IMF programmes and stigma in Emerging Markets Economies

Claudia Maurini*

January 17, 2020

Abstract

This paper investigates the existence and estimates the magnitude of a financial market stigma associated with the International Monetary Fund's non-concessional programmes. In particular, it focuses on the impact of IMF non-concessional loans on Emerging Markets' sovereign spreads, using the propensity score matching methodology to deal with the selection bias problem. We find evidence of higher spreads for countries supported by a non-concessional IMF programme with respect to comparable countries that are not supported by such a programme. This effect may be linked to both a pure financial stigma and the (low) probability of the programme succeeding, as it tends to dissipate towards the end of the programme and to be smaller and less significant if we restrict the sample to non-repeated programmes (more likely to be successful). Finally, we find that precautionary programmes (such as the Flexible Credit Line) have a negative impact on sovereign spreads.

1 Introduction[1]

IMF stigma has been gaining increasing attention over the past few years as part of the debate on global financial safety nets (GFSNs), because the perceived existence of this stigma is viewed as a factor that precludes an efficient and timely use of the resources allocated to the GFSNs, and more specifically to the Fund. In general terms, stigma refers to the negative perception attached to a country's decision to ask for IMF assistance. This negative perception may affect both the political process, by turning the electorate against the incumbent government (political stigma), and the country's access to financial markets, by sending a negative signal about the country's solvency (economic or financial stigma). In the case of political stigma, governments fear that by accepting the conditionality attached to IMF programmes they may be perceived as having given up sovereignty and that the electorate will punish them for having chosen to resort to the IMF; political

*Bank of Italy, International Relations Department.

[1]This work has benefited from feedback by Giuseppe Parigi, Pietro Catte, Francesco Paternó, Marco Committeri, Flavia Corneli, and Federico Cingano. All remaining errors are my own. The views expressed in the paper are those of the author and do not necessarily reflect those of the Bank of Italy.

stigma became a prominent problem following the Asian crisis of the 1990s [Ito (2002)]. Financial stigma, on the other hand, originates from the fact that 'so much of IMF lending in recent decades is longer term and connected to long run solvency problems' [Reinhart and Trebesch (2016)]. It should also be noted that political stigma and financial stigma may sometimes be interlinked: specifically, the fear that implementing an IMF programme will lead to a defeat at the following general election may decrease market confidence in the success of any IMF programme which is expected to end after the election; this, in turn, will jeopardize access to financial markets. The concept of stigma is important because it has been identified as one of the reasons why countries are reluctant to resort to the IMF, ultimately triggering, on the one hand, excessive self-insurance through reserve accumulation and, on the other hand, excessive delays in engaging with the Fund, which in turn makes programmes more complicated to design and less successful in restoring economic and financial stability [IMF (2017)]. Stigma is also often identified as one of the causes of the recent growth in the number and size of Regional Financing Arrangements [IRC, Task force on IMF issues (2018)].

In this paper we focus on the notion of financial stigma, by using the EMBIG sovereign spread as a proxy measure for it, which is a standard measure of the creditworthiness that financial markets assign to emerging market economies. In this sense, the financial stigma is expected to be revealed by financial markets demanding higher sovereign spreads for countries supported by an IMF non-concessional programme than for countries in a comparable economic and financial condition not supported by the IMF. We check the impact on sovereign spread of both traditional lending programmes (such as the Stand-By Arrangements and Extended Fund Facilities) and precautionary programmes (such as the Flexible Credit Line) and find that the two types of intervention have a different impact on sovereign spreads.[2]

Every assessment of the effect of IMF programmes is faced with the problem of selection bias, arising from the fact that participation in an IMF programme is not random; in fact, the economic conditions and surrounding circumstances of a programme country differ systematically from those of a non-programme country, and any meaningful comparison of the post-programme performance must take such differences into account. In this paper we employ the propensity score matching (PSM) methodology to address the selection bias problem. This allows us to consider the IMF programme as a treatment variable and to build a matched control sample based on some observable covariates, against which is possible to test the difference in outcome with the treated sample. The idea underlying this approach is that controlling for the observable covariates is sufficient to remove the dependence between the treatment assignment and the treatment-specific outcome (selection on observables), also given that the unobservable covariates are likely correlated to those that are observed. As is discussed further in the dedicated section, the use of PSM in assessing the effects of IMF programmes is relatively new but has proven to be a useful way to go beyond the traditional instrumental variables setting that has not led to a convincingly

[2]In the robustness section, we also consider the role of precautionary Stand-By Arrangements.

unambiguous narrative up to now. Our sample consists of monthly observations for 29 emerging market economies over the period January 1997 - August 2017, the choice of the countries in the sample being constrained by the availability of EMBIG spreads data.

With respect to the existing literature, the main contribution of this paper is to use a the PSM approach to assess the impact of IMF programmes on spreads with monthly data, distinguishing the role of non-precautionary and precautionary programmes. Our results point to the following main conclusions: (a) emerging market economies supported by an IMF non-concessional traditional programme (i.e SBA or EFF) record on average a sovereign spread that is about 180 basis points higher compared to similar emerging economies not supported by the IMF and this effect tends to dissipate towards the end of the programme; (b) the impact on spreads may be due both to pure financial stigma (stemming from the mere existence of an IMF programme) as well as to the low probability of success of IMF programmes, revealed by the fact that a large number of them is followed (or replaced) by another IMF programme; we provide empirical evidence consistent with this interpretation by showing that the impact on spreads is smaller and less robust when we restrict our sample to stand-alone programmes, i.e. the IMF non-concessional programmes which are not followed by any other programme; (c) emerging market economies resorting to IMF precautionary facilities record lower EMBIG spreads, a result consistent with the fact that the assignment of these facilities, as it is reserved to strong and sound economies, can hardly send a negative signal about the country's solvency.

The paper is organized as follows: in the next section we review the relevant literature; in section 3 we present our data; in section 4 we outline the main features of the propensity score matching methodology and in section 5 we explain our empirical strategy; section 6 discusses the main results and their robustness; section 7 concludes.

2 Related literature

This paper can be viewed as the first one in the literature to use propensity score matching to assess the impact of IMF programmes on sovereign spreads. With respect to most previous works, it also introduces the use of monthly data and the inclusion in the sample[3] of the most recent wave of programmes approved after the global financial crisis, including precautionary programmes. This work is linked to the large body of literature on the effects of IMF programmes, and more specifically to the relatively smaller subsets of works studying the IMF's catalytic effect and the IMF stigma.

The economic literature investigating the impact of IMF programmes on the supported economies dates back to the 1970s, even though it is only since 2000 that this issue has been tackled using more refined econometric techniques.[4] The main problem in analysing the effectiveness of IMF programmes originates from the fact that countries self-select into the programmes when they face economic difficulties; therefore it is not easy to distinguish

[3]Exceptions in this respect are Andone and Scheubel (2017) and Chapman *et al.* (2015).

[4]For a comprehensive review of the early literature, see Bird (2001)

whether their post-programme economic performance is due to the previous conditions or to the programme itself. The approach most commonly used in the literature to address this issue makes use of instrumental variables, as in the widely cited work by Barro and Lee (2005).[5] The instrumental variables framework assumes that there are omitted factors that affect both the outcome variable (for example, GDP growth) and the country's decision to seek an IMF programme, and proposes as a solution the search for appropriate instruments that are correlated with the IMF variable but uncorrelated with the outcome variable. Barro and Lee (2005) came up with the idea of using political economy variables, such as political proximity to the US or the country's weight within the IMF, as instruments to explain participation in IMF programmes. This strand of the literature, essentially identified by the adoption of the instrumental variable framework to investigate the impact of IMF programmes on different dimensions of the supported economy, is quite wide and generally fails to reach broadly consistent results.[6] For example, Barro and Lee (2005) find no significant effects on investment, inflation, government consumption or international openness, but conclude that higher participation in IMF loans reduces economic growth. Dreher (2006) draws the same conclusion on economic growth, but discovers that greater compliance with the conditionality attached to the programmes mitigates this negative effect. Dreher and Walter (2010) look at the impact of IMF programmes on the likelihood of currency crises and find that IMF involvement reduces the probability of a crisis. Papi *et al.* (2015) show that countries with IMF non-concessional lending programmes are significantly less likely to experience a banking crisis than other countries.

In general, the instrumental variables approach based on a country's political proximity to the US seems to have become increasingly outdated in a more multipolar world, where the euro area and the large emerging market countries (such as BRICS) have begun to have a larger informal influence on IMF decisions (even though this is not reflected in their formal voting power in the case of BRICS), as proven by the massive IMF support given to large euro area countries during the sovereign debt crisis of 2010-2011. This is one of the reasons why more recent works address the selection bias by applying alternative techniques borrowed from impact evaluation studies, such as the propensity score matching methodology.[7] As is explained in more detail in section 4, propensity score matching provides for the construction of a counter-factual sample against which it is possible to compare the performance of the units that receive a treatment, in this case countries that receive an IMF loan. Atoyan and Conway (2006) use PSM to evaluate the impact of IMF programmes on GDP growth, comparing the results obtained using the propensity score methodology and those obtained using the instrumental variables methodology. They find weaker evidence consistent with an improvement in economic growth in countries supported by the IMF during

[5] An alternative technique for addressing selection bias is the Heckman two-stage methodology, used in a smaller number of studies, such as Przeworski and Vreeland (2000) and Bas and Stone (2014).

[6] As also pointed out by Atoyan and Conway (2006).

[7] In the last few years the use of the synthetic control method has also been gaining popularity as a means of studying the impact of IMF policies on member countries. This technique proves useful in a time series setting, with only a few treated individuals and a long time span. See for example Al Sadiq (2015) and Essers and Ide (2017).

the programme, but stronger evidence of an improvement after the programme; they also argue that the heterogeneity in the results obtained with the two techniques depends on the different samples included in the analysis, since matching excludes country episodes associated with extreme values for the propensity score. Other studies focus on smaller (and more homogeneous) groups of countries; for example, Bal Gunduz (2016) and Bird and Rowlands (2017) both focus on Low Income Countries (LICs) using PSM and find that IMF engagement is positively associated with a wide range of macroeconomic outcomes, such as economic growth, current account balances, reserve coverage and inflation.

The idea that an IMF loan may induce markets to apply a higher risk premium to borrowing countries - the main idea tested in the paper - is the reverse of the so-called catalytic effect of IMF financing, which refers to the Fund's ability to facilitate market access (both in terms of prices and quantities) for countries receiving financial assistance; therefore the literature on the catalytic role of the Fund is also relevant to our study.[8] In this context, to the best of our knowledge, the only works that look at the impact of IMF programmes on sovereign spreads are Mody and Saravia (2003), Eichengreen and Mody (2006) and Chapman *et al.* (2015) and none of them use a PSM approach to address the selection bias, as we do. By estimating a two-equation model (a probit for the decision to issue a bond and a spread equation for the determinants of the spreads charged on individual bond issuances), Mody and Saravia (2003) find that the Fund's catalytic role depends crucially on countries' fundamentals, on the features of the IMF programme and on the credibility of the reform effort carried out by the country. Eichengreen and Mody (2006) look at both bond markets and bank lending, following the same approach as in Mody and Saravia (2003), and find that spreads on bonds are lower when they are issued in conjunction with IMF-supported programmes (the exact opposite of a stigma effect). However, the above mentioned studies do not take into account the endogeneity of IMF programmes, as also pointed out by Gehring and Lang (2018).[9]

Two studies that represent an important benchmark for us are Chapman *et al.* (2015) and Gehring and Lang (2018), which pursue a similar research question with a different methodology, namely instrumental variables. The findings of Chapman *et al.* (2015) resonate with ours, since they find that the implementation of an IMF programme is associated with higher bond spreads, and they also find that the size of the IMF loan, the extent of the conditionality and the political economy features give rise to important heterogeneities. Gehring and Lang (2018) look at the impact of IMF programmes on sovereign credit ratings and reach the conclusion that a positive signaling effect prevents creditworthiness from deteriorating despite economic contractions under IMF programmes; they adopt an instrumental variable approach with monthly data, using as an instrument the interaction of a country's probability of having participated in an IMF programme in the past (measured as the number of years the country was in a programme) with the IMF's liquidity ratio

[8]See Cottarelli and Giannini (2002) for a discussion of the channels through which the catalytic effect may take place and for a review of the previous empirical findings.

[9]Mody and Saravia (2003), state that due to the difficulty of modeling selection into IMF programmes and finding a suitable instrument 'explicit consideration of the selection bias is not undertaken' (p.852).

(defined as its liquid resources divided by its liquid liabilities). Their instrument is different from the traditional ones and gives a very superficial characterization of the participation process, since it only considers IMF liquidity (a global variable with a very slow dynamic) and any previous participation in IMF programmes (which may not be totally exogenous with respect to creditworthiness, since a country's repeated engagement with the Fund may signal a bad track record in programme implementation). In addition, their sample is very different from ours since it encompasses a much larger number of countries (100 vs 29 countries) with a lower number of periods (annual data from 1987 to 2013 vs monthly data between 1997 and 2017); in our opinion, monthly data are better suited to grasping the short-term signalling effect triggered by an IMF intervention, while annual data are more useful in assessing the adjustment effects of an IMF programme on the macroeconomic factors determining creditworthiness[10]. Lastly, a big difference which may help in explaining the divergent results is the fact that we only take into account non-concessional programmes, while Gehring and Lang (2018) also consider Poverty Reduction and Growth Trust (PRGT) programmes, which are meant for low income countries and whose main goals are to promote growth and reduce poverty.

Finally, more recently Andone and Scheubel (2017) and Scheubel *et al.* (2018) explicitly studied the issue of stigma, exploiting the recidivism that the data show in the use of IMF resources.[11] Scheubel *et al.* (2018) check whether a potentially adverse market reaction may deter a country in need of financial assistance from asking for an IMF loan; to do so they estimate the market reaction to IMF events in an event study set-up and use the bad market reaction as a treatment variable in a subsequent propensity score matching model to evaluate the impact on the likelihood of asking for an IMF programme. The first stage analysis is performed on treasury billa rather than on the EMBIG spread, to allow for a larger country sample, and finds that an IMF programme can have both positive and negative short-term effects on the t-bill rate (the programme marked a raise in the rate in 62 cases and a decrease in 39 cases); the second stage finds that a previous IMF programme raises the probability of another programme, irrespective of the signs of the previous market reaction. Andone and Scheubel (2017) focus on the effect of a country's past experience with IMF conditionality on the likelihood of entering into a new programme and find that a country that experienced an above-average number of disbursement-relevant conditions in the past is less likely to approach the IMF for help again.

3 Data

One of the main novelties of this paper is the choice of using monthly data. This enriches the sample and allows higher frequency information to be exploited, but at the same time reduces the number of covariates that can be included in the analysis, due to data con-

[10] See Scheubel *et al.* (2018) for a detailed explanation of the advantages of using monthly data over annual data for this kind of analysis.

[11] See Scheubel *et al.* (2018) on a discussion for the possible sources of this phenomenon.

straints. The country sample is made up of countries for which there are sufficiently long time-series of data on the EMBIG spreads, namely 29 emerging market countries with monthly observations from January 1997 to August 2017 (248 periods).[12] For each country, the outcome variable is the monthly-averaged Emerging Market Bond Index Global (EMBIG) sovereign spreads provided by J. P. Morgan. The EMBIG tracks total returns for traded external debt instruments (foreign currency denominated fixed income instruments) issued by sovereign or quasi-sovereign entities in emerging market economies; specifically, we consider stripped spreads which show the yield differential in basis points over US Treasuries, stripping out any credit enhancements such as principal and/or interest collateral. The treatment variable that we consider is a dummy variable indicating whether a country is supported by an IMF programme in a given month. We initially focus on Stand-By Arrangements (SBA), the standard non-concessional IMF programmes[13] In order to check the effect of the programme at different stages of its development, we use different definitions of the programme dummy variables. The baseline SBA variable considers the entire duration of the programme, $SBAinitial$ and $SBAfinal$ consider just the first and last six months of a programme (or a series of consecutive programmes), respectively.[14] Overall in our sample there are 52 SBAs regarding 21 countries (Table 1), which reveals that almost all countries that resort to IMF financial assistance have had more than one programme in the time span of the analysis (8 countries of the sample have had no SBA). Since some countries have a series of consecutive SBAs, which may signal problems in the implementation or in the design of the programmes, we repeat our matching analysis by restricting the sample to those countries with just one programme or two stand-alone (i.e. non-consecutive) programmes[15]; in the following sections we call this the $SBAreduced$ model.

In addition we consider the Extended Fund Facilities as an alternative treatment variable. EFFs are meant for countries with serious medium-term balance of payments problems due to structural weaknesses; compared with SBAs, they generally feature a longer programme engagement and a longer repayment period and are less frequent in our sample, with 15 programmes approved for 11 countries (Table 2).[16]

[12]The country sample includes Argentina, Brazil, Bulgaria, Chile, China, Colombia, Croatia, Cote d'Ivoire, the Dominican Republic, Ecuador, Egypt, El Salvador, Hungary, Lebanon, Malaysia, Mexico, Morocco, Nigeria, Pakistan, Panama, Peru, the Philippines, Poland, Russia, South Africa, Turkey, Ukraine, Uruguay and Venezuela.

[13]The SBA is frequently referred to as the IMF workhorse lending instrument for emerging and advanced market countries; it provides normal access up to a cumulative 435 percent of a members' quota in the IMF, for a duration of no more than 36 months; the resources are disbursed in tranches following regular reviews by the IMF's Executive Board which assesses the overall performance and the country's compliance with the conditionality set out in the adjustment programme attached to the loan, meant to overcome the problems that led the country to seek funding in the first place.

[14]To properly identify $SBAinitial$ ($SBAfinal$) we check whether there are programmes that start (end) before (after) the first (last) date of our sample and exclude the corresponding observations.

[15]In particular, we exclude Argentina, Bulgaria, Croatia, Dominican Republic, Ecuador, El Salvador, Pakistan, Peru, Turkey, Ukraine and Uruguay from the analysis.

[16]EFFs provide assistance to countries: (i) experiencing serious payment imbalances because of structural impediments; or (ii) characterized by slow growth and an inherently weak balance of payments position. Under an EFF, the conditionality featured in the adjustment programme is expected to have a strong focus

Finally, we run our model using the precautionary programmes, Flexible Credit Line (FCL) and Precautionary and Liquidity Line (PLL) as treatment variables. The FCL is designed to give (pre-qualified) countries with very strong policy frameworks and track records in economic performance access to a large amount of resources that can be drawn on upfront, without additional conditions attached to them.[17] Since its introduction in 2009, the FCL has been used by just three countries (Mexico, Poland and Colombia) for a continuous period of time up to the end of our sample (August 2017, through repeated programmes (Table 3).[18] A PLL is meant to give precautionary support to countries with sound fundamentals but with some remaining vulnerabilities; it has a shorter duration and less access to resources than the FCL and so far has been used only by two countries (Morocco and Macedonia).[19]

The approval process for all IMF programmes starts with the request from the member country. Following such a request, an IMF staff team holds discussions with the local government to assess the economic and financial situation, the size of the country's overall financing needs, and agree on the appropriate policy response. Once the IMF and the country authorities reach an agreement on a policy programme, in most cases this is presented to the Fund's Executive Board in a Letter of Intent and further detailed in a Memorandum of Understanding. The programme is subsequently submitted to the Executive Board which meets weekly aat the IMF headquarters and once it gets there it is approved in 100 percent of the cases, according to the historical records. There are no data on the average length of this process, and it may vary according to the specific circumstances; however, it is reasonable to assume that it is not a long one, since it is in the interests of both the IMF and the country to act as swiftly as possible to reduce uncertainty and restore conditions for a stable economy and sustainable growth. The news of a forthcoming IMF loan normally breaks to the public once the IMF delegation has flown out to the country and therefore the anticipation of a programme by the markets does not usually predate the actual approval by too long. In any case, we address possible anticipation effects in the empirical analysis.

The covariates are the composite risk indicators taken from the International Country Risk Guide (ICRG) database. The ICRG database comprises 22 risk components grouped into three major risk categories: political risk (12 components), financial risk (5 compo-

on structural reforms to address institutional or economic weaknesses, in addition to the maintenance of macroeconomic stability. The borrowing limits are the same as for the SBA (cumulatively up to 435 percent of a member's quota) but the duration is longer (up to four years, instead of three) and so is the repayment period (4.5 to 10 years, instead of 3 to 5 years).

[17] Countries with very strong economic fundamentals and policy track records can apply for an FCL when faced with potential or actual balance of payments pressures. Qualified countries have the flexibility to draw on the credit line at any time within a pre-specified period (one or two years); access to IMF resources is upfront (without disbursements in tranches) and does not imply additional conditionality, given the strength of the policy frameworks of the eligible countries; there is no cap on access to IMF resources, and the need for resources is assessed on a case-by-case basis by the Executive Board.

[18] At the time of writing, Colombia and Mexico still have an FCL in place, while Poland exited from the programme in November 2017.

[19] In our sample we have all three FCL countries but only Morocco as a PLL user, we therefore pool these observations in one joint FCL-PLL variable.

nents) and economic risk (5 components). Each component is assigned a maximum numerical value (risk points), with the highest number of points indicating the lowest potential risk for that component and the lowest number (zero) indicating the highest potential risk. The maximum points for any risk component are pre-set and depend on the importance of that component in the overall risk of a country. Political information and economic and financial data are converted into risk points for each individual risk component on the basis of a consistent evaluation pattern.[20]

- The **Economic Risk Rating (ERR)** measures the soundness of the macroeconomic fundamentals and includes five components: per capita GDP, real GDP growth rate, inflation, and the fiscal and current account balances as a percentage of GDP (Table 4). The ERR can take a value of between 0 and 50; low scores signal weak macroeconomic fundamentals, while high scores are associated with sound fundamentals.
- The **Financial Risk Rating (FRR)** evaluates the ability of a country to pay its external debt obligations and is based on: external debt as a percentage of GDP, external debt as a percentage of exports of goods and services, the current account balance as a percentage of exports of goods, the ratio of official reserves holdings to monthly imports, and a measure of nominal exchange rate stability (Table 4). Like the ERR, the FRR can take values of between 0 and 50, with low values indicating a high degree of external vulnerability and high values indicating low vulnerability (or better resilience to external shocks).
- The **Political Risk Rating (PRR)** measures the degree of political stability in a given country. It is based on the following sub-components: government stability, socio-economic conditions, investment profile, internal and external conflict, corruption, military persons in power, religious tensions, law and order, ethnic tensions, democratic accountability and bureaucracy quality (Table 4). The PRR ranges from 0 to 100, with low scores associated with high political risk.

The advantage of using ICRG data lies in the fact that they are available with a monthly frequency and account for different aspects of the countries' situation that have all been found relevant in explaining both spreads [Comelli (2012), Csonto and Ivaschenko (2013)] and participation in IMF programmes [Bal Gunduz (2016)].
In addition to these country-specific factors, the global factors that are used in the analysis are the global risk appetite as measured by the Chicago Board Options Exchange Volatility Index (VIX), which accounts for the implied volatility of S&P index options and the short term US interest rate, as a proxy for global liquidity conditions. These global variables, which are frequently used as spread determinants in the literature,[21] help to take into account the external conditions that countries face, in order to achieve a better matching

[20] The ICRG database is compiled by the PRS Group. A detailed description of the methodology is available at https://www.prsgroup.com/about-us/our-two-methodologies/icrg.
[21] See, for example, Gonzales-Rosada and Levy-Yeyati (2008).

between treated and control units. [22]

Table 5 reports the summary statistics for the variables used in the baseline specification of the model in the overall sample. Table 6 shows mean EMBIG spread values and the risk ratings for countries with an SBA and without any other programme, Table 7 shows the same for countries with an EFF and without any other programme and Table 8 for countries with an FCL or PLL and without any other programme. In general, sovereign spreads are higher on average for countries under an SBA and under an EFF and lower for those under an FCL. Symmetrically, risk ratings are on average worse (i.e. lower) for the former and better (i.e. higher) for the latter. As regards the geographical distribution of the data, Table 9 shows the mean EMBIG spreads and risk ratings across regions, while Tables 10 and 11 show the distributions of SBA and EFF programmes in the different regions.

4 Propensity Score Matching

This paper uses propensity score matching to deal with the selection bias problem that arises when assessing the impact of IMF programmes. The selection bias originates from the fact that countries self-select into an IMF programme, given their pre-programme characteristics; this makes it difficult to attach a causal interpretation to the different outcomes observed in countries supported by a programme and countries not supported.

The PSM technique has been used in the design of non-experimental (or quasi experimental) studies as a means for evaluating the impact of a programme or an intervention.[23] Matching techniques use the information from the units that do not participate in the intervention to identify what would have happened to participating units in the absence of intervention; the comparison between the outcome for participants and observationally similar non-participants allows us to estimate the effect of the intervention. What is crucial in this process is to find a convincing way to pair participants and non-participants which are similar in their observable pre-intervention characteristics, in order to be able to identify the causal effect of the intervention as the difference in the outcome for the two groups after the intervention. This corresponds to creating two well-balanced groups of units, i.e. two groups that are only randomly different with respect to their pre-treatment (observable and unobservable) characteristics. The requirement is that the common variables that affect treatment assignment and treatment-specific outcomes are observable (selection on observables).[24] Given that conditioning on a large number of covariates may prove difficult (curse of dimensionality), the propensity score - the probability of receiving the treatment given a set of observable variables - can be used to reduce the matching problem to a

[22] Alternative measures of US interest rate and volatility are used as robustness checks, as discussed in section 6.

[23] The description of PSM that follows is based on Stuart (2010) and Stuart and Rubin (2008) who provide an excellent basis for understanding and implementing the technique.

[24] As Stuart (2010) puts it, 'this assumption is often more reasonable than it may sound at first since matching on or controlling for the observed covariates also matches on or controls for the unobserved covariates, in as much as they are correlated with those that are observed' (p. 3).

single dimension. Indeed, Rosenbaum and Rubin (1985) show that if potential outcomes are independent of treatment conditional on covariates X, they are also independent of treatment conditional on a balancing score $b(X)$, such as the propensity score.
If all the information that is relevant for participation is observable and can be used for calculating the propensity score, it will be able to produce convincing matching for estimating the impact of the intervention. Therefore two key assumptions for this type of analysis are the following ones: 1) the assignment of the treatment is independent of the potential outcomes, given the propensity score (uncounfoundedness or conditional independence assumption, CIA), and 2) there is a positive probability of receiving the treatment for all levels of the propensity score (common support).
In analytical terms, the propensity score, the conditional independence assumption (CIA) based on the propensity score and the common support can be described, respectively, using the three following equations:

$$p(X) = Pr(D = 1|X) \tag{1}$$

$$Y_0, Y_1 \perp D|p(X); \forall X \tag{2}$$

$$0 < Pr(D = 1|X) < 1 \tag{3}$$

where Y_1 and Y_0 are the outcome for the treated and the untreated units, respectively, and $D = 1$ indicates the treatment status and X the covariates.
In this set-up, the parameter of interest in most cases is the Average Treatment effect on the Treated (ATT), i.e. the difference between the average level of the outcome of interest between treated units and statistically comparable (matched) untreated units. In general, it can be expressed as:

$$ATT = E(Y_1|D = 1) - E(Y_0|D = 1) \tag{4}$$

while its PSM estimator has the following form:

$$\tau_{PSM}^{ATT} = E_{p(X)|D=1} \{E[Y_1|D = 1; p(X)] - E[Y_0|D = 0; p(X)]\} \tag{5}$$

In other words, the PSM estimator is simply the mean difference in outcomes over the common support, appropriately weighted by the propensity score distribution of participants.[25]
There are a series of steps to be taken in order to be able to give a convincing causal interpretation to the ATT estimated through propensity score matching [Stuart and Rubin (2008)]:

- estimate the propensity score using covariates that are related to both the treatment assignment and the outcome, but are not affected by the treatment assignment;

[25]This expression for the PSM estimator for the ATT is taken from Caliendo and Kopeinig (2005).

- select the matches to be used for each treated unit (e.g. use the n nearest neighbours, or a weighted average of all different untreated units with the weights depending on the similarity between the untreated units and the treated one):
- check that the matching is able to balance the matched sample according to the covariates used in the estimation of the propensity score.

Only when the balance is good, i.e. the pools of treated and untreated units in the matched sample are statistically equivalent with respect to the covariates chosen in the analysis (and therefore the treatment assignment can be considered quasi-random), is it possible to exclusively attribute the difference in the outcome between the two groups to the treatment and not to pre-existing differences. This is the reason why in what follows we carefully assess the covariates balancing derived from our matchings.

5 Empirical strategy

In this work we use propensity score matching to assess the impact of IMF programmes on EMBIG spreads, using monthly data for a pool of emerging market countries. As noted above, there are a few studies that use the PSM approach to assess the impact of the IMF programmes, but none of them focus on spreads or use monthly data. In this set-up, the treatment is represented by the programme and the propensity score is the probability of receiving financial assistance from the IMF. It involves the estimation of a participation equation which takes into account most of the variables that may affect the probability of participating in a programme. The literature on the determinants of participation in IMF programmes is reviewed by Bal Gunduz (2016) and Bird *et al.* (2015); their general finding is that economic, financial and political variables all play a role in a country's decision to resort to the IMF and in the IMF's decision to approve a programme. We take this into account and estimate a participation equation using the ICRG economic, financial and political risk ratings as country-specific variables; to obtain better estimates of the propensity scores, we also consider the role of global variables that are relevant for the demand for IMF resources, such as the VIX and the US short-term interest rate.

Here the key assumption is that the three risk ratings are able to account for all the observable (and unobservable) differences among countries before an IMF loan is made and that no major additional characteristic that matters for IMF programme participation and EMBIG spreads is left out. The fact that the risk ratings consider many aspects of a country's economic, financial and political situation, together with the fact that our sample comprises broadly homogeneous countries help in this sense and makes this assumption not too ambitious.

Our strategy is to start from a baseline model and then assess the sensitivity of its results to different specifications of the matching algorithm. In a dedicated section, we also present the robustness tests performed by changing the covariates and the sample coverage. The baseline model is the PSM estimated using the three risk indexes (ERR, PRR and

FRR), the VIX index and the short-term US interest rate as covariates for the participation equation. The risk ratings are lagged by one period to account for potential endogeneity between macroeconomic outcomes and programmes, and the results of the selection equation estimated with a logit are summarized in Table 12, column 1.[26] For the matching, we use the 20 nearest neighbours within a caliper of 0.01 (the matches for the treated units are the 20 untreated units with the closest propensity score within a 0.01 standard deviation points radius) with replacement (one control unit can be used as a match for multiple treated units).[27]

The estimates are performed using the STATA command `psmatch2` by Leuven and Sianesi (2003), based on Rosenbaum and Rubin (1985). The standard errors of the estimated ATT are retrieved through a bootstrapping procedure with 1000 repetitions.[28]

As noted above, since we do not condition on all covariates but only on the propensity score, it has to be checked if the matching procedure is able to balance the distribution of the covariates in both the control and treatment group. The basic idea is to compare the situation before and after matching and check if any significant difference remains after conditioning on the propensity score. To this end, we present tables with detailed diagnostics on the covariates balance,[29] checking the balance for the single covariates and for the overall model. In particular we show the difference in the mean of each covariate in the treated and in the control group before (unmatched) and after the matching (matched), the standardized bias in the two groups and its percentage reduction, the t-test, and the relative p-value for the statistics on the equality of the means in the two groups and the variance ratio. Following Rosenbaum and Rubin (1985), the standardized bias for each covariate X is defined as the difference of sample means in the treated and matched control subsamples as a percentage of the square root of the average of sample variances in both groups. The standardized bias before matching is given by:

$$SB_{before} = 100 * \frac{(\overline{X}_1 - \overline{X}_0)}{\sqrt{0.5 * (V_1(X) + V_0(X))}} \tag{6}$$

[26]The results of the selection equation have the expected signs and are mostly significant, however, as pointed out in Stuart and Rubin (2008), '[..] with propensity score estimation, concern is not with the parameter estimates of the model but rather with the quality of the matches and [..] the key diagnostic is covariate balance in the resulting matching sample' (pag.160).

[27]The results of the baseline model are not significantly sensitive to the change in the number of neighbours used for matching and to the change to the caliper. In addition, the results do not change significantly when we replace the logit with the probit model for the selection equation.

[28]The econometric literature is not definitive on how to calculate the standard errors for the ATT obtained through an estimated propensity score. Abadie and Imbens (2016) show that the matching estimator under the estimated propensity score is consistent and asymptotically normal, but at the same time Abadie and Imbens (2008) argue that the conventional iid bootstrap does not consistently estimate the distribution of pair or one-to-many matching estimators due to the inability of the i.i.d. bootstrap to correctly reproduce the distribution of the number of times each unit is used as a match. In the applied literature, bootstrapped standard errors are widely used to estimate the confidence intervals for the ATT. In our study we calculate the bootstrapped standard errors for our one-to-many matching a cross check of the significance of the same ATT by calculating the bootstrapped errors also for the one-to-one estimator (Table 13).

[29]For the sensitivity and robustness tests, we just present one statistic on the overall bias to avoid an excess of detail.

After the matching it is given by:

$$SB_{after} = 100 * \frac{(\overline{X}_{1M} - \overline{X}_{0M})}{\sqrt{0.5 * (V_{1M}(X) + V_{0M}(X))}} \tag{7}$$

where $\overline{X}_1$ and $V_1(X)$ are the mean and variance for the treated group before the matching, $\overline{X}_0$ and $V_0(X)$ the same statistics for the control group, and $\overline{X}_{1M}$ and $V_{1M}(X)$, $\overline{X}_{0M}$ and $V_{0M}(X)$ are the same items calculated after the matching. A reduction in the bias to under 5 percent is deemed as a good balance by Rubin (2001) and by most empirical studies [Caliendo and Kopeinig (2005)].
Another way to check the balance is through a two-sample t-test to test for differences in the covariate means before and after the matching (Rosenbaum and Rubin (1985)). Differences are expected before matching, but after matching, the covariates should be balanced in both groups and hence no significant differences should be found (p-value close to zero). Finally, we also present the variance ratio for each covariate, which is the ratio between the variance of the treated and control group before and after the matching; in this case, the desired result is a ratio that gets closer to one after matching.
As diagnostics on the overall covariates balance, we report two statistics on the joint standardized bias, Rubin's B and Rubin's R.[30] Rubin's B represents the difference in the means of the propensity scores in the two groups being compared (matched treated and matched untreated), where the difference is measured in terms of the percentage of standard deviations. Rubin's R represents the ratio of the variances of the propensity scores in the two groups (matched treated and matched untreated). Rubin (2001) recommends that Rubin's B should be less than 25 percent while Rubin's R should be between 0.5 and 2 for a satisfactory covariate balance.

6 Results

We start by estimating the baseline model as described above for the SBA variable, which is set equal to 1 at time t if the country is supported by a Stand-By Arrangement.[31] This model is based on 5,342 (on-support) observations, of which 986 are treated; 27 (treated) observations are discarded from the analysis (Table **??**), since they are off-support (the propensity score is so high that it is not possible to find any good match for them in the control group).[32] In this set-up we are able to find a good balance of covariates in

[30] In some parts of the paper we just report Rubin's B to account for the quality of the covariate balance, for the sake of clarity and simplicity; in those cases the single covariate balance is assessed and does not raise concerns.

[31] This model is run excluding all the observations that have another non-concessional IMF programme in place, such as the Extended Fund Facility (EFF).

[32] Note that not all the observations that are on support are used as matches, but only those that are closer to the treated variables they match with, according to the parameters of the matching algorithm. In other words, the off-support observations are not the only ones not used in the matching but there are also other on-support ones that are not used as matches, because they are too far from the treated units. For example, in the baseline SBA model, one of the nearest neighbours for Mexico in July 1997 is Panama in March 2003, or Peru in March 2001 is matched with Mexico in autumn 2004.

the treated and control groups after matching, as shown in Tables 14 and 15. Table 14 reports the diagnostics on the overall balance for each model. One of these diagnostics (first column) is based on the comparison of the pseudo-R2 of the probit re-estimation of the propensity score on the matched and unmatched sample; the pseudo-R2 indicates how well the regressors X explain the participation probability, and after matching there should be no systematic differences in the distribution of covariates between the two groups and the pseudo-R2 should therefore be low. Furthermore, the table shows the p-values of the likelihood-ratio test on the joint insignificance of all regressors (column 2), which should be rejected for the estimation before the matching and accepted afterwards. The same table reports the median and average bias among covariates before and after the matching (columns 3 and 4) and, more importantly, Rubin's B and Rubin's R (column 5 and 6).
Having a good balance is important because it ensures that the treated and untreated groups do not systemically differ in their pre-programme characteristics and enables us to attribute the post-programme differences in the outcome of interest to the programme itself and not to the selection bias. Therefore, after having checked that the balance is good, we can give a causal interpretation to the ATT, which in this case is equal to 182 basis points and is significant with bootstrapped standard errors (first column and first row of Table 13).[33] This means that during the life of the programme, countries supported by an IMF programme record on average an EMBIG spread that is 1.8 percentage points higher with respect to countries in a comparable economic, financial and political condition not supported by the IMF.[34] This result suggests that the Fund's financial assistance does actually carry some stigma for financial markets. One possible explanation for this result is that countries under an IMF programme face significant domestic political challenges and increased political uncertainty that may lead to higher risk premiums charged by the international investors. Another possible reason is the scepticism surrounding the probability of IMF programmes succeeding, linked to the observed track record of a prolonged use of IMF resources (IEO (2012)) and to the criticism on the institution's legitimacy and governance (Reinhart and Trebesch (2016)).
Some countries in our sample do have a set of contiguous programmes over a long time span and this may signal problems in the programmes as well as structural problems that take longer to resolve. To check how much the above reported results are driven by cases implying some sort of failure in the design and/or the implementation of the programme, we re-estimate the model keeping only the countries with just one programme or two stand-alone programmes at different points in time in what we call a $SBAreduced$ model.[35] In this case, the estimated ATT decreases to 63 basis points for the whole programme (Rubin's

[33] As noted before, we cross check that the ATT is of a similar magnitude and still significant with bootstrapped standard errors estimating the model with one-to-one matching (Table 13, column 2).

[34] Following Forbes *et al.* (2015), we also tried to consider a three-month exclusion window which prevents the three months previous to the start of the programme from being used as matches, and the results do not change significantly.

[35] This means dropping Argentina, Brazil, Bulgaria, Colombia, Croatia, the Dominican Republic, Ecuador, El Salvador, Peru, Turkey, Ukraine and Uruguay from the analysis. Therefore in this case our sample comprises 17 countries.

B equal to 5.7 percent) but the result is less robust across different matching techniques (Table 13). Therefore, the result showing the existence of some stigma effect continues to hold, with a smaller magnitude and less significance, also when excluding countries with repeated use of IMF resources from the analysis, whose spreads may be higher due to a lack of credibility on the programme's success.

In order to check the different magnitudes of stigma along different phases of an IMF programme, we repeat the same exercise using different treatment variables: one for the initial six months of the programme ($SBAinitial$) and one for the final six months ($SBAfinal$).[36]

By running the same model for the $SBAinitial$ variable, we again obtain a good balance after matching (see tables 14 and 17; see Table **??** for the common support) and an estimated ATT of 92 basis points; therefore in the initial months of the programme the stigma is already positive even though smaller than over the entire life of the programme.[37]

When using $SBAfinal$ as treatment variable, which accounts for the last six months of a programme (or of a series of programmes), we find a negative ATT of -77 basis points, based on a good balanced sample (see Tables 14 and 18; see Table **??** for the common support). A possible interpretation of this result is that towards the end of the programme, when the economic adjustment promoted by the IMF has taken place and the uncertainty surrounding the success of the intervention dissipates, markets are ready to charge less for lending to countries that are about to exit from IMF support.[38] The stigma actually seems to turn negative at this point and the sovereign spreads for countries in the final six months of an SBA programme have a spread of around 0.8 percentage points lower on average than countries in a comparable economic, financial and political situation, but without IMF support.[39] The estimated ATTs commented above are summarized in Table 13, column 1.

In order to address concerns over a possible anticipation of the concession of an IMF loan by the markets, we check that there are no significant changes in the results when we use an SBA variable lagged by one or three months.

We now turn our attention to precautionary programmes, namely the FCL and the PLL, and we restrict our sample to observations starting from April 2009 (the date when the FCL was included in the Fund's toolkit). The propensity score matching model run for the FCL-PLL variable yields an ATT of -80 basis points using the same matching algorithm and parameters as in the SBA model (20 nearest neighbours matching within a caliper of 0.01)

[36] In the event of consecutive programmes, we consider the last six months of the last programme.

[37] In an alternative specification, we allow the initial six months of a programme to be compared only with months without a programme. In this case the ATT is higher (equal to 188 bp), because we are comparing months under a programme with months without a programme, which tend to be associated with more stable economic conditions and hence relatively lower spreads.

[38] In this case, we compare the last six months of the programme with all periods, including the other months of the programme itself. If we prevent the months under the programme from being used as matches, we obtain a smaller but still negative ATT, of around -14 bp. This result is consistent with the idea that in the last months of the programme the spread is more similar to normal times than to programme times.

[39] The bootstrapped standard errors of the ATT for the $SBAfinal$ model point to a large variability in the estimates, suggesting that the stigma effect towards the end of the programme may be negative or also nil.

with a barely satisfying balance of covariates (B equal 20.6).[40] The matching procedure results in a better balance if performed with one-to-one matching and the resulting ATT does not change significantly, being equal to -78 bp and significant (Table 13); in Tables 14 and 19 we show the balance diagnostics for the one to one matching and in Table **??** the common support. This result supports the idea that the IMF precautionary programmes do succeed in their intention to signal the creditworthiness of the subscribing country and reducing its borrowing costs. In other words, this analysis provides evidence against the existence of a financial market stigma associated with IMF precautionary lending and supports instead the idea of a catalytic effect of this kind of programmes. Nonetheless, this does not rule out other sources of concern, for example the possibility that political stigma may make countries reluctant to resort to precautionary lending. This finding is consistent with the evidence presented in IMF (2014)[41] and in Essers and Ide (2017), who use the synthetic control method to evaluate the impact of FCL on sovereign spreads; however, results on the precautionary facilities should be read keeping in mind that they are obtained from a very restricted sample of countries, given their very low pick-up record.

We refine our analysis by also looking at a special kind of SBA: the precautionary ones. Precautionary SBAs have the same features as traditional SBAs, but are intended to be used in a precautionary way by the country's authorities, who decide not to receive the actual disbursements associated with the programme. In our sample, 14 out of 52 SBAs are precautionary, therefore we focus our analysis only on them and, at the same time, repeat our baseline exercises only on non-precautionary SBAs. The model for precautionary SBAs yields an ATT of -134 bp (with a Rubin's B of 6 percent and a good individual covariate balance), while the rerun of the model excluding the precautionary SBAs does not significantly change the magnitude or the significance of the treatment effects estimated on the full sample (probably because precautionary SBAs represent less than one fifth of the entire SBAs sample in terms of monthly observations). Therefore, the effect on sovereign spreads of precautionary SBAs can be assimilated into those of fully precautionary programmes such as the FCL, thereby helping emerging countries to lower their borrowing costs.

The difference in the effect on the spreads of traditional and precautionary programmes is one important result of this analysis and can be interpreted by looking at the different circumstances in which a country finds itself asking for IMF intervention in the two cases. A country asks for an SBA or an EFF when it has an actual balance of payment problem that it seeks to resolve through the financial help and the adjustment programme provided by the IMF, and this may give the markets a negative signal on te country's circumstances; the FCL is only granted to countries that are deemed by the IMF to be in very good shape

[40]In this case we preclude the observations either with an SBA or an EFF to be used as matches. For example, the nearest neighbour for Colombia in the first month of use of the FCL, April 2009, is Croatia in January 2017, while for Colombia in the same month it is Brazil in January 2004.

[41]IMF (2014) features a brief analysis of the impact of FCLs on sovereign spreads. They run a simple panel regression and the selection bias problem is not addressed. The resulting impact on spreads has a negative sign but is only significant if the analysis is restricted to Mexico and Colombia.

in terms of economic fundamentals and policies and this is the information that is conveyed to the markets when an FCL is granted.

6.1 Robustness

6.1.1 Alternative specifications of the PSM model

The results presented so far are based on a peculiar specification of the PSM model, which is found to ensure a good balance between covariates and therefore gives credible estimates of the ATT. Robustness tests on the ATT have to be based on alternative models that ensure an equally good balance, otherwise we would be comparing a meaningful ATT with an unreliable one.

In an initial robustness test we tried different matching algorithms, as was also done in Forbes *et al.* (2015). In particular, we compare the baseline results with the simpler 1:1 nearest neighbour without replacement (meaning that a control can be used as a match for just one treated individual), radius matching with caliper, local-linear regression and kernel weighted matching.[42] Table 13 compares the estimated ATT and the Rubin's B resulting for these 5 models (including the baseline), for the different definitions of the programme variable.[43] It shows that our main results, namely a positive ATT with a magnitude of between 1 and 2 percentage points for the whole programme, a negative ATT of between 0.5 and 1 percentage points for $SBAfinal$ and a negative ATT around one percentage point for FCL-PLL, hold across the use of different matching algorithms, except for the kernel weighting, which in some cases does not allow us to find a satisfying covariate balance.

In addition, we look at the data for Extended Fund Facilities (EFFs) instead of SBAs. Repeating the PSM model with EFFs yields results consistent with our baseline with SBAs: the estimated ATT is 87 basis points and the balance is satisfactory (B is equal to 10.2 percent). We cannot repeat the analysis for the last months of EFF programmes, since the treated observations are very few and it is not possible to find a satisfactory balance after matching.[44]

Our main results also hold when pooling observations for SBAs and EFFs in a more general IMF programme dummy variable. In this case the resulting ATT is 147 basis points, and the associated B is 2.5 percent.

[42] Radius matching uses all the comparison groups within a maximum propensity score distance (the caliper in our case is set at 0.01); local-linear and kernel matching are nonparametric estimators that compare the outcome of each treated observation with a weighted average of the outcomes of all the untreated ones, with the highest weight being placed on those with scores closest to the treated individual. See Heinrich *et al.* (2010) for a discussion of these different methodologies.

[43] We only show the Rubin's B as a balance diagnostic for ease of presentation, but in each model we also checked the balance for each covariate.

[44] In this case we also have to exclude countries that were still under an EFF at the end of the sample (August 2017), since they could not be considered as being in the final months (Egypt, Ukraine and Cote d'Ivoire).

6.1.2 Using indices' components

In order to assess the robustness of our results, we also try to run the model using the single risk components of the composite ICRG indices (ERR, FRR and PRR). We estimate multiple models using the components of each index one at a time and obtain results in line with our baseline specification, although the covariate balance is less satisfying than in the baseline exercise, given the increased difficulty of balancing the samples over a greater number of variables.[45] When using the components of the ERR, together with the FRR and PRR composite indices and the global variables, we get an ATT of around 100 bp and an overall balance diagnostic (Rubin's B) of slightly below 25 percent for the SBA and an ATT of -73 bp for the FCL (B is equal to 24 percent). With the FRR components and the ERR and PRR composite indices the result is again an ATT of around 100 basis points with a Rubin's B of 21 percent (for the respective FCL model, the ATT equals -76 bp and B is 18 percent). The model with the PRR components has a larger number of covariates, therefore we are not able to find a satisfying balance after matching (in this case we get, for the SBA, an ATT of around 140 bp and a B of over 30 percent; for the FCL, an ATT of -116 bp and a B of almost 50 percent). We also try to pick only the risk components of the indices that, according to the literature on IMF programme participation, seem to matter most for the probability of requesting and receiving the Fund's support, namely those related to international reserves, GDP Growth, foreign debt, budget balance, government stability, democratic accountability and bureaucracy quality; in this specification we obtain an ATT of 230 bp (not too far from our baseline) and a Rubin's B of around 10 percent (for the FCL model, the ATT is -76 bp and Rubin's B is 18 percent).

6.1.3 Entropy balancing methodology

The use of a larger number of covariates is possible when switching to an alternative empirical model to address the selection bias, namely the entropy balancing methodology developed by Hainmueller (2012).[46] Entropy balancing may be seen as a generalization of the conventional propensity score model, since it is a scheme to preprocess data in order to obtain a balance in covariates between the treated and control groups. However, instead of searching for a balance after the preprocessing, the entropy balancing approach searches directly for a set of unit weights that balances the covariate distributions with respect to the specified moments. In this way, the estimation of the propensity score is no longer needed, as it is not the balance check, and it is possible to perform the analysis on a larger set of covariates. Once the set of weights that balances the covariates distribution

[45] If instead of this step-by-step approach we consider a model with all the components of the indices at one time, we cannot find a satisfying balance and are also likely to encounter collinearity problems.

[46] This methodology has recently been used by Neuenkirch and Neumeier (2016) to assess the impact of U.S. sanctions on poverty, by Balima (2017) to analyse the effect of domestic sovereign bond market participation on financial dollarization and by Balima and Sy (2019) to evaluate the impact of IMF programmes on the likelihood of sovereign defaults.

according to the specified moments is found, it is possible to perform a weighted regression on the outcome variable and look at the coefficient of the treatment indicator.[47] In our case we found a set of weights able to balance the mean of all the components of the risk indices used in the baseline analysis (22 variables, see Table 4) plus the global variables (VIX and US interest rate); the coefficient of the SBA dummy variable in the EMBIG regression on the reweighted data is significant and equal to 232 bp; the same coefficient is significant and equal to -138 bp for the $SBAfinal$ dummy and -134 bp for the FCL-PLL dummy.[48] Our main results are confirmed in this new set-up, with a larger set of covariates.

6.1.4 Quarterly macro data

As a further robustness check, we repeat the analysis using a different sample, which comprises macro variables at a lower frequency, with quarterly data; the country sample and the time span remain the same. Specifically we use the quarterly average of EMBIG spreads as the outcome variable and the following variables as covariates: GDP growth, real GDP, the reserves-to-imports ratio, the public debt-to-GDP ratio, the current account balance-to-GDP ratio, Political Risk Rating, VIX and the US Federal Funds rate.[49] By using the same baseline specification of the PSM model as in section 4[50] with the new data, we obtain a significant ATT for the SBA of 351 bp (with a good covariate balance, B is equal to 13.8 percent and a negative and significant ATT for the FCL-PLL, equal to -154 bp (B is equal to 8.8 percent).

6.1.5 Other covariates and subsamples

Other robustness checks include the use of other measures of US interest rates instead of the three-month one and a measure of the global risk appetite alternative to the VIX. Our main results do not change when using the Federal Funds rate or the shadow federal funds rate as calculated by Wu and Xia (2015).[51] Similarly, our results continue to hold if we use the global financial cycle indicator developed by Miranda-Agrippino and Rey (2015).[52] As a further check, we pool countries by region to enable a comparison between countries that belong to the same region and should in principle be more similar. In this case we are only able to find a good balance for the South and Central America region, for which we obtain a positive and significant ATT for SBA programmes and a negative effect for the last six month of the programmes. In addition, we run our model after having excluded

[47] For details on the implementation of entropy balance see also Hainmueller and Xu (2013).

[48] The same covariates used for the balancing are included as controls in the EMBIG regression, as suggested by Hainmueller (2012).

[49] The macro variables' source is the IMF's International Financial Statistics; public debt is the only variable with an annual frequency and comes from the International Debt Statistics of the World Bank.

[50] Nearest neighbour matching with 20 neighbours within a caliper of 0.01.

[51] Unlike the observed short-term interest rate, the shadow rate is not bounded below by 0 percent. Whenever the Wu-Xia shadow rate is above 1/4 percent, it is equal to the Federal Fund rate.

[52] This global factor indicator is only available up until 2012, so in this case we shorten the time horizon of the analysis to the years from 1997 to 2012.

some outlier countries, namely Argentina, Cote d'Ivoire, Turkey and Ukraine.[53] After these exclusions, for the SBA model the ATT remains positive but turns out smaller (106 basis points) and the balance is good (B equal to 5.5 percent); for the $SBAfinal$ the ATT is very small and negative (-8 basis points) with a B of 8.3 percent; for FCL-PLL the ATT keeps being negative and is equal to -86 basis points (with a B of 11.7 percent).
Finally, if we split the sample in two subperiods (1997 to 2008 and 2009 to 2017) the overall results of a positive ATT for the SBA holds in both periods, the ATT for the last six months of the programmes is still negative in the first subperiod, while no conclusion can be drawn for the second subperiod, due to a lack of covariate balance after the matching.

7 Conclusions

This paper reports robust and significant evidence on the existence of some financial market stigma for traditional non-concessional IMF lending programmes (namely, SBAs and EFFs). We find that financial markets apply an additional risk premium to countries supported by an IMF programme, as compared with countries that are in very similar economic, financial and political conditions but are not supported by the Fund. However, such a risk premium tends to dissipate towards the end of the programme, when it is more likely that the economic adjustment promoted by the IMF has taken place and there is less uncertainty over the success of the programme. Furthermore, such an additional risk premium has a smaller magnitude and is less significant when we restrict the analysis to countries that did not have consecutive IMF programmes, suggesting that part of the estimated financial stigma is linked to the lack of success of the programme itself, rather than to its mere existence. At the same time, we find evidence of a positive effect of IMF precautionary programmes on sovereign spreads, in the sense that countries with an FCL or a PLL enjoy lower borrowing costs than comparable countries. This result supports the idea that IMF precautionary lending is successful in its declared intent of reducing market pressure in those countries that are eligible for precautionary support.
It has to be noted that this kind of study, like all those on impact evaluation, does not explore the mechanism through which an IMF loan may give rise to financial market stigma; consequently, it cannot support specific policy recommendations to deal with this issue. c

[53]These are countries whose time series of sovereign spreads behave very differently with respect to countries in the same region.

Table 1: Month of approval and duration of SBAs in the sample

Country	**Time**	**Duration** (months)
Argentina	1997m1	13
Argentina	2000m3	34
Argentina	2003m1	8
Argentina	2003m9	28
Brazil	1998m12	33
Brazil	2001m9	12
*Brazil	2002m9	30
Bulgaria	1997m1	18
Bulgaria	2002m2	26
*Bulgaria	2004m8	32
*Colombia	2003m1	28
*Colombia	2005m5	18
*Croatia	2001m3	15
*Croatia	2003m2	15
*Croatia	2004m8	28
Dominican Republic	2003m8	16
Dominican Republic	2005m2	36
Dominican Republic	2009m11	28
Ecuador	2000m4	20
Ecuador	2003m3	13
Egypt	1997m1	21
*El Salvador	1997m2	16
*El Salvador	1998m9	18
*El Salvador	2009m1	14
*El Salvador	2010m3	36
Hungary	1997m1	14
Hungary	2008m11	23
Mexico	1999m7	16
Nigeria	2000m8	14
Pakistan	1997m1	9
Pakistan	2000m11	10
Pakistan	2008m11	24
Panama	2000m7	20
*Peru	2001m3	10
*Peru	2002m2	24
Peru	2004m6	26
Peru	2007m1	27
Philippines	1998m4	32
Russia	1999m7	17
Turkey	1999m12	26
Turkey	2002m2	36
Turkey	2005m5	37
Ukraine	1997m8	12
*Ukraine	2004m3	12
Ukraine	2008m11	20
Ukraine	2010m7	29
Ukraine	2014m4	11
Uruguay	1997m6	26
Uruguay	1999m3	12
Uruguay	2000m5	22
Uruguay	2002m4	36
Venezuela	1997m1	7

* indicates a programme treated as precautionary by the country's authorities

Table 2: Month of approval and duration of EFFs in the sample

Country	**Time**	**Duration** (months)
Argentina	1998m2	25
Bulgaria	1998m9	37
Cote d'Ivoire	2016m12	(9)
Croatia	1997m3	37
Egypt	2016m11	(10)
Pakistan	1997m10	37
Pakistan	2002m1	36
Pakistan	2013m9	37
Panama	1997m12	31
Peru	1997m1	27
Peru	1999m6	21
Philippines	1997m1	15
Russia	1997m1	27
Ukraine	1998m9	49
Ukraine	2015m3	(30)

In parentheses programmes still ongoing at the end of the sample (2017m8).

Table 3: Month of approval and duration of FCL-PLLs in the sample

Country	**Time**	**Duration** (months)
Colombia	2009m5	12
Colombia	2010m5	12
Colombia	2011m5	24
Colombia	2013m6	24
Colombia	2015m6	12
Colombia	2016m6	(14)
Mexico	2009m4	12
Mexico	2010m3	10
Mexico	2011m1	23
Mexico	2012m11	24
Mexico	2014m11	19
Mexico	2016m5	(15)
Poland	2009m5	12
Poland	2010m7	5
Poland	2011m1	24
Poland	2013m1	24
Poland	2015m1	24
Poland	2017m1	(8)
Morocco (PLL)	2012m8	23
Morocco (PLL)	2014m7	24
Morocco (PLL)	2016m7	(13)

In parentheses programmes still ongoing at the end of the sample (2017m8).

Table 4: Risk Rating Components

Component	**Points**
Economic Risk Rating Components	
Per capita GDP	5
Real GDP growth rate	10
Annual Inflation Rate	10
Budget balance as a percentage of GDP	10
Current account as a percentage of GDP	15
Total	**50**
Financial Risk Rating Components	
Foreign Debt as a percentage of GDP	10
Foreign Debt Service as a percentage of Exports of Goods and Services	10
Current Account as a percentage of Exports of Goods and Services	15
Official Reserves Holdings as a percentage of monthly imports	5
Exchange Rate Stability	10
Total	**50**
Political Risk Rating Components	
Government Stability	12
Socioeconomic Conditions	12
Investment Profile	12
Internal Conflict	12
External Conflict	12
Corruption	6
Military in Politics	6
Religious Tensions	6
Law and Order	6
Ethnic Tensions	6
Democratic Accountability	6
Bureaucracy Quality	4
Total	**100**

Table 5: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
EMBIG	496.782	677.560	0	7078	6754
ERR	35.03	4.358	12.5	45.5	7154
FRR	37.235	4.922	15	49	7154
PRR	64.191	9.736	36.5	87	7154
VIX	20.629	8.066	10.31	62.25	7192
US3m	2.092	2.119	-0.01	6.37	7192

Table 6: Variables means by programme: SBA

	EMBIG	FRR	ERR	PRR
0	441.8	37.8	35.2	63.8
1	806.4	34.0	33.7	65.7
Total	496.7	37.2	35.0	64.1

Table 7: Variables means by programme: EFF

	EMBIG	FRR	ERR	PRR
0	481.2	37.4	35.1	64.4
1	754.8	34.4	33.2	60.1
Total	496.7	37.2	35.0	64.1

Table 8: Variables means by programme: FCL and PLL

	EMBIG	FRR	ERR	PRR
0	513.9	37.1	34.9	64.0
1	195.5	38.6	35.6	67.3
Total	496.7	37.2	35.0	64.1

Table 9: Variables means by region

	EMBIG	FRR	ERR	PRR
Asia	375.4	38.19	35.09	60.81
South and Central America	585.8	37.02	35.43	65.53
East Europe	336.3	35.19	34.54	72.39
Africa	618.8	38.37	34.43	57.75
Total	496.8	37.23	35.02	64.19

Table 10: SBA programme by region

SBA	Asia	S.C. America	EastEurope	Africa	Total
0	1530	2312	979	1201	6025
1	206	664	261	36	1167
Total	1736	2976	1240	1240	7192

Table 11: EFF programme by region

EFF	Asia	S.C. America	EastEurope	Africa	Total
0	1584	2835	1087	1221	6727
1	152	141	153	19	465
Total	1736	2976	1240	1240	7192

Table 12: Selection equation

	(1) SBA	(2) SBAinitial	(3) SBAfinal	(4) FCL-PLL	(5) SBAreduced
lagFRR	-0.160***	-0.0993***	-0.0742***	0.0332**	-0.04***
	(0.00854)	(0.0152)	(0.0133)	(0.0140)	(0.0075)
lagPRR	0.0223***	-0.0116	0.0118	0.0886***	-0.0078**
	(0.00389)	(0.00726)	(0.00764)	(0.00872)	(0.0036)
lagERR	-0.0212**	-0.000114	0.0104	-0.0400**	0.0337***
	(0.00946)	(0.0172)	(0.0176)	(0.0162)	(0.0769)
VIX	0.00112	0.0484***	-0.00954	-0.0242**	0.033***
	(0.00445)	(0.00818)	(0.0110)	(0.0111)	(0.0031)
US3m	0.191***	0.260***	0.228***	-0.0164	0.197***
	(0.0157)	(0.0361)	(0.0312)	(0.266)	(0.0184)
_cons	3.036***	-1.014	-2.114***	-6.806***	0.375
	(0.356)	(0.660)	(0.651)	(0.647)	(0.285)
pseudo R^2	0.137	0.101	0.050	0.060	0.183
N	6685	6685	6685	2661	3999

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Logit model for columns (1) to (4), probit model for column (5)

Table 13: Estimated ATT for different matching procedures

	Baseline	Nearest neighbor	Radius caliper	Local-linear	Kernel
SBA					
ATT	**182***	184***	150***	172***	174***
Rubin's B (%)	5.2	4.4	6.8	5.0	8.0
SBA reduced					
ATT	**67**	62	71	84**	122***
Rubin's B (%)	4.9	7.0	4.5	20.6	27.1
SBAinitial					
ATT	**92**	95	96	123*	229
Rubin's B (%)	8.2	17.5	6.0	13.6	42.6
SBAfinal					
ATT	**-77**	-61	-69	-102***	19
Rubin's B (%)	3.2	19.8	7.1	19.9	49.7
FCL-PLL					
ATT	**-80***	-78***	-76***	-124***	-102***
Rubin's B (%)	20.6	15.0	20.8	21.2	11.9

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, bootstrapped standard errors with 1000 repetitions

Group	Off-support	On-support	Total
		SBA	
Untreated	0	5342	5342
Treated	27	986	1013
Total	27	6328	6355
		SBA reduced	
Untreated	0	3667	3667
Treated	0	176	176
Total	0	3843	3843
		SBAinitial	
Untreated	0	6194	6194
Treated	1	160	161
Total	1	6354	6355
		SBAfinal	
Untreated	0	6155	6155
Treated	1	199	200
Total	1	6354	6355
		FCL-PLL	
Untreated	0	2297	2297
Treated	16	348	364
Total	16	2645	2661

Table 14: Overall balance diagnostics

Sample	Pseudo R2	p-value LR test	Mean bias	Median bias	Rubin's B(%)	Rubin's R
			SBA			
Unmatched	0.148	0.000	42.1	44.4	101.6	1.14
Matched	0.000	0.931	1.6	0.8	5.2	0.74
			SBA reduced			
Unmatched	0.180	0.000	63.4	65.6	144.7	0.61
Matched	0.000	0.999	1.9	2.9	4.9	0.67
			SBAinitial			
Unmatched	0.102	0.000	47.1	43.7	102.4	1.17
Matched	0.001	0.990	3.2	3.5	8.2	0.66
			SBAfinal			
Unmatched	0.050	0.000	25.5	11.1	71.1	0.88
Matched	0.000	1.000	1.2	1.1	3.2	0.67
			FCL-PLL			
Unmatched	0.06	0.000	21.3	9.6	66.5	0.62
Matched	0.004	0.559	4.8	6.4	15.1	0.59

Table 15: Covariate balance for the SBA model

Variable	Unmatched Matched	Mean Treated	Mean Control	bias %	bias %reduction	t-test p>t	$V(T)/V(C)$
lagFRR	U	33.56	38.15	-95.6		0.00	1.18
	M	33.98	34.02	-0.8	99.2	0.85	0.91
lagPRR	U	65.09	63.98	12.4		0.01	0.6
	M	65.14	65.10	0.4	96.9	0.93	0.57
lagERR	U	33.46	35.42	-44.4		0.00	1.13
	M	33.71	33.70	0.2	99.5	0.96	0.78
VIX	U	21.14	20.16	12.2		0.01	0.86
	M	21.04	21.30	-3.3	73.3	0.46	0.89
US3m	U	2.56	1.65	45.7		0.00	1.06
	M	2.52	2.59	-3.2	93.1	0.51	0.82

Table 16: Covariate balance for the SBA model with a reduced sample

Variable	Unmatched Matched	Mean Treated	Mean Control	bias %	bias %reduction	t-test p>t	$V(T)/V(C)$
lagFRR	U	34.61	38.82	-102.6		0.00	0.46
	M	34.61	34.62	-0.4	99.6	0.96	0.45
lagPRR	U	61.46	63.98	-21.2		0.00	1.23
	M	61.46	61.45	0.1	99.5	0.99	1.12
lagERR	U	33.05	35.60	-61.1		0.00	0.63
	M	33.05	32.91	3.2	94.7	0.78	0.49
VIX	U	25.44	20.25	65.6		0.00	0.86
	M	25.44	25.68	-3.0	95.5	0.80	0.59
US3m	U	3.34	1.83	66.7		0.00	1.57
	M	3.34	3.2	2.9	95.7	0.79	1.39

Table 17: Covariate balance for the SBAinitial model

Variable	Unmatched Matched	Mean Treated	Mean Control	bias %	bias %reduction	t-test p>t	$V(T)/V(C)$
lagFRR	U	33.38	37.52	-84.7		0.00	0.98
	M	33.46	33.70	-5.1	94.0	0.67	0.73
lagPRR	U	63.49	64.17	-7.3		0.34	0.86
	M	63.53	63.42	1.2	84.2	0.92	0.86
lagERR	U	33.21	35.16	-43.3		0.00	1.14
	M	33.27	33.46	4.1	90.6	0.73	0.84
VIX	U	24.37	20.21	43.7		0.01	1.69
	M	24.36	24.69	-3.5	92.0	0.78	1.08
US3m	U	2.97	1.77	56.6		0.00	1.32
	M	2.95	3.00	-2.2	96.1	0.85	1.08

Table 18: Covariate balance for the SBAfinal model

Variable	Unmatched Matched	Mean Treated	Mean Control	bias %	bias %reduction	t-test p>t	$V(T)/V(C)$
lagFRR	U	34.99	37.50	-54.1		0.00	0.74
	M	35.09	35.08	0.2	99.5	0.98	0.56
lagPRR	U	65.12	64.13	11.1		0.15	0.67
	M	65.07	65.25	-2.1	80.8	0.82	0.70
lagERR	U	34.80	35.12	-7.6		0.30	0.87
	M	34.86	34.92	-1.1	85.1	0.91	0.78
VIX	U	19.91	20.33	-5.2		0.48	0.91
	M	19.90	19.82	1.0	80.9	0.91	1.33
US3m	U	2.79	1.77	49.7		0.00	1.15
	M	2.77	2.74	1.6	96.9	0.87	1.03

Table 19: Covariate balance for the FCL-PLL model

Variable	Unmatched Matched	Mean Treated	Mean Control	bias %	bias %reduction	t-test p>t	$V(T)/V(C)$
lagFRR	U	38.58	38.93	-9.6		0.13	0.38
	M	38.67	38.89	-6.4	33.0	0.39	0.39
lagPRR	U	67.42	61.98	64.5		0.34	0.60
	M	66.86	66.92	-0.7	98.9	0.91	0.98
lagERR	U	35.61	34.86	20.6		0.00	0.31
	M	35.57	35.64	-1.8	91.1	0.78	0.46
VIX	U	17.65	18.11	-7.9		0.17	0.98
	M	17.70	18.12	-7.1	9.3	0.35	0.96
US3m	U	0.17	0.16	4.2		0.44	1.12
	M	0.16	0.18	-8.1	-93.3	0.30	0.82

References

Abadie, A. and **Imbens, G. W.**, 2008. "On the failure of the bootstrap for matching estimators". *Econometrica*, (76): 1537–1557.

Abadie, A. and **Imbens, G. W.**, 2016. "Matching on the estimated propensity score". *Econometrica*, volume 84(2): 781–807.

Al Sadiq, A. J., 2015. "The impact of IMF-supported programs on FDI in low-income countries". *IMF Working paper no. 157.*

Andone, I. and **Scheubel, B.**, 2017. "IMF stigma: the role of own and neighbour's experience". *mimeo.*

Atoyan, R. and **Conway, P.**, 2006. "Evaluating the impact of IMF programs: a comparison of matching and instrumental-variable estimators". *Review of International Organizations*, volume 1: 99–124.

Bal Gunduz, Y., 2016. "The economic impact of short-term IMF engagement in low-income countries". *World Development*, volume 87: 30–49.

Balima, H., 2017. "Do domestic bond markets participation help reduce financial dollarization in developing countries?" *Economic Modelling*, volume 66: 360–377.

Balima, H. and **Sy, A.**, 2019. "The impact of bailouts on the probability of sovereign debt crises: Evidence from IMF-supported programs". *IMF working paper no. 2.*

Barro, R. J. and **Lee, J.-W.**, 2005. "IMF Programs: Who is chosen and what are the effects?" *Journal of Monetary Economics*, volume 52: 1245–1269.

Bas, M. and **Stone, R.**, 2014. "Adverse selection and growth under IMF programs". *The review of international organizations*, volume 9(1): 1–28.

Bird, G., 2001. "IMF Programs: Do They Work? Can They Be Made To Work Better?"". *World Development*, volume 29(11): 1849–1865.

Bird, G., **Mylonas, J.** and **Rowlands, D.**, 2015. "The political economy of participation in IMF programs: A disaggregated empirical analysis". *Journal of Economic Policy Reform*, volume 18(3): 221–243.

Bird, G. and **Rowlands, D.**, 2017. "The effect of IMF programmes on economic growth in low income countries: An empirical analysis". *The Journal of Development Studies.*

Caliendo, M. and **Kopeinig, S.**, 2005. "Some practical guidance for the implementation of propensity score matching". *IZA Discussion Paper No. 1588.*

Chapman, T., **Fang, S.**, **Li, X.** and **Stone, R. W.**, 2015. "Mixed signals: IMF lending and capital markets". *British Journal of Political Science*, volume 47: 329–349.

Comelli, F., 2012. "Emerging market sovereign bond spreads: Estimation and back-testing". *Emerging Markets Review*, volume 13(4): 598–625.

Cottarelli, C. and **Giannini, C.**, 2002. "Bedfellows, hostages or perfect strangers? Global capital markets and the catalytic effect of IMF crisis lending". *IMF working paper no. 193.*

Csonto, B. and **Ivaschenko, I.**, 2013. "Determinants of sovereign bond spreads in emerging markets: Local fundamentals and global factors vs. ever-changing misalignments". *IMF Working Paper No. 164/2013.*

Dreher, A., 2006. "IMF and economic growth: The effects of programs, loans, and compliance with conditionality". *World Development*, volume 34(5): 769–788.

Dreher, A. and **Walter, S.**, 2010. "Does the IMF help or hurt? the effect of IMF programs on the likelihood and outcome of currency crises". *World Development*, volume 38(1): 1–18.

Eichengreen, K. K., Barry and **Mody, A.**, 2006. "The IMF in a world of private capital markets". *Journal of Banking and Finance*, volume 30(5): 1335–1357.

Essers, D. and **Ide, S.**, 2017. "The IMF and precautionary lending: an empirical evaluation of the selectivity and effectiveness of the flexible credit line". *National Bank of Belgium Working Paper Research no. 323.*

Forbes, K., **Fratzscher, M.** and **Straub, R.**, 2015. "Capital-flow management measures: What are they good for?" *Journal of International Economics*, volume 96(S1): 76–97.

Gehring, K. and **Lang, V.**, 2018. "Stigma or cushion? IMF programs and sovereign creditworthiness". *CIS Working Paper no. 98,.*

Gonzales-Rosada, B. and **Levy-Yeyati, E.**, 2008. "Global factors and emerging markets spreads". *The Economic Journal*, volume 118(533): 19171936.

Hainmueller, J., 2012. "Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies". *Political Analysis*, volume 20(1): 25–46.

Hainmueller, J. and **Xu, Y.**, 2013. "ebalance: a stata package for entropy balancing". *Journal of Statistical Software*, volume 54(7).

Heinrich, C., **Maffioli, A.** and **Vazquez, G.**, 2010. "A primer for applying propensity-score matching". *Inter-american Development Bank Technical notes no. IDB-TN-161.*

IEO, 2012. "Evaluation of prolonged use of IMF resources". *IMF Independent Evaluation Office Report.*

IMF, 2014. "Review of flexible credit line, the precautionary and liquidity line, and the rapid financing instrument".

IMF, 2017. "Adequacy of the global financial safety net - considerations for fund toolkit reform".

IRC, Task force on IMF issues, 2018. "Strengthening the global financial safety net". *ECB occasional paper no. 207.*

Ito, T., 2002. "Can asia overcome the IMF stigma?" *American Economic Review: Papers and Proceedings*, volume 102(3): 198–202.

Leuven, E. and **Sianesi, B.**, 2003. "PSMATCH2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. This version: 4.0.12, 30jan2016".

Miranda-Agrippino, S. and **Rey, H.**, 2015. "Us monetary policy and the global financial cycle". *NBER, Working Paper n. 21722.*

Mody, A. and **Saravia, D.**, 2003. "Catalyzing capital flows: do IMF-supported programs work as commitment devices?" *IMF Working paper no. 100.*

Neuenkirch, M. and **Neumeier, F.**, 2016. "The impact of US sanctions on poverty". *Journal of Development Economics*, volume 121: 110–119.

Papi, L., **Presbitero, A.** and **Zazzaro, A.**, 2015. "IMF lending and banking crises". *IMF Economic Review*, volume 63(3).

Przeworski, A. and **Vreeland, J. R.**, 2000. "The effect of IMF programs on economic growth". *Journal of Development Economics*, volume 62: 385–421.

Reinhart, C. M. and **Trebesch, C.**, 2016. "The International Monetary Fund: 70 years of reinvention". *Journal of Economic Perspectives*, volume 30(1): 3–28.

Rosenbaum, P. and **Rubin, D.**, 1985. "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score". *The American Statistician*, volume 39(1).

Rubin, D., 2001. "Using propensity scores to help design observational studies: Application to the tobacco litigation". *Health Services and Outcomes Research Methodology*, volume 2(3-4): 169–188.

Scheubel, B., **Tafuro, A.** and **Benjamin, V.**, 2018. "Stigma? What stigma? A contribution to the debate on financial market effects of IMF lending". *ECB Working paper no. 2198.*

Stuart, E. A., 2010. "Matching methods for causal inference: A review and look forward". *Statistical Science*, volume 25(1): 1–21.

Stuart, E. A. and **Rubin, D.**, 2008. "Best practices in quasi experimental designs: matching methods for causal inference". In *Best practices in quantitative methods*, chapter 11, pp. 155–176. Thousand Oaks, CA: SAGE Publications Ltd.

Wu, J. C. and **Xia, F. D.**, 2015. "Measuring the macroeconomic impact of monetary policy at the zero lower bound". *Chicago Booth Research Paper No. 13-77.*