? Determinants of successful adjustment with IMF programs. *OECD Journal: Economic Studies*, 2016(1):37–69.
- Liu, W. and Moench, E. (2016). What predicts US recessions? *International Journal of Forecasting*, 32(4):1138–1150.
- Marchesi, S. and Sabani, L. (2007). Prolonged Use and Conditionality Failure: Investigating IMF Responsibility. In *Advancing Development*, pages 319–332. Springer.
- Oeking, A. and Sumlinski, M. A. (2016). Arrears to the IMF - A Ghost of the Past? *IMF Working Paper. International Monetary Fund*.
- Oka, C. (2003). Anticipating Arrears to the IMF Early Warning Systems. *IMF Working Paper. International Monetary Fund*.
- Poulain, J.-G. and Reynaud, J. (2017). IMF Lending in an Interconnected World. *IMF Working Paper. International Monetary Fund*.
- Przeworski, A. and Vreeland, J. R. (2000). The effect of IMF programs on economic growth. *Journal of Development Economics*, 62(2):385–421.

Annex 1: Figures and Tables

Figure 1: IMF-supported GRA programs per year

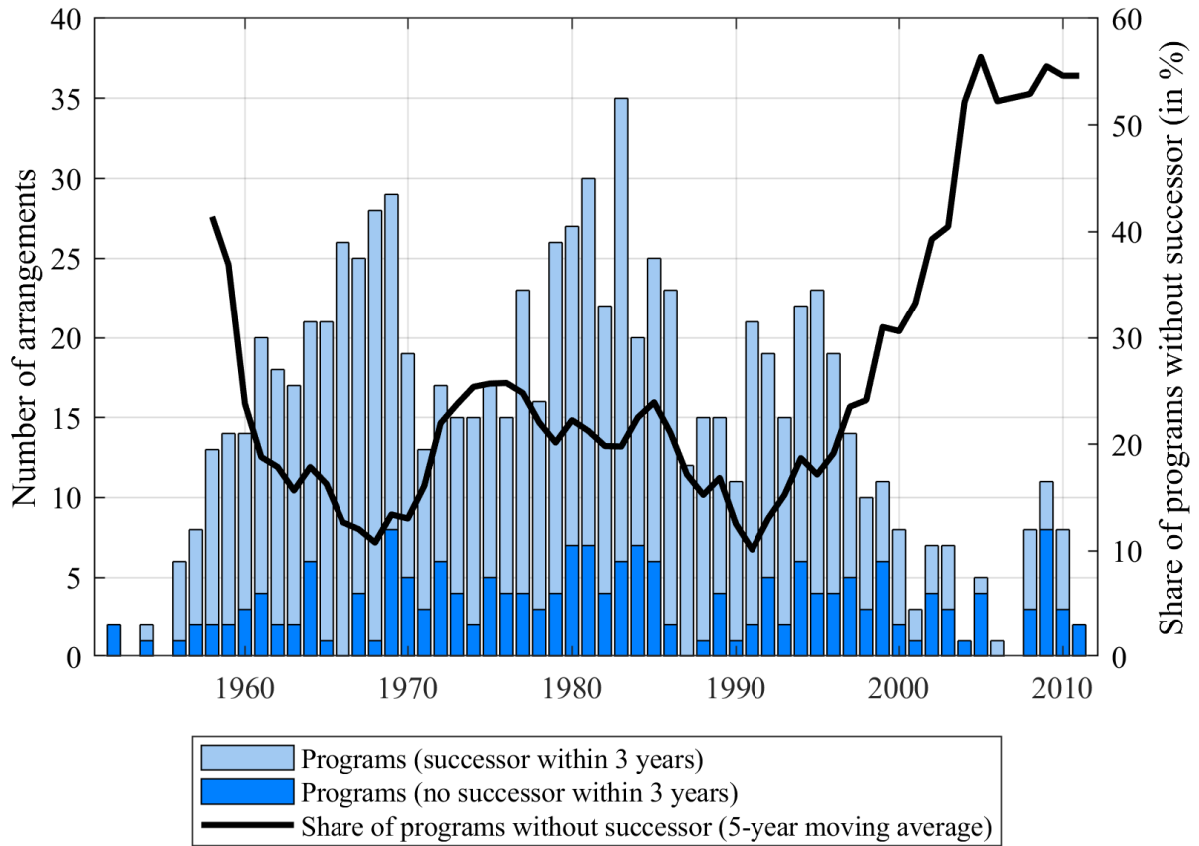


Figure 2: Key macroeconomic variables: pre-program, program, and post-program period for programs without successor (left column) and with successor (right column) within three years

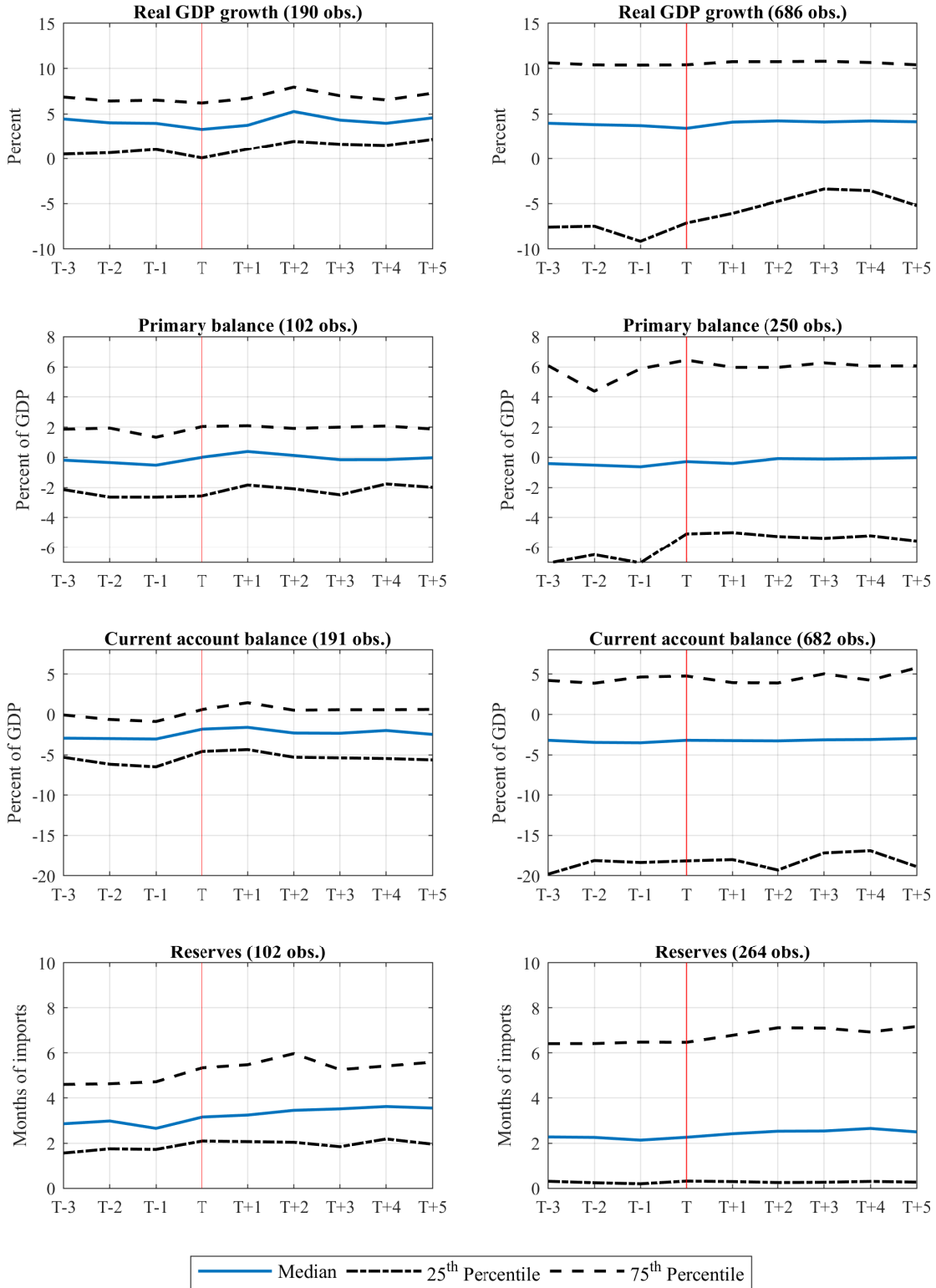


Table 1: Baseline model specifications including public debt, full sample

Variable	AME	β (se $_{\beta}$)	AME	β (se $_{\beta}$)	AME	β (se $_{\beta}$)
GDP_growth	0.70	0.031(0.015) **	0.82	0.036(0.015) **	0.75	0.033(0.015) **
Public_debt	-0.03	-0.001(0.003)	-0.03	-0.001(0.003)	-0.04	-0.002(0.003)
Public_debt_adj	-0.08	-0.004(0.006)	-0.06	-0.003(0.006)	-0.10	-0.004(0.006)
Primary_balance	-0.89	-0.039(0.027)	-1.20	-0.052(0.027)*	-1.31	-0.057(0.028) **
Primary_balance_adj	0.34	0.015(0.034)	0.04	0.002(0.034)	-0.15	-0.006(0.035)
CA_balance	0.75	0.033(0.010)***	0.84	0.036(0.011)***	0.83	0.036(0.011)***
CA_balance_adj	0.40	0.018(0.016)	0.54	0.023(0.016)	0.43	0.019(0.017)
Reserves	2.54	0.112(0.038)***	2.34	0.101(0.038)***	2.22	0.097(0.038) **
Reserves_adj	3.13	0.137(0.068) **	2.39	0.103(0.067)	3.12	0.137(0.069) **
Openness	0.17	0.008(0.002)***	0.18	0.008(0.002)***	0.17	0.007(0.002)***
World_GDP_growth			-1.00	-0.043(0.047)	-0.33	-0.014(0.048)
US_IR			-0.24	-0.010(0.020)	-0.06	-0.003(0.021)
Loans_out					0.02	0.001(0.000) **
Share_cov_past					-0.05	-0.002(0.002)
Dum_long					1.91	0.082(0.185)
Dum_short					-1.51	-0.067(0.178)
Obs. (complete)	890 (263)		890 (263)		890 (263)	
R ²	0.1232		0.1288		0.1410	
AUROC	0.7077		0.7147		0.7141	

Notes: This table reports the estimated coefficients of probit regressions and the corresponding standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. A constant is included in each model but not reported. AME is the average marginal effect, i.e. the average expected change in the predicted probability given a one unit change of the corresponding explanatory variable. For dummy variables, the AME is measured as the change in the average predicted probability for a change from 0 to 1. As a measure of model fit we report the pseudo R² of Efron (1978). AUROC is the area under the receiving operating characteristic curve.

Table 2: Baseline model specifications, full sample

Variable	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})
GDP_growth	0.81	0.037(0.012)***	0.89	0.039(0.013)***	0.89	0.040(0.013)***
Primary_balance	-0.62	-0.028(0.024)	-0.86	-0.038(0.024)	-0.82	-0.037(0.025)
Primary_balance_adj	0.29	0.013(0.032)	0.06	0.002(0.033)	-0.06	-0.003(0.033)
CA_balance	0.73	0.033(0.010)***	0.81	0.036(0.010)***	0.80	0.036(0.010)***
CA_balance_adj	0.50	0.023(0.015)	0.62	0.028(0.016)*	0.55	0.024(0.016)
Reserves	2.78	0.127(0.037)***	2.60	0.116(0.037)***	2.54	0.114(0.037)***
Reserves_adj	2.76	0.126(0.066)*	2.04	0.091(0.065)	2.34	0.105(0.067)
Openness	0.16	0.007(0.002)***	0.17	0.008(0.002)***	0.16	0.007(0.002)***
World_GDP_growth			-0.73	-0.032(0.046)	0.00	0.000(0.047)
US_IR			-0.20	-0.009(0.020)	0.01	0.000(0.020)
Loans_out					0.02	0.001(0.000)*
Share_cov_past					-0.07	-0.003(0.002)
Dum_long					1.75	0.076(0.178)
Dum_short					-1.20	-0.055(0.170)
Obs. (complete)	890 (279)		890 (279)		890 (279)	
R ²	0.1233		0.129		0.1405	
AUROC	0.7204		0.7295		0.7279	

Notes: See Table 1.

Table 3: Baseline model specifications, sample since 1990

Variable	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})
GDP_growth	2.72	0.098(0.027)***	2.15	0.080(0.029)***	1.92	0.072(0.031) **
Primary_balance	-0.89	-0.032(0.031)	-0.47	-0.017(0.032)	-0.54	-0.020(0.034)
Primary_balance_adj	1.15	0.042(0.042)	0.78	0.029(0.044)	0.66	0.024(0.045)
CA_balance	1.27	0.046(0.018) **	1.18	0.044(0.019) **	1.20	0.045(0.020) **
CA_balance_adj	1.43	0.052(0.028)*	1.26	0.047(0.028)*	1.39	0.052(0.029)*
Reserves	3.98	0.143(0.056) **	3.03	0.113(0.058)*	2.81	0.105(0.058)*
Reserves_adj	1.51	0.055(0.100)	1.36	0.051(0.103)	2.21	0.082(0.106)
Openness	0.23	0.008(0.003)***	0.17	0.006(0.003)*	0.17	0.006(0.003)*
World_GDP_growth			7.23	0.269(0.158)*	6.88	0.256(0.164)
US_IR			-3.92	-0.146(0.060) **	-4.65	-0.173(0.071) **
Loans_out					0.00	0.000(0.000)
Share_cov_past					0.05	0.002(0.004)
Dum_long					-10.25	-0.404(0.260)
Dum_short					-11.75	-0.470(0.333)
Obs. (complete)	226 (173)		226 (173)		226 (173)	
R ²	0.1821		0.2392		0.2601	
AUROC	0.7469		0.7978		0.8133	

Notes: See Table 1.

Table 4: Model specifications including exchange rate (regime) and alternative measure of track record, full sample

Variable	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})
GDP_growth	0.72	0.033(0.013) **	0.81	0.036(0.014)***	0.74	0.033(0.018)*
Primary_balance	-0.56	-0.025(0.025)	-0.83	-0.037(0.025)	-1.05	-0.047(0.029)
Primary_balance_adj	0.47	0.021(0.033)	0.20	0.009(0.033)	-0.33	-0.015(0.040)
CA_balance	0.74	0.034(0.010)***	0.82	0.036(0.010)***	0.84	0.037(0.013)***
CA_balance_adj	0.49	0.022(0.015)	0.61	0.027(0.016)*	0.51	0.023(0.018)
Reserves	2.78	0.126(0.037)***	2.58	0.114(0.037)***	2.33	0.104(0.040)***
Reserves_adj	3.14	0.142(0.070) **	2.43	0.108(0.068)	3.76	0.168(0.074) **
Openness	0.16	0.007(0.002)***	0.17	0.007(0.002)***	0.16	0.007(0.002)***
NEER_adj	0.16	0.007(0.005)	0.15	0.007(0.005)	0.17	0.007(0.005)
ER_regime	0.44	0.020(0.014)	0.48	0.021(0.014)	0.50	0.023(0.016)
World_GDP_growth			-0.71	-0.031(0.046)	-0.33	-0.015(0.055)
US_IR			-0.22	-0.010(0.020)	-0.23	-0.010(0.024)
Loans_out					0.01	0.000(0.000)
Share_suc_past					0.20	0.009(0.003)***
Dum_long					0.13	0.006(0.200)
Dum_short					-2.92	-0.136(0.196)
Obs. (complete)	890 (279)		890 (279)		752 (255)	
R ²	0.1310		0.1377		0.1760	
AUROC	0.7302		0.7381		0.7584	

Notes: See Table 1. For the specification in the last column, the first program of each country has been dropped to allow the computation of the share of successfully implemented programs in the past for each country-program in the sample.

Table 5: Model specifications including exchange rate (regime) and alternative measure of track record, sample since 1990

Variable	AME	β (se $_{\beta}$)	AME	β (se $_{\beta}$)	AME	β (se $_{\beta}$)
GDP_growth	2.69	0.097(0.027)***	2.18	0.081(0.029)***	2.60	0.092(0.039) **
Primary_balance	-0.88	-0.032(0.031)	-0.49	-0.018(0.032)	-0.55	-0.019(0.041)
Primary_balance_adj	1.27	0.046(0.043)	0.85	0.031(0.044)	0.34	0.012(0.055)
CA_balance	1.27	0.046(0.018) **	1.16	0.043(0.019) **	1.19	0.042(0.023)*
CA_balance_adj	1.45	0.052(0.028)*	1.22	0.045(0.029)	1.34	0.047(0.033)
Reserves	3.92	0.141(0.056) **	2.93	0.109(0.058)*	2.40	0.085(0.067)
Reserves_adj	1.37	0.049(0.101)	1.12	0.042(0.104)	2.20	0.078(0.111)
Openness	0.24	0.009(0.003)***	0.18	0.007(0.003) **	0.19	0.007(0.004)*
NEER_adj	0.06	0.002(0.007)	0.02	0.001(0.007)	0.02	0.001(0.007)
ER_regime	0.26	0.009(0.025)	0.43	0.016(0.026)	1.01	0.036(0.031)
World_GDP_growth			7.49	0.277(0.160)*	8.22	0.290(0.183)
US_IR			-3.95	-0.146(0.061) **	-3.50	-0.124(0.076)
Loans_out					0.00	0.000(0.001)
Share_suc_past					0.29	0.010(0.006)
Dum_long					-7.62	-0.279(0.282)
Dum_short					-7.13	-0.262(0.363)
Obs. (complete)	226 (173)		226 (173)		187 (155)	
R ²	0.18		0.2447		0.2207	
AUROC	0.75		0.8031		0.7932	

Notes: See Table 1. For the specification in the last column, the first program of each country has been dropped to allow the computation of the share of successfully implemented programs in the past for each country-program in the sample.

Table 6: Model specifications including inflation, institutions and exceptional access dummy, sample since 1990

Variable	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})
GDP_growth	2.66	0.115(0.032)***	2.39	0.111(0.033)***	2.28	0.104(0.036)***
Primary_balance	-1.28	-0.055(0.043)	-0.64	-0.030(0.045)	-0.44	-0.020(0.046)
Primary_balance_adj	0.29	0.013(0.061)	-0.04	-0.002(0.062)	0.40	0.018(0.062)
CA_balance	1.18	0.051(0.023) **	0.98	0.045(0.023)*	0.90	0.041(0.024)*
CA_balance_adj	1.40	0.060(0.030) **	1.23	0.057(0.031)*	1.26	0.057(0.032)*
Reserves	2.82	0.121(0.062)*	1.96	0.091(0.064)	1.99	0.091(0.068)
Reserves_adj	-0.70	-0.030(0.113)	-0.75	-0.035(0.116)	-0.82	-0.037(0.119)
Openness	0.20	0.009(0.003) **	0.15	0.007(0.004) **	0.15	0.007(0.004)*
Inflation	-0.01	-0.000(0.001)	0.01	0.000(0.001)	0.01	0.000(0.001)
Institutions	10.91	0.471(0.176)***	11.53	0.533(0.188)***	11.86	0.541(0.194)***
World_GDP_growth			7.04	0.325(0.167)*	6.99	0.319(0.177)*
US_IR			-3.61	-0.167(0.069) **	-3.80	-0.174(0.083) **
Loans_out					-0.01	-0.000(0.001)
Dum_access					6.04	0.262(0.446)
Share_cov_past					0.00	-0.000(0.004)
Dum_long					-5.01	-0.236(0.285)
Dum_short					-8.02	-0.388(0.348)
Obs. (complete)	226 (139)		226 (139)		226 (139)	
R ²	0.2407		0.3058		0.3154	
AUROC	0.7995		0.8420		0.8456	

Notes: See Table 1.

Table 7: Robustness checks

Variable	(1)		(2)		(3)		(4)		(5)		(6)	
	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})
GDP_growth	0.89	0.040(0.013)***	0.85	0.038(0.013)***	0.89	0.040(0.013)***	2.09	0.095(0.035)***	2.27	0.088(0.033)***	0.76	0.042(0.016)***
Primary_balance	-0.81	-0.036(0.025)	-0.74	-0.033(0.025)	-0.84	-0.037(0.025)	-0.41	-0.018(0.047)	-0.41	-0.016(0.046)	-1.32	-0.072(0.029) **
Primary_balance_adj	-0.05	-0.002(0.033)	0.01	0.000(0.033)	-0.06	-0.003(0.033)	0.72	0.033(0.064)	0.82	0.032(0.061)	-0.62	-0.034(0.037)
CA_balance	0.78	0.035(0.010)***	0.79	0.035(0.010)***	0.82	0.037(0.010)***	0.91	0.041(0.024)*	1.13	0.044(0.023)*	0.74	0.041(0.013)***
CA_balance_adj	0.55	0.025(0.016)	0.53	0.024(0.016)	0.57	0.025(0.016)	1.34	0.061(0.033)*	1.32	0.051(0.030)*	0.61	0.033(0.020)*
Reserves	2.54	0.115(0.037)***	2.50	0.112(0.037)***	2.52	0.113(0.037)***	1.36	0.061(0.066)	1.88	0.073(0.065)	3.54	0.193(0.053)***
Reserves_adj	2.14	0.096(0.067)	2.32	0.104(0.067)	2.28	0.102(0.067)	-0.73	-0.033(0.120)	0.14	0.005(0.116)	2.45	0.134(0.082)
Openness	0.17	0.008(0.002)***	0.16	0.007(0.002)***	0.16	0.007(0.002)***	0.14	0.006(0.004)*	0.17	0.007(0.004)*	0.18	0.010(0.002)***
World_GDP_growth	0.03	0.001(0.047)	0.11	0.005(0.047)	-0.01	-0.000(0.047)	4.37	0.197(0.180)	8.58	0.333(0.177)*	-0.06	-0.003(0.056)
US_IR	0.09	0.004(0.020)	-0.20	-0.009(0.021)	-0.01	-0.000(0.020)	-3.46	-0.156(0.080)*	-4.12	-0.160(0.081)*	-0.04	-0.002(0.024)
Loans_out	0.02	0.001(0.000) **	0.02	0.001(0.000)*	0.02	0.001(0.000)*	-0.01	-0.000(0.001)	-0.01	-0.000(0.001)	0.01	0.001(0.000) **
Share_cov_past	-0.07	-0.003(0.002)	-0.08	-0.003(0.002)*	-0.07	-0.003(0.002)	-0.01	-0.000(0.004)	0.05	0.002(0.004)	-0.04	-0.002(0.002)
Dum_long	2.13	0.093(0.179)	1.40	0.061(0.179)	1.54	0.067(0.180)	-6.30	-0.296(0.279)	-7.44	-0.301(0.273)	2.04	0.107(0.193)
Dum_short	-0.98	-0.045(0.170)	-1.48	-0.068(0.171)	-1.41	-0.064(0.176)	-9.47	-0.459(0.351)	-9.73	-0.402(0.341)	-1.11	-0.062(0.192)
Quota_share	0.66	0.030(0.030)										
Vola			0.40	0.018(0.013)								
Dum_EFF					2.39	0.103(0.189)						
Inflation							0.02	0.001(0.001)	0.01	0.000(0.001)		
Gov_stability							3.93	0.178(0.102)*				
Corruption									2.43	0.094(0.153)		
Dum_access							7.12	0.307(0.442)	11.39	0.411(0.435)		
External_debt											-0.02	-0.001(0.002)
External_debt_adj											-0.01	-0.001(0.004)
Time effects		No		No		No		No		No		No
Excluding FCTA		No		No		No		No		No		No
Five-year threshold		No		No		No		No		No		No
Obs. (complete)		890 (279)		890 (279)		890 (279)		226 (139)		226 (139)		890 (237)
R ²		0.1411		0.1393		0.1371		0.2751		0.2738		0.1677
AUROC		0.7296		0.7258		0.7248		0.8145		0.8179		0.7529

Table 7 (continued): Robustness checks

Variable	(7)		(8)		(9)		(10)		(11)		(12)	
	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})	AME	β (se_{β})
GDP_growth	0.98	0.041(0.025)	1.07	0.047(0.013)***	1.20	0.053(0.015)***	0.91	0.041(0.013)***	2.67	0.126(0.039)***	0.96	0.046(0.015)***
Primary_balance	-1.35	-0.056(0.030)*	-1.03	-0.045(0.025)*	-0.98	-0.043(0.026)	-0.98	-0.044(0.026)*	-0.68	-0.032(0.048)	-1.06	-0.051(0.028)*
Primary_balance_adj	-0.04	-0.002(0.037)	-0.29	-0.013(0.034)	0.22	0.010(0.035)	-0.10	-0.004(0.034)	-0.23	-0.011(0.066)	-0.86	-0.041(0.037)
CA_balance	1.34	0.056(0.018)***	0.88	0.039(0.010)***	0.69	0.030(0.012)***	0.80	0.036(0.011)***	0.82	0.039(0.025)	0.91	0.044(0.012)***
CA_balance_adj	1.39	0.058(0.027)**	0.72	0.031(0.016)**	0.44	0.019(0.017)	0.58	0.026(0.016)	1.43	0.067(0.033)**	0.92	0.044(0.019)**
Reserves	2.32	0.097(0.042)**	2.47	0.108(0.037)***	3.59	0.159(0.041)***	2.48	0.112(0.037)***	2.08	0.098(0.071)	0.96	0.046(0.039)
Reserves_adj	3.40	0.142(0.099)	2.28	0.100(0.067)	2.40	0.106(0.069)	1.96	0.088(0.067)	-1.48	-0.070(0.124)	0.89	0.043(0.069)
Openness	0.18	0.008(0.003)**	0.12	0.005(0.002)***	0.19	0.008(0.002)***	0.17	0.008(0.002)***	0.15	0.007(0.004)*	0.16	0.007(0.002)***
World_GDP_growth	-1.10	-0.046(0.106)	0.69	0.030(0.048)	1.10	0.049(0.052)	-0.05	-0.002(0.047)	11.91	0.562(0.203)***	-1.73	-0.083(0.054)
US_IR	-1.42	-0.059(0.034)*	-0.21	-0.009(0.021)	0.29	0.013(0.022)	-0.06	-0.003(0.021)	-4.83	-0.228(0.088)**	-2.23	-0.107(0.034)***
Loans_out	0.02	0.001(0.000)*	0.01	0.000(0.000)	0.01	0.001(0.000)	0.02	0.001(0.000)*	-0.01	-0.001(0.001)	0.02	0.001(0.000)*
Share_cov_past	-0.03	-0.001(0.003)	-0.08	-0.004(0.002)*	-0.03	-0.001(0.002)	-0.07	-0.003(0.002)	-0.04	-0.002(0.004)	0.00	-0.000(0.002)
Dum_long	-5.05	-0.227(0.234)	-2.13	-0.097(0.186)	2.46	0.105(0.181)	2.97	0.128(0.182)	-5.72	-0.279(0.299)	-3.64	-0.187(0.204)
Dum_short	-8.25	-0.385(0.262)	-1.96	-0.088(0.171)	-0.61	-0.027(0.179)	-0.48	-0.022(0.174)	-7.31	-0.363(0.363)	2.20	0.102(0.180)
Dum_EFF									-2.64	-0.127(0.330)		
Dum_blend							-12.65	-0.850(0.680)	-2.89	-0.140(1.808)		
Inflation									0.01	0.001(0.001)		
Institutions									11.68	0.551(0.201)***		
Public_debt	-0.06	-0.003(0.003)										
Public_debt_adj	-0.38	-0.016(0.012)										
Public_debt (IA)	0.00	0.000(0.000)										
Primary_balance (IA)	0.01	0.001(0.004)										
CA_balance (IA)	0.03	0.001(0.002)										
Reserves (IA)	-0.25	-0.010(0.010)										
Time effects		No		Yes		No		No		No		No
Excluding FCTA		No		No		Yes		No		No		No
Five-year threshold		No		No		No		No		No		Yes
Obs. (complete)		890 (263)		890 (279)		781 (269)		890 (279)		226 (139)		890 (276)
R ²		0.1877		0.1701		0.1706		0.1465		0.3140		0.1372
AUROC		0.7644		0.7437		0.7427		0.7376		0.8485		0.7539

Notes: See Table 1. *IA* refers to interaction term and is defined as the product of the initial condition and the adjustment over the program period of the corresponding variable.

Table 8: Out of sample forecasting (period: 01/2000 - 01/2012)

	True (no repeated use)	True (repeated use)	Total
Pred. (no repeated use)	19	12	31
Pred. (repeated use)	5	16	21
Total	24	28	52
Share (correct)	0.79	0.57	0.67
Obs. (est. / pred.)	280 / 52		
Cut-off probability	0.28		
AUROC (p-value)	0.6994 (0.00)		

Included predictors: GDP_growth, Primary_balance_adj, CA_balance, CA_balance_adj, Reserves, Openness, World_GDP_growth, US_IR, Share_suc_past. Estimation period: 01/1980 - 01/1997 (latest completion date).

Table 9: Out of sample forecasting (period: 01/2005 - 01/2012)

	True (no repeated use)	True (repeated use)	Total
Pred. (no repeated use)	12	5	17
Pred. (repeated use)	1	8	9
Total	13	13	26
Share (correct)	0.92	0.62	0.77
Obs. (est. / pred.)	134 / 26		
Cut-off probability	0.43		
AUROC (p-value)	0.7870 (0.00)		

Included predictors: GDP_growth, Primary_balance_adj, CA_balance, CA_balance_adj, Institutions, US_IR, Share_suc_past. Estimation period: 01/1990 - 01/2002 (latest completion date).

Annex 2: Other Dataset Figures and Tables

Figure A-1: GRA programs starting between 1952 and 2012 per country (A-Gab) with (red) and without (blue) successor within three years.



Figure A-1 (continued): GRA programs per country (Gam-New) with (red) and without (blue) successor within three years.

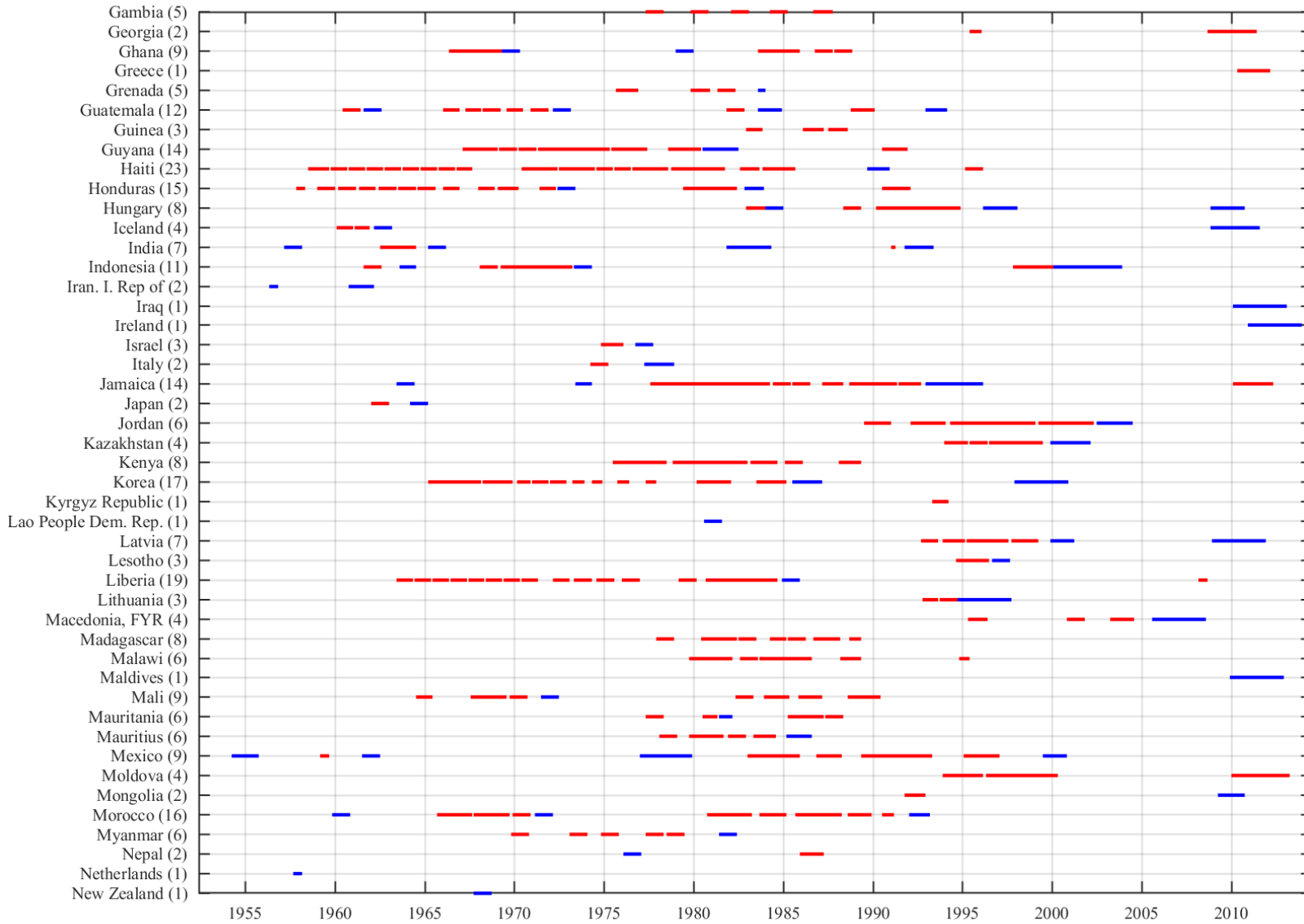
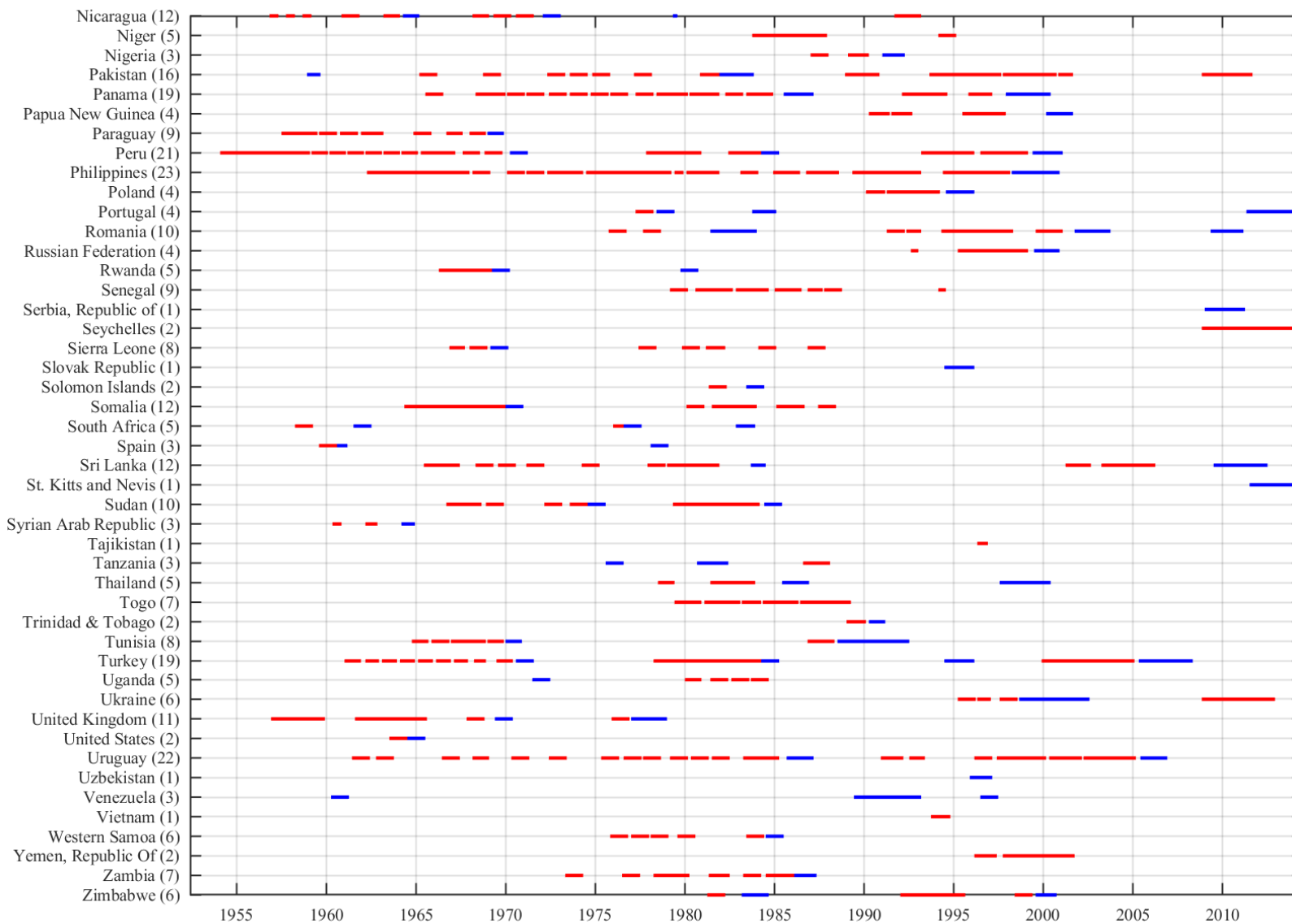


Figure A-1 (concluded): GRA programs per country (Nic-Z) with (red) and without (blue) successor within three years.



Note: Red bars not followed by another bar indicate GRA programs that were succeeded by a PRGT program within three years.

Figure A-2: Distribution of GRA programs across countries I

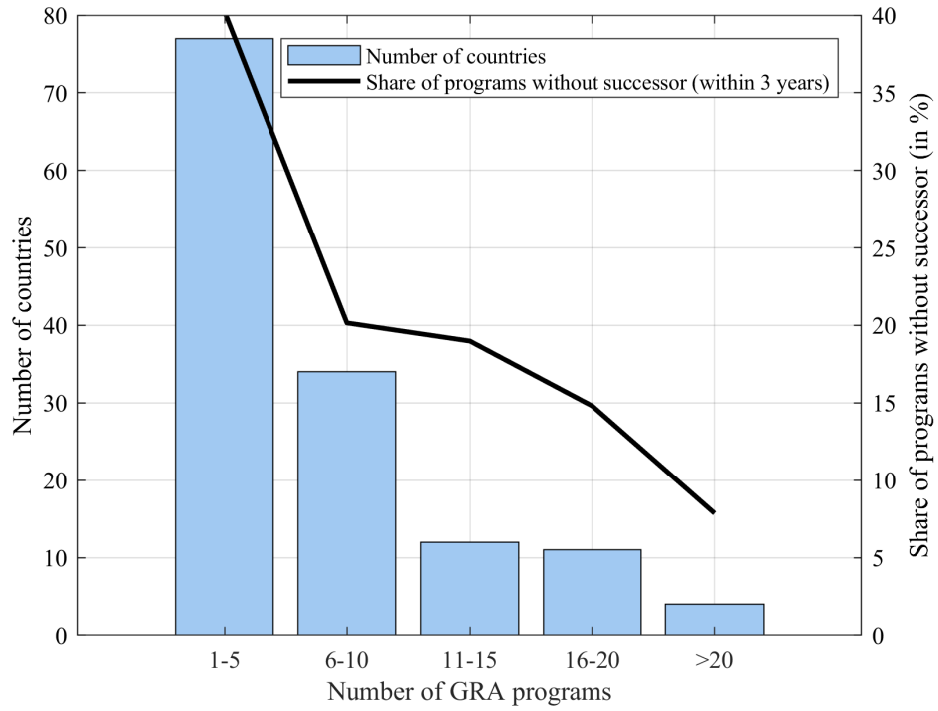


Figure A-3: Distribution of GRA programs across countries II

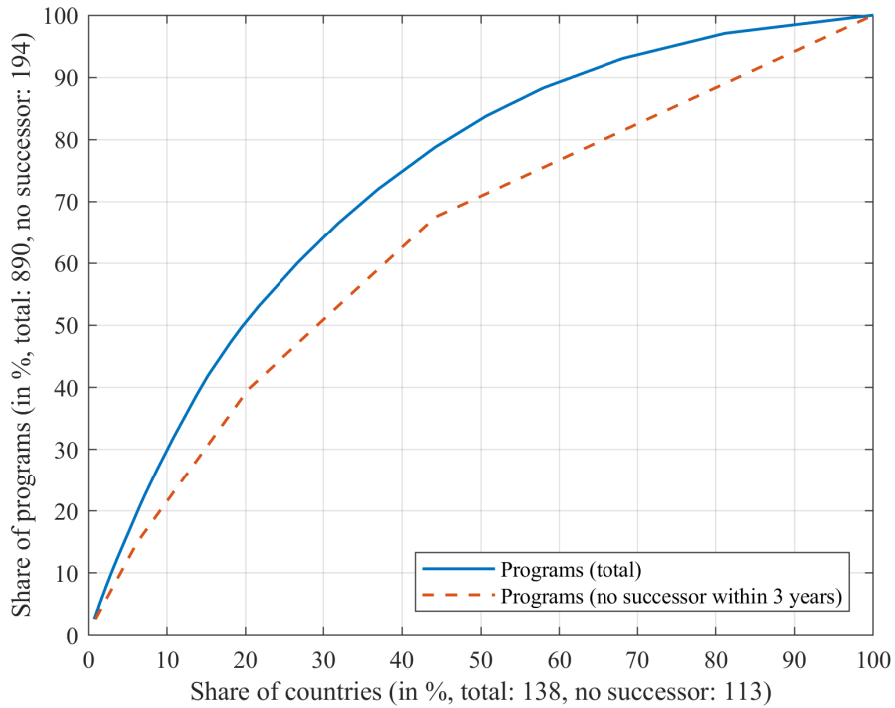
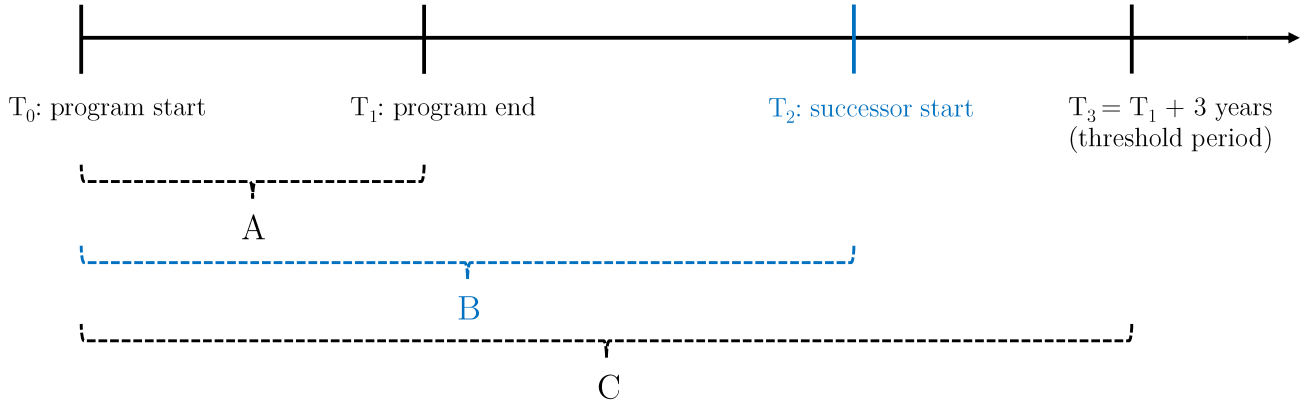


Table A-1: Variable definitions and sources

Variable	Description	Measurement (details in Figure A-4)	Source	Share missing
(No) repeated use	Binary variable (1 if a program was not followed by a successor program within three years, 0 otherwise)	—	Internal IMF program database, authors' calculations	0%
GDP_growth	Real gross domestic product (GDP), percentage change	Average	World Economic Outlook (WEO), Penn World Table (PWT)	0%
Public_debt	General government debt, % of GDP	Initial level at year of program start; change over program period (adjustment)	WEO, IMF Historical Debt Database	22%
External_debt	Total external debt, % of GDP	Initial level at year of program start; change over program period (adjustment)	WEO	42%
Primary_balance	General government primary balance, % of GDP	Initial level at year of program start; change over program period (adjustment)	WEO, OECD, Inter-American Development Bank, IMF Historical Public Finance Dataset, IMF Government Finance Statistics	51%
CA_balance	Current account balance (for 30% of the dataset only trade balance), % of GDP	Initial level at year of program start; change over program period (adjustment)	WEO, PWT, Worldbank World Development Indicators (WDI)	0%
Reserves	Total reserves in months of imports	Initial level at year of program start; change over program period (adjustment)	Worldbank WDI	48%
Inflation	Percentage change of the Consumer Price Index (CPI)	Average	WEO, International Financial Statistics (IFS), Worldbank WDI	6%
Openness	Share of total trade (EX+IM) in GDP	Year of program start	PWT, Worldbank WDI	0%
NEER	Nominal effective exchange rate	Percentage change over program period (adjustment)	Darvas (2012)	11%
ER_regime	Exchange rate regime	Year of program start	Ilzetzki et al. (2019)	0%
Institutions	Bureaucracy quality index (scale from 0 to 4)	Average	International Country Risk Guide (ICRG)	70%
World_GDP_growth	Real World GDP growth, prior to 1963: real U.S. GDP growth	Average	WEO	0%
Vola	VXO index, prior to 1986: 12-month centered rolling window standard deviation of S&P500, level difference removed when merging both series	Average	Cboe Global Markets, FRED Database, authors' calculations	0%
US_IR	Real U.S. interest rate (FED Fundsrate adjusted by CPI inflation)	Average	FRED Database	0%
Quota_share	Share of country's quota in total IMF quota (in %)	Year of program start	IFS, authors' calculations	0%
Loans_out	IMF GRA credit outstanding relative to quota (in %)	Year of program start	IFS, authors' calculations	0%
Dum_access	Binary variable (1 if program is an exceptional access case, 0 otherwise)	—	Internal IMF program database	0%
Share_suc_past	Share of a country's successful preceding GRA programs (in %)	Year of program start	Internal IMF program database, authors' calculations	16%
Share_cov_past	Fraction of previous five years spend under a GRA program (in %)	Year of program start	Internal IMF program database, authors' calculations	0%
Dum_long	Binary variable (1 if program lasted longer than originally planned, 0 otherwise)	—	Internal IMF program database, authors' calculations	0%
Dum_short	Binary variable (1 if program lasted shorter than originally planned, 0 otherwise)	—	Internal IMF program database, authors' calculations	0%
Dum_EFF	Binary variable (1 if program is an EFF arrangement, 0 otherwise)	—	Internal IMF program database, authors' calculations	0%
Dum_blend	Binary variable (1 if program is a 'blended' arrangement, 0 otherwise)	—	Internal IMF program database, authors' calculations	0%

Figure A-4: Variable measurement and timing

- A: period of the program under consideration
- B: period between program start and the start of the successor program (if existing)
- C: program period + 3-year threshold period (relevant if no successor program exists)

Measurement	Program with successor	Program without successor
Initial level (year of program start)	T_0	T_0
Adjustment over program period	A	A
Average	B	C

Notes: Following Larch et al. (2017), in cases where the program ended between the first and eighth month of a year, the previous year has been used as the year of program end (T_1). If the adjustment program only lasted one year, this rule was not applied, i.e. the actual year of program end has been used to compute the relevant adjustments and averages of the explanatory variables.

Table A-2: Summary statistics

Variable	mean (total)	mean (no repeated use)	mean (repeated use)	p-value
GDP_growth	3.14	3.96	2.91	0.01
Public_debt	57.26	52.23	58.78	0.16
Public_debt_adj	1.61	-0.26	2.16	0.13
Primary_balance	-0.21	-0.27	-0.18	0.88
Primary_balance_adj	0.05	0.77	-0.22	0.00
CA_balance	-4.24	-2.40	-4.76	0.00
CA_balance_adj	0.10	0.57	-0.03	0.14
Reserves	3.14	3.83	2.90	0.01
Reserves_adj	0.03	0.35	-0.09	0.14
Openness	52.61	58.23	51.04	0.01
NEER_adj	-7.05	-6.46	-7.23	0.62
ER_regime	7.16	7.32	7.11	0.60
Inflation	40.10	35.30	41.52	0.64
Institutions	1.85	2.01	1.79	0.06
External_debt	56.17	58.93	55.27	0.59
External_debt_adj	0.83	-0.50	1.26	0.54
World_GDP_growth	3.94	3.90	3.95	0.70
Vola	20.87	21.29	20.75	0.16
US_IR	5.92	5.50	6.03	0.04
Quota_share	0.61	0.84	0.54	0.07
Loans_out	101.04	116.70	96.68	0.06
Share_suc_past	11.25	17.10	9.70	0.00
Share_cov_past	40.11	34.29	41.73	0.00
Dum_access	0.04	0.09	0.03	0.00
Dum_long	0.10	0.13	0.09	0.09
Dum_short	0.13	0.11	0.14	0.26
Dum_EFF	0.09	0.10	0.09	0.51
Dum_blend	0.03	0.01	0.03	0.04

Table A-3: Correlation matrix of the explanatory variables

	GDP_growth	Public_debt	Public_debt_adj	Primary_balance	Primary_balance_adj	CA_balance	CA_balance_adj	Reserves	Reserves_adj	Openness	NEER_adj	ER_regime	Inflation	Institutions	World_GDP_growth	Vola	US_IR	Quota_share	Loans_out	Share_suc_past	Share_cov_past	
GDP_growth	1.00																					
Public_debt	-0.06	1.00																				
Public_debt_adj	-0.31*	-0.03	1.00																			
Primary_balance	0.03	0.13	-0.19*	1.00																		
Primary_balance_adj	0.09	-0.02	-0.05	-0.54*	1.00																	
CA_balance	0.09	0.02	-0.13	0.44*	-0.17*	1.00																
CA_balance_adj	0.00	-0.15	-0.10	-0.22*	0.38*	-0.42*	1.00															
Reserves	0.06	-0.14	0.00	-0.13	0.07	0.09	-0.03	1.00														
Reserves_adj	-0.05	0.11	-0.03	0.03	0.07	-0.03	0.39*	-0.20*	1.00													
Openness	-0.09	0.08	-0.12	0.12	-0.09	-0.09	0.04	-0.20*	0.08	1.00												
NEER_adj	0.28*	0.02	-0.12	-0.01	-0.08	-0.04	-0.21*	-0.01	-0.19*	0.27*	1.00											
ER_regime	-0.13	0.07	-0.13	-0.08	0.09	-0.03	0.10	0.11	0.15	-0.15	-0.46*	1.00										
Inflation	-0.17*	0.06	0.00	0.06	-0.07	0.10	-0.06	0.10	0.05	-0.20*	-0.56*	0.25*	1.00									
Institutions	0.04	0.08	0.07	0.06	0.01	0.15	0.05	-0.03	0.16*	0.02	0.01	-0.01	0.06	1.00								
World_GDP_growth	0.32*	0.00	-0.14	0.03	0.10	0.21*	0.05	0.04	0.06	-0.05	-0.01	0.02	-0.01	0.08	1.00							
Vola	-0.18*	-0.08	0.07	-0.16*	0.08	0.05	0.07	0.09	0.12	-0.02	0.01	-0.02	-0.05	0.05	0.15	1.00						
US_IR	0.05	0.09	-0.05	0.22*	-0.12	0.12	-0.06	-0.14	0.02	-0.33*	-0.21*	0.17*	0.20*	0.02	0.14	0.00	1.00					
Quota_share	-0.02	-0.04	-0.02	0.01	0.09	0.20*	0.09	0.10	0.15	-0.39*	-0.30*	0.21*	0.31*	0.26*	-0.05	0.04	0.06	1.00				
Loans_out	0.04	0.01	-0.02	-0.09	0.13	-0.06	0.12	0.10	0.07	0.01	0.05	0.07	-0.05	0.13	0.15	0.16	-0.27*	0.14	1.00			
Share_suc_past	-0.07	-0.12	0.05	-0.13	0.03	-0.10	0.03	-0.17*	-0.01	0.03	-0.07	-0.03	0.01	-0.01	-0.07	-0.07	-0.12	0.21*	0.17*	1.00		
Share_cov_past	0.23*	-0.02	-0.10	0.31*	-0.10	0.23*	0.00	0.14	-0.02	-0.14	0.08	-0.03	-0.04	0.17*	0.20*	0.10	0.25*	0.02	0.04	-0.33*	1.00	

Notes: Statistical significance of the pairwise correlations at the 5% level is indicated by *.