Better two eyes than one: a synthesis classification of exchange rate regimes

Cécile Couharde* and Carl Grekou†

Abstract:

This paper proposes a new de facto classification of exchange rate regimes, the synthesis classification. The proposed framework has several advantages over existing de facto classifications. First, it offers a unified framework based on the most divergent classifications, the RR and LYS classifications, leading not only to a broader coverage but also to encompass a broad spectrum of exchange systems. Second, it fits better with the known history of exchange rate regimes developments in the post-Bretton Woods era. Among others, it brings an interesting nuance to the so-called hollowing-out hypothesis by showing that the evolution of de facto regimes —especially in emerging economies since the late 1990s— has essentially involved movement toward more tightly "managed" intermediate regimes and not a shift away from such regimes. As an illustration of the insightfulness of our classification, we empirically revisit the nexus between currency crises and exchange rate regimes. In addition to associate a higher probability of currency crisis to both intermediate and floating regimes, our classification, also displays better statistical performances than other classifications in predicting currency crises.

JEL codes: E52; F33; F4; O24.

Keywords: Currency crisis; De facto classifications; Exchange rate regimes; Probit model; ROC analysis.

*EconomiX-CNRS, University of Paris Nanterre, France. Email: cecile.couharde@parisnanterre.fr.

†CEPII and EconomiX-CNRS, University of Paris Nanterre, France. Email: carl.grekou@cepii.fr.

The synthesis classification is available here.

The online appendix is available here.

We thank Anne-Laure Delatte, Anthony Edo, Thomas Grjebine, Jean Imbs, Sébastien Jean, Valérie Mignon, Gianluca Orefice and Farid Toubal for very helpful discussions and comments. We also thank Anderson Harald and Carmen Reinhart for generously sharing their data with us. We are also indebted to seminar participants at CEPII, Banque de France, EconomiX (University Paris Nanterre), and LEM (University of Lille). All remaining errors are ours.

An earlier version of this paper has been circulated under the title "The Fata Morgana of Exchange Rate Regimes: Reconciling the *LYS* and the *RR* classifications".

1. Introduction

Since the collapse of the Bretton Woods system, academics and policy-makers have always scrutinized exchange rate regimes (ERR) as governments' essential tools for avoiding excessive imbalances and fostering growth. However, since the early 2000s, there has been a clear ramp-up in this scrutiny owing not only to the emergence of several financial crises —especially in emerging markets in the 1990s and early 2000s[1], but also to the progress made in understanding exchange rate policies across the globe and in assigning currencies to an ERR category. Indeed, considerable light has been shed on what countries effectively did for two decades with the development of the *de facto* ERR classifications. Nevertheless, no modus operandi has emerged from these different approaches about the way to assign a currency to a specific ERR category.[2] Consequently, the empirical literature —on the choices and effects of ERR— is still marked by a severe lack of robustness or even substantial disagreements. This situation results in part from the absence of unified definitions of ERR categories. Even if there is a kind of agreement on what an ERR category is, the underlying state of a regime (fixed/intermediate/float) is not always directly observable and has to be inferred. Meanwhile, the methods for inferring a regime category diverge across classifications.

Since the choice of an exchange rate regime is at the crux of several debates in international economics, steps toward untangling this knot, this imbroglio, are critical. As Rose (2011) notes, *"... it is scary that one can no longer say with confidence that currency x at time y was fixed, floating or whatever."* Following on from this, there are still today, only sparse —frankly, hardly any— certainties about the macroeconomic consequences of ERR choices. The current argument that the different *de facto* ERR classifications measure different things and are therefore useful in different contexts is clearly unsatisfactory since crucial questions remain open. Is ERR flexibility a better shield against crises? Does it foster growth? Does it facilitate macroeconomic adjustments?[3] ... While the theoretical literature pullulates with incriminating and exculpatory evidence for each regime, the empirical literature is still missing a "tuned instrument".

Against this background, we introduce an original approach to determine ERR by providing a comprehensive and unified framework. Precisely, whether or not a country must be classified under an ERR category is assessed through a probabilistic analysis consisting in evaluating probabilities of disagreement *vis-à-vis* the alternative categories —postulated by the different classifications— in a unified framework. Using model selection criteria, one

[1] Mexico in 1994, Thailand, Indonesia, and Korea in 1997, Russia in 1998, Brazil in 1999, Turkey in 2000-2001, and Argentina in 2001-2002.

[2] See Tavlas et al. (2008) for a survey.

[3] See Ghosh et al. (2019) for a recent controversy on this issue.

can select the most probable ERR category for each controversial observation, that is, the ERR category *vis-à-vis* which the probability of disagreement is the lowest. By combining the information contained in alternative classifications, such a unified framework can highlight patterns of exchange rate behavior that would otherwise remain undetected. We provide preliminary evidence on the usefulness of this synthesis classification as an input in the design of exchange rate regimes and by examining the extent to which it performs better than traditional classifications in detecting currency crisis.

In this paper, we thus combine alternative classifications to exploit their complementarity with the idea that none of these classifications is better than another one. We propose an original methodology to evaluate the disagreements among the classifications and derive a shared conception of ERR categories as a common reference frame from this evaluation. We focus on two *de facto* classifications: the Reinhart and Rogoff (2004, thereafter *RR*) and the Levy-Yeyati and Sturzenegger (2005, thereafter *LYS*) classifications. We have several reasons for focusing on this pair. First, owing to their country/year coverage, these two classifications are the most used in the literature and are furthermore freely and publicly available. Second, they are also the most discordant classifications from a methodological point of view, so that they indirectly cover the whole spectrum of the existing *de facto* classifications.

Several results emerge from our analyses. First, regarding the joint examination of the *RR* and *LYS* classifications, we evidence that around 40% of the observations —where each observation corresponds to a given country's regime in a particular year— are not directly comparable even when there is an agreement on the ERR category. A direct implication is that many empirical studies using both classifications for robustness purposes may be plagued by artifacts. Second, we show that relatively few disagreements (around one-fourth) are directly attributable to specific variables well-identified. Instead, most of them originate from the different thresholds used by the classifications in the definition of the ERR categories and the interactions between several variables. These complexities further advocate the need for a consistent and unified framework as brought by the synthesis classification. With this latter classification, we propose a historical reinterpretation of ERR developments around the globe. Among other findings, the synthesis classification provides a more nuanced picture of the so-called bipolar view. Indeed, we show that the evolution of *de facto* regimes —especially in emerging economies since the late 1990s— has not involved a shift away from softly pegged exchange rate regimes toward floating rates and hard pegs but instead a movement toward more tightly "managed" intermediate regimes. Finally, our synthesis classification, by displaying the highest agreement rates with the most popular classifications, appears to convey more information and thus allows for greater reliability. An empirical exercise on the relationship between ERR and currency crises supports this feature and associates exchange rate flexibility with a higher and statistically significant risk of currency crisis —for a comparable macroeconomic environment.

This paper therefore contributes to the literature along several dimensions. First, we develop an original methodological framework that allows the empirical testing and identification of a rich set of sources of disagreement between the *LYS* and *RR* classifications. As such, we offer some explanations for the lack of clear-cut results on the choices and consequences of ERR. Second, we provide the first *de facto* classification based on a synthesis of ERR classifications. From a methodological point of view, we open the way to extend this framework to a more general setting. Finally, our paper opens up several perspectives for future empirical works on the consequences associated to ERR choices.

The paper proceeds as follows. Section 2 presents the background and describes the methodology used. Section 3 provides an historical landscape of exchange rate regimes deriving from the synthesis classification. Section 4 analyzes the properties and the empirical scope of the synthesis classification. Finally, Section 5 concludes.

2. Background and methodology

There have been many attempts to adopt a holistic approach to classify exchange rate regimes. If progress has been made, significant problems remain due to the complex nature of ERRs so that this issue still remains one of the most challenging research program in international economics.

As noted by Tavlas et al. (2008), more than a dozen *de facto* classifications have been proposed —e.g., Bubula and Ötker-Robe (2002), Ghosh et al. (2003), Bailliu, Lafrance and Perrault (2003), Reinhart and Rogoff (2004), Shambaugh (2004), Levy-Yeyati and Sturzenegger (2005), Bénassy-Quéré et al. (2006) and Obstfeld, Shambaugh, and Taylor (2010). Since 1999, the IMF has also adopted a *de facto* classification (IMF, 1999).[4] While these *de facto* classifications have in common to classify exchange rate regimes using a system based on actual behavior, they disagree (often considerably) regarding the countries' exchange rate regimes (see Table 1).

The low levels of agreement are, in fact, not very surprising. By drawing on different conceptions of ERR, the classifications capture various aspects of exchange rate regimes. As aforementioned, this represents the Achilles heel of the empirical literature on ERR marked by a lack of robustness. Surprisingly, however, the profession has not taken the step forward, that is, exploiting the complementarity between the different ERR classifications

[4] Up to 2008, the IMF *de facto* classification coincided with the Bubula, Ötker-Robe and Anderson classification (BORA; see Anderson, 2012).

Table 1 — Agreements between the *de facto* ERR classifications

	IMF	BORA	LYS	OST	RR
IMF (*de jure*)	100%				
Bubula, Ötker-Robe & Anderson (BORA)	81.63% (5089)	100%			
Levy-Yeyati & Sturzenegger (LYS)	44.95% (3766)	46.21% (3367)	100%		
Obstfeld, Shambaugh & Taylor (OST)	47.04% (5153)	46.33% (4675)	65.79% (4779)	100%	
Reinhart & Rogoff (RR)	46.95% (3766)	55.39% (3367)	57.69% (5011)	71.37% (4779)	100%

Notes: The entries correspond to the percentages of observations on which the classifications agree. The total number of observations used for the pairwise comparisons is reported in parentheses. We dropped all the observations prior to 1973. The IMF *de jure* and the BORA classifications cover the 1973-2006 period. All the classifications are composed of three categories: fixed, intermediate, and floating.

to derive a richer conception of ERR categories. However, reaching this objective implies addressing some related issues, among which, that of the classifications to be used.

2.1. Selecting the classifications

Among the different existing *de facto* classifications, we select the classifications developed by Reinhart and Rogoff (2004; thereafter RR) —also known as IRR since its extension (see Ilzetzki et al., 2017)— and Levy-Yeyati and Sturzenegger (2005; thereafter LYS). These classifications are the most comprehensive ones and form the most dissimilar pair from a methodological point of view. This latter feature thus allows us to indirectly cover the whole spectrum of the existing *de facto* classifications.

Specifically, the two classifications differ considerably regarding: (i) the data, (ii) the key statistic(s), and (iii) the methodology they use for categorizing the different ERR. The *LYS* classification combines available information on the official exchange rate and reserves' movements to capture the effect of interventions on the exchange rate and determine the *de facto* flexibility of ERR. On the methodology side, it builds on a cluster analysis which partitions data points (a data point corresponding to a given country's currency x at particular time t) into different ERR categories according to their similarity across the following variables: (i) changes in the nominal exchange rate —measured as the average of the absolute monthly percentage changes in the nominal exchange rate during a calendar year, (ii) the volatility of these changes —computed as the standard deviation of the monthly percentage changes in the exchange rate, and (iii) the volatility of the net-reserves-to-the monetary base ratio. The principle underlying this clustering is that countries experiencing low volatility of their exchange rates (both in levels and changes) and high volatility of their reserves should be classified under a *Fixed* ERR. Floaters should be associated with highly volatile exchange rates (both in changes and levels) and stable reserves. By definition, intermediate regimes fall between these two extreme regimes. The *RR* classification is based exclusively on exchange rate variations (the parallel market exchange rates when available), calculated as the absolute percent changes in the monthly nominal exchange rate averaged over a five-year rolling window —two-year in some cases.[5]

Due to their different frameworks, the two classifications exhibit significant divergences in the categorization of exchange rate regimes. To illustrate this, we collapse the *RR* and *LYS* classifications into three categories to fit the traditional three-way classification. Following the literature, we aggregate the different ERR categories of the *RR* classification as follows. The *Fixed* ERR comprises the categories 1 to 4 (fine classification), the *Intermediate* ERR includes categories 5 to 11, and the *Floating* ERR consists of the remaining categories.[6] The *LYS* classification differentiates only four categories of regimes (plus one associated with inconclusive determinations) that can be converted into the usual tripartite categorization by grouping dirty float and dirty float/crawling pegs into the *Intermediate* RR. Table 2 presents the two-way contingency table between the *RR* and *LYS* classifications. The observed rate of agreement between the *RR* and *LYS* classifications reaches 57.7%.[7] On average, the agreement between the two classifications is the highest for the *Fixed* regime category, followed by the *Intermediate* category.

Table 2 — Two-way contingency table

		LYS			
		Fix.	*Inter.*	*Float*	Total
RR	*Fix..*	**2080**	187	289	2556
	Inter.	481	**497**	888	1866
	Float	151	124	**314**	589
	Total	2712	808	1491	5011
Pearson $\chi^2(4) = 1.6e+03 \mid Pr = 0.000$					

Note: Pearson $\chi^2(.)$ displays the statistics and p.value associated to the independence test of rows and columns –in a two-way table.

[5]The *RR* classification also considers the inflation rate in its procedure to differentiate the "freely falling" category —composed of countries whose twelve-month inflation rate is above 40%— from the others.

[6]Following the literature, we exclude the "freely falling" category from the empirical analysis. This omission represents a loss of 397 observations. Furthermore, note that the "separating line" between the *Intermediate* and the *Floating* ERRs is itself a source of disagreements. The selected "line" maximizes the concordance (a gain of 89 points) between the two classifications —and is in line with the literature.

[7]The agreement rate corresponds to the sum of observations along the diagonal divided by the total number of observations.

2.2. Deriving the synthesis classification

We propose in this paper a synthesis classification aiming at exploiting the complementarity between the *LYS* and *RR* classifications to derive a richer conception of ERR categories. This synthesis classification first draws on the upstream examination of the divergences' causes between the *LYS* and *RR* classifications. The rationale behind this first analysis is simple. To obtain a clear picture of the ERR followed by a country in a particular year, we have to integrate the *LYS* and *RR* classifications into a unified framework. Consequently, we first need to understand the reasons behind their disagreements in order to exploit them and extract the common conception of ERR categories. To save space, the methodology underlying this exercise and the findings are reported in Appendix B.1. In substance, we derive the synthesis classification by inferring, for a given country in a particular year t, the closest ERR category, i.e., the ERR category *vis-à-vis* which the probability of disagreement is the lowest once unified the *LYS* and *RR* ERR conceptions. Operationally, we proceed by estimating different probit models and predict the associated probabilities, each indicating the likelihood —in the unified framework— of the ERR category suggested by each of the two classifications. Our dependent variable scores 1 in case of divergences between the *LYS* and *RR* classifications; 0 otherwise. The set of explanatory variables include both variables used by the classifications as well as other determinants.[8]

It is worthwhile noting that we depart, in this first exercise, from the overlapping sample between the *LYS* and *RR* classifications. Indeed, several observations recorded by the two classifications derive from judgmental decisions: *(i)* observations not classified by the clustering algorithm and labelled as "Uncontroversial" and "Fixed inconclusive" in the *LYS* classification, and/or *(ii)* observations with a difference in the reference currency between the two classifications. These "conditional" observations blur the perimeters of the different ERR categories defined by the two classifications so that it is impossible to derive a shared conception of regime categories.[9] Therefore, we remove in a first step these observations to ensure that the classifications are directly comparable and that such a comparison can

[8]Note that we also conduct an exploratory analysis of the disagreement observations, which highlights the set of explanatory variables retained. See the online appendix.

[9]As discussed in the online appendix, this is namely the case of the euro area countries. Indeed, despite the lack of volatility *vis-à-vis* the reference currency —i.e., the Deutsche mark, the *LYS* classification classifies the ERR of euro area countries under the *Float* category from 1999 onwards (except in 2008 they enter into the *Intermediate* category). Levy-Yeyati and Sturzenegger (2016) deliberately choose to classify the Eurozone member countries as *Float* —given the euro's behavior *vis-à-vis* other currencies. But, they acknowledge that the ERR's classification for the euro area countries (i.e. *Fixed* or Float) remains an open question and that the answer depends on the issue at stake. Note that this treatment was only applied to the eurozone, leaving the ERR categorization of other monetary union member countries at the mercy of their algorithm.

be performed within a consistent framework. These "conditional" observations represent 38.8% of the initial observations (1945 among the initial 5011 observations). The remaining observations (3066) thus constitute our working sample for which we infer, for each observation, the ERR in the unified framework, i.e. the core synthesis classification —using the above methodology.

In a second step, we extend the core synthesis classification to include additional observations, especially the aforementioned conditional observations. Indeed, from a practical point of view, the relatively low coverage of the core synthesis classification seriously weakens its potential benefits. To considerably gain in scope, we need to extend the core *SC*, i.e., filling the gaps and updating it to recent years. For brevity, we present in Appendix B.2 an extensive discussion on these extensions —issues and methodology. In essence, we rely on a discriminant analysis, the k^{th}-nearest-neighbor (KNN) algorithm. This method allows us to assign to each unclassified observation the most suitable regime category of the core *SC*, e.g., the regime category that includes the most similar observations to the data point to be classified. In addition to the observations initially left apart, we also cover some "blind spots" in the *LYS* and *RR* classifications. The final classification, the extended synthesis classification, thus also enriches the core *SC* with observations not "classified" in the *LYS* classification (i.e., "undisclosed baskets") and in the *RR* classification (i.e., "freely falling").[10]

3. The evolution of exchange rate regimes: evidence from the synthesis classification

Figure 1 compares the changes —since 1974— in the distribution of aggregated ERR categories (fixed, intermediate, and float) between the —extended— synthesis classification and the *LYS* and *RR* classifications.

As visible, the *synthesis classification* offers a discernible picture in the composition of ERR compared to the other classifications. Specifically, the number of intermediate regimes is significantly higher under the synthesis classification than the *LYS* classification. On the contrary, intermediate regimes are less prevalent under the synthesis classification than suggested by the *RR* classification. Differences across the three classifications also exist regarding shifts in the composition of ERR.

Among advanced economies, the synthesis and the *LYS* classifications suggest that floating currencies principally marked the breakup of the Bretton Woods system. In contrast, for the *RR* classification, intermediate regimes were more prevalent. However, the synthesis classification yields a similar trend to the *RR* classification towards an expanded

[10]The whole methodology underlying the synthesis classification is summarized in Figure B.1 in Appendix B.

fraction of fixed regimes in the 1980s and 1990s with the move of the euro area countries toward monetary union. In contrast, the *LYS* classification picks up less fixed regimes. This result is due mainly to the classification of euro area currency regimes as *Intermediate* in 2008 and *Float* otherwise in the *LYS* classification.[11]

Among emerging markets (EMEs), intermediate regimes have been and continue to be considerably more prevalent under the *RR* classification than suggested by the other classifications. 45% —on average— of the observations are recorded as *Intermediate* ERR, while in the synthesis classification (resp. LYS), the average share is equal to 33% (resp. 26%). Inversely, the *LYS* classification records many more floaters in EMEs than the other classifications. According to this classification, the proportion of floaters represented 25% of all exchange rate regimes in 1980 and has nearly doubled over the whole period. In contrast, under the *RR* classification, floaters almost disappeared among EMEs in the mid-1990s before gradually reappearing (14% on average from 2000 onwards). The *synthesis classification* provides a more balanced perspective on the evolutions of both the *Intermediate* and *Float* categories. Besides, it also indicates a lower proportion of fixers in EMEs from the mid-1980s to the mid-1990s.

[11] Suppose that the *LYS* classification classifies euro area countries under the *Fixed* category as the synthesis and *RR* classifications. In that case, the *LYS* classification would have recorded 66.7% of observations under the *Fixed* category—against 55.7% for the synthesis classification (see the dashed red line in the figure indicating the *Fixed* category's share under this alternative (re)classification).

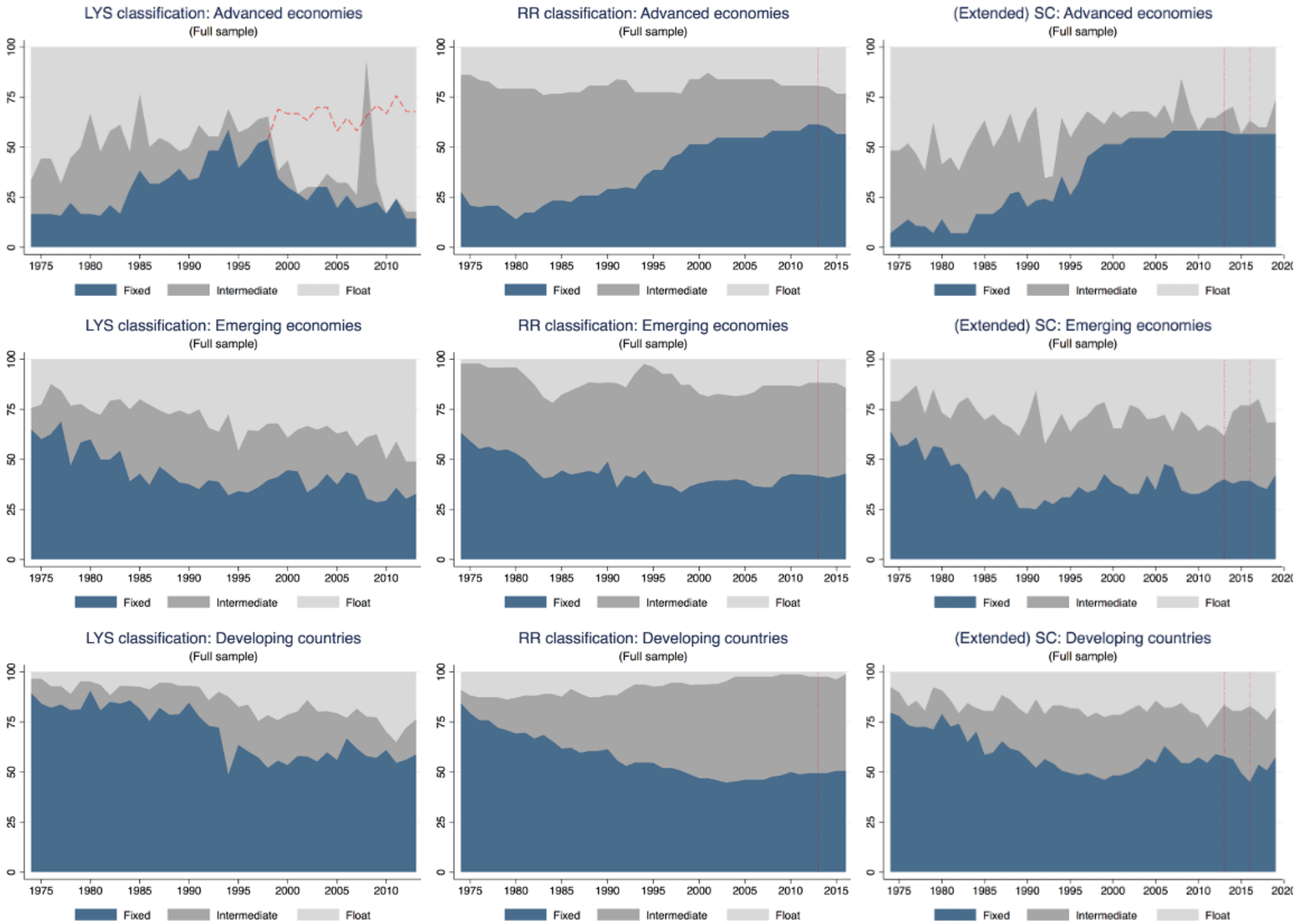

Figure 1 — ERR category distributions over time (by classification; % of annual observations; full samples)

Notes: The dotted (resp. dashed) vertical line indicates the end of the *LYS* (resp. *RR*) classification coverage, i.e. the year 2013 (resp. 2016). The dashed red line (LYS, advanced economies) indicates the *Fixed* category's share under the reclassification of euro area countries within this category.

The distribution across ERR categories among developing countries (DCs) also shows a shift of both the *Fixed* and *Float* categories towards the *Intermediate* category under the *RR* classification. The *Intermediate* category's share in the *RR* classification has increased from 6.5% in 1974 to 48% in 2016, mainly fueled by the steady decline in the percentage of the *Fixed* category (from 84% in 1974 to 50.6% in 2016). The *LYS* classification again describes an expansion of the *Float* category contrary to the trend suggested by the *RR* classification. In contrast, the synthesis classification provides a more balanced view but with specific changes. Finally, the predominance of the *Fixed* ERR is noticeable in all classifications. However, in contrast with the other classifications displaying a continued decline in the *Fixed* category —at least before the 2000s, the *LYS* classification records a major regime transition before and after the 1994 Franc CFA devaluation. This reflects the tendency of this classification in recording short-term currency market pressures as regime changes. Following the CFA franc devaluation in early 1994, this classification assigns to the CFA franc zone countries an *Intermediate* regime. Consequently, devaluation episodes further exacerbate the difference between the *LYS* classification and the others.

Overall, the increase of *Floaters* against the *Intermediate* category among EMEs and DCs since the 1990s suggested by the *LYS* classifications seems to support the bipolar view. In contrast, the picture given by the *RR* classification does not show a move to the polar extremes of exchange rate flexibility: the *Intermediate* category has remained significant in EMEs and has even increased in DCs. In contrast, the *synthesis classification* suggests an evolution in the composition of exchange rate regimes more stable than that recorded by the *RR* and *LYS* classifications. This apparent stability also questions the general validity of the bipolar view.

Figure 2 depicts the evolution of regime categories derived from a more detailed classification to gain additional insights on this issue. According to thresholds reached by changes in the nominal exchange rate and the volatility of these changes (Table 3), we split the *Fixed* ERR into four sub-categories and the *Intermediate* ERR into five sub-categories, the *Floating* ERR remaining unchanged.[12]

[12]We do not distinguish categories within the *Floating* ERR because it is a perilous exercise. Indeed, given the plurality of the intervention means and the scarcity of the data to control, the distinction often made between free float and managed float does not refer to freely floaters and "floaters" that intervene actively or frequently on the foreign exchange market. Instead, "managed" refers to the fact that for whatever reason —e.g., a random lack of volatility— the exchange rate variability index does not behave like the indices for the freely floaters (see Reinhart and Rogoff, 2004, p.46). Note further that we rely on a descriptive analysis to select the thresholds (see Figure B.2.3.2.1 in Appendix B.2).

Table 3 — The synthesis classification: coarse and fine grids

Regime	Coarse grid	Type	Exchange rate volatility (ΔE)	Volatility of exchange rate changes ($\sigma_{\Delta E}$)	Fine grid
Fixed	1	1	$\Delta E = 0$	$\sigma_{\Delta E} = 0$	1
		2	$\Delta E < 1\%$	$\sigma_{\Delta E} > 0$	2
		3	$1\% \leq \Delta E < 2\%$	$\sigma_{\Delta E} > 0$	3
		4	$\Delta E \geq 2\%$	$\sigma_{\Delta E} > 0$	4
Intermediate	2	1	$\Delta E < 2\%$	$\sigma_{\Delta E} < 2\%$	5
		2	$2\% \leq \Delta E < 5\%$	$\sigma_{\Delta E} < 2\%$	6
		3	$\Delta E < 2\%$	$\sigma_{\Delta E} \geq 2\%$	7
		4	$2\% \leq \Delta E < 5\%$	$\sigma_{\Delta E} \geq 2\%$	8
		5	$\Delta E \geq 5\%$	$\sigma_{\Delta E} \geq 2\%$	9
Float	3		—		10

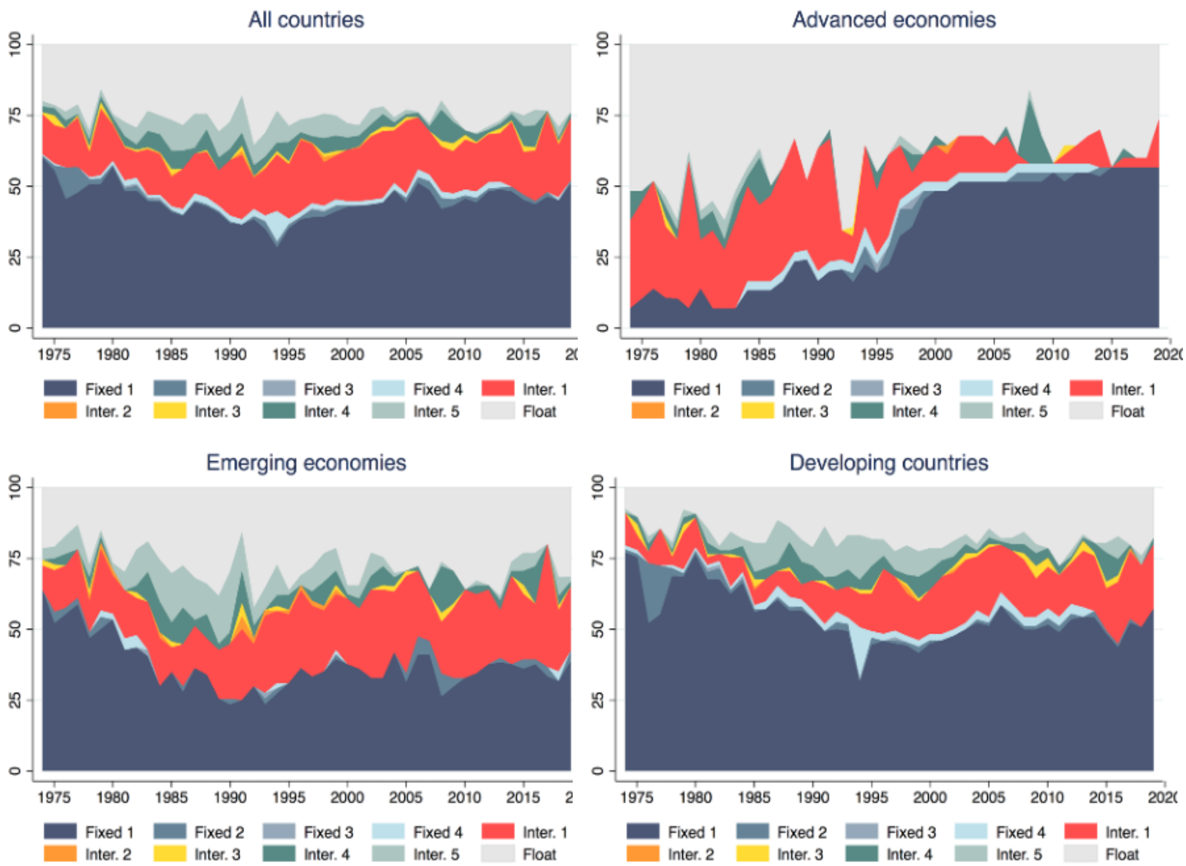

Figure 2 — The fine *SC*: evolutions of the categories (% of annual observations

Note: Full extended *SC* sample. The figures in front of the regime indicate the type.

Except for AEs that gradually moved towards the *Fixed* regime due to the monetary union in Europe, changes in countries' exchange rate regimes have been more prevalent within than between aggregated regime categories. Among EMEs, *Intermediate* regimes have recorded the most significant changes. Within this category, the number of more flexible arrangements have declined to the benefits of more rigid arrangements since the mid-1980s. These opposite trends indicate that EMEs have not hollowed out the intermediate regime. Instead, this regime has kept a rather stable share over time, as suggested by Figure 1.[13] More specifically, the evolution of *de facto* regimes —in EMEs since the late 1990s— has involved a movement toward more tightly "managed" intermediate regimes. This observation offers a new perspective on exchange rate management in EMEs during the 1980s and 1990s. It also provides the motivation for a new investigation of regimes' vulnerability to currency crisis.

4. Evaluating the scope of the synthesis classification

4.1. Properties of the synthesis classification

Overall, the extended synthesis classification consists of 7780 observations covering 184 countries over the 1974-2019 period. Compared to alternative classifications, it thus displays the broadest data coverage.

Table 4 reports the percentage of agreements between the synthesis classification and alternative *de facto* classifications. As visible, the *synthesis classification* displays a significantly higher agreement rate with the *LYS* and the *RR* classifications than the rate between the two classifications. The agreement rate reaches 80% *vis-à-vis* the *LYS* classification and between 66% and 68% *vis-à-vis* the *RR* classification, against an agreement rate of 57% between the two classifications.

Our classification also shows high rates of agreement with the IMF *de facto* classification (between 66.3% and 73.2%). This latter also shows relatively high proximity *vis-à-vis* the *RR* classification. Overall, these findings demonstrate that our methodology, consisting of combining the most discordant classifications into a unified framework, results in a classification that exhibits, on average, the highest agreement rates *vis-à-vis* alternative ones.

[13]The tightening of the *Intermediate* ERR is also noted for DCs —but with a steep gradient.

Table 4 — Agreements between the *de facto* ERR classifications

	IMF *de jure*	IMF *de facto*	LYS	RR	SC
Common sample					
IMF *de jure*	100% (2641)				
IMF *de facto*	57.7% (2641)	100% (2641)			
LYS	43.1% (2641)	69.3% (2641)	100% (2641)		
RR	46.7% (2641)	69.6% (2641)	59.2% (2641)	100% (2641)	
SC	44.9% (2641)	73.2% (2641)	87.4% (2641)	68.1% (2641)	100% (2641)
Pairwise overlapping sample					
IMF *de jure*	100% (4023)				
IMF *de facto*	60.4% (3959)	100% (7516)			
LYS	40.9% (3106)	60.1% (5413)	100% (5517)		
RR	51.1% (3410)	67.6% (6447)	57.7% (5011)	100% (7054)	
SC	42.4% (3915)	66.3% (7380)	83.6% (5494)	66.1% (6597)	100% (7780)

Notes: The entries correspond to the percentages of observations on which the classifications agree. The total number of observations used for the comparisons is reported in parentheses. We dropped observations before 1974. The IMF *de jure* (resp. *de facto*) classification(s) covers the 1974-1999 (1974-2018) period. We collapsed all the classifications into the usual three-way classification, i.e., *Fixed*, Intermediate, and *Float* (see Table A.2 in Appendix A for further details).

4.2. Exchange rate regimes and currency crises: what insights from the synthesis classification?

As a final step, we carry a brief empirical exercise to illustrate the insightfulness of the synthesis classification (SC). Specifically, we reexamine the relationship between exchange rate regimes and currency crises in emerging and developing countries, a long-standing debate on the vulnerabilities of the different exchange rate regimes that has not yet been closed despite numerous empirical studies. On the theoretical side, a consensus is far from being reached owing partly to the diversity of historical episodes that fueled different interpretations, thus making this issue an empirical one. However, on the strictly empirical side, the lack of consensus among classifications in categorizing exchange rate regimes leaves open the question of which ERR is more prone to crisis.

Figure 3 illustrates this point by showing the frequencies of currency crises associated with the three aggregated regimes for the IMF, LYS, RR, and synthesis classifications. For each classification, we present the frequency of crisis subdivided according to the exchange rate regime in place *(i)* the year of the crisis (left panel) and *(ii)* the year before the crisis (right panel). This latter subdivision is generally preferred in empirical exercises to mitigate endogeneity concerns, while the former serves for descriptive purposes. Different interesting observations can be made from Figure 3. Across all classifications, the *Fixed* and the *Intermediate* regimes appear more prevalent the year before a crisis. This pattern contrasts with the left panels presenting the *Intermediate* category as the regime recording significantly most crises —except in the IMF classification and, to a lesser extent, the *RR* classification. We observe similar charts in this latter classification, suggesting a higher degree of regimes' inertia. But, the striking feature of the *RR* classification is its number of unclassified observations—i.e., crisis episodes for which exchange rate regimes are "not classified".[14] With very few unclassified observations, the IMF and synthesis classifications provide a different picture, namely by associating a higher frequency to the *Floating* regime the year before a crisis —compared to the *LYS* and *RR* classifications. Overall, this visual inspection provides clues indicating that the *SC* could unveil previously non-visible aspects/information.[15]

[14]Indeed, the methodology underlying the *RR* classification does not allow to "really" characterize all the regimes, namely those of economies (i) with a high inflation rate (above 40%) like it is was the case for several Latin American countries from the 1970s to the 1990s, or (ii) switching from a *Fixed* to a *Floating* regime following a crisis. Observations meeting one of these criteria are grouped in the "Freely falling" category, which is not operational.

[15]One should interpret these outcomes with suitable caution rather than evidence of a causal relationship between the ERR and currency crises.

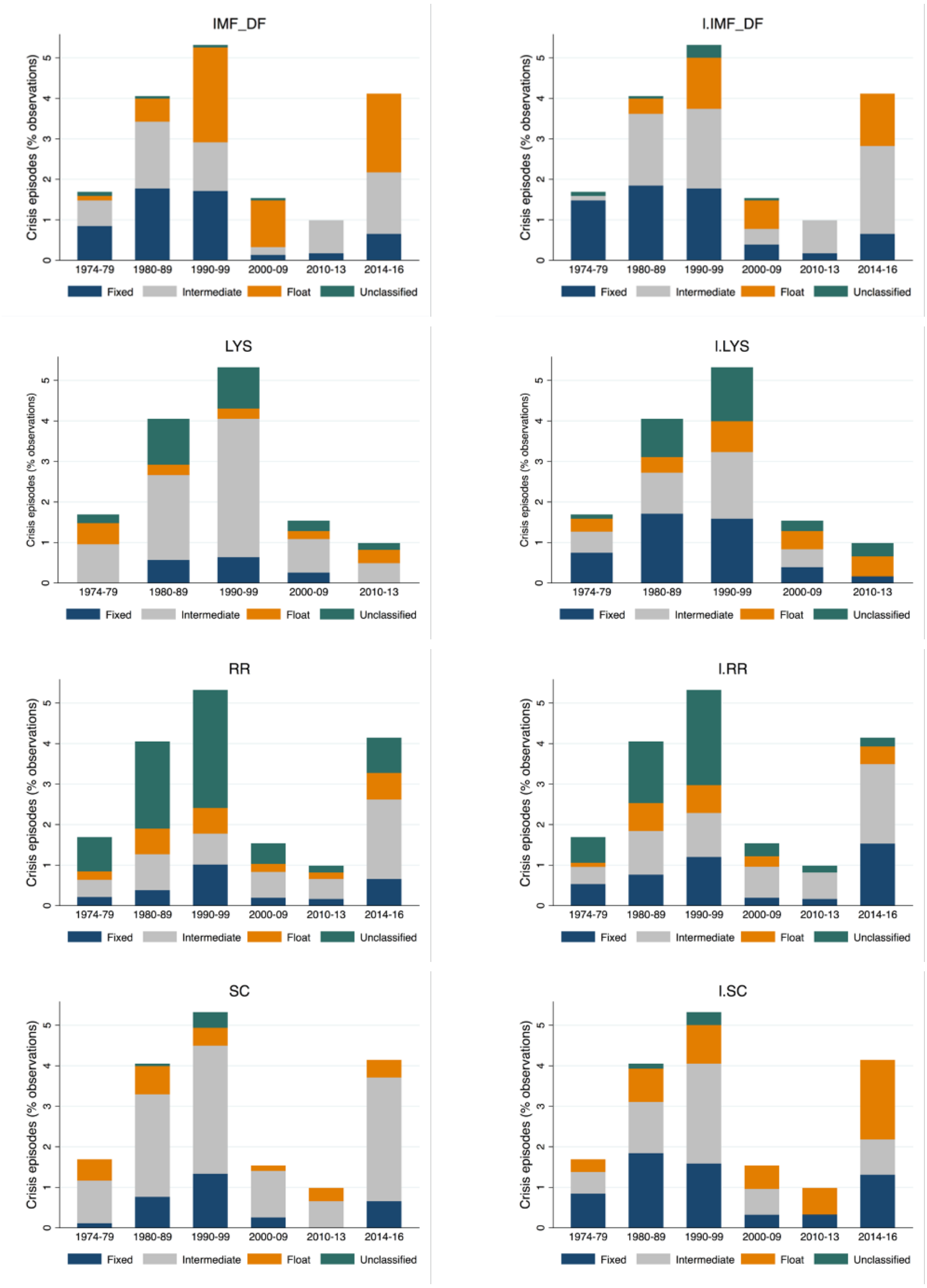

Figure 3 — Exchange rate regimes and currency crisis frequencies (DCs & EMEs)

Notes: The bars indicate the frequencies of currency crises, i.e. the number of crises in proportion of the observations (country-years). This frequency is subdivided according to the exchange rate regime in place (*i*) the year of the crisis (left chart) and (*ii*) the year before the crisis (right chart). "Unclassified" indicates crisis episodes for which the exchange rate regime is not determined. Advanced economies are excluded from the sample. A crisis is defined as a depreciation of the nominal exchange rate against the dollar of at least 30 percent and that is at least 10 percentage points greater than the depreciation in the previous year (see Laeven and Valencia, 2013). The crisis database does not cover the period after 2016.

We propose an in-depth analysis of the ERRs vulnerability by estimating probit models relying specifically on the framework proposed by Ghosh et al. (2015).[16] For brevity, results are reported in Appendix C. Figure 4 shows the main takeaways from the analysis. Indeed, it displays the estimated odds ratios obtained for each regime (relative to the *Fixed* category) by classification. As visible, the synthesis classification indicates that, once we control for the macroeconomic environment, the *Fixed* regime appears less vulnerable to currency crises. More specifically, estimates suggest that other things being equal, the risk of crisis in both *Intermediate* and *Floating* regimes is —on average— around three times higher compared to the *Fixed* regime. Hence, the synthesis classification associates flexibility with a higher risk of currency crisis occurrence. This result contrasts with those based on the other classifications —and so those in the empirical literature relying mainly on the *LYS* and *RR* classifications. Indeed, as can be seen, the *LYS* classification identifies the *Intermediate* regime as being more prone to crises than the *Fixed* category, while with the *RR* classification, it is rather the *Floating* regime that tends to be associated with a higher probability of currency crisis. However, for both *LYS* and *RR* classifications, the level of uncertainty is such that one cannot establish any ranking regarding the insulating properties of the different regimes —see the confidence intervals' overlap or the coefficients equality tests in Table C.1. The picture is also the same for the IMF *de facto* classification. This lack of robustness/clear-cut results is typical of most empirical studies relying on these classifications. The outcomes derived from the *SC* reflect the interest of such classification. It improves the analysis by increasing the spectrum of existing ERR and enhancing the way these ERR are assigned across countries and over time.[17] In all likelihood, the results derived using the *SC* appear as a —less noisy— superimposition of images of the different classifications. Say differently, the *SC* appears to combine and reflect both the *LYS* and *RR* classifications' specificities.

[16]The underlying assumption is that the probability of an exchange rate crisis occurring in a country is a function of its macroeconomic characteristics such as the exchange rate misalignment (i.e., the deviation of the exchange rate from its equilibrium level), the current account balance, the GDP growth rate, the inflation rate, and the real GDP per capita as well as that of the official reserves. The exchange rate regime is introduced as an additional variable to identify its effect once controlling for these other determinants—other things being equal, in sum. We retained the same set of regressors which have been one-year lagged to limit endogeneity concerns.

[17]As an illustrative example, once we restrict the sample to the common sample, the *SC* loses its global vision and is close to the *LYS* classification.

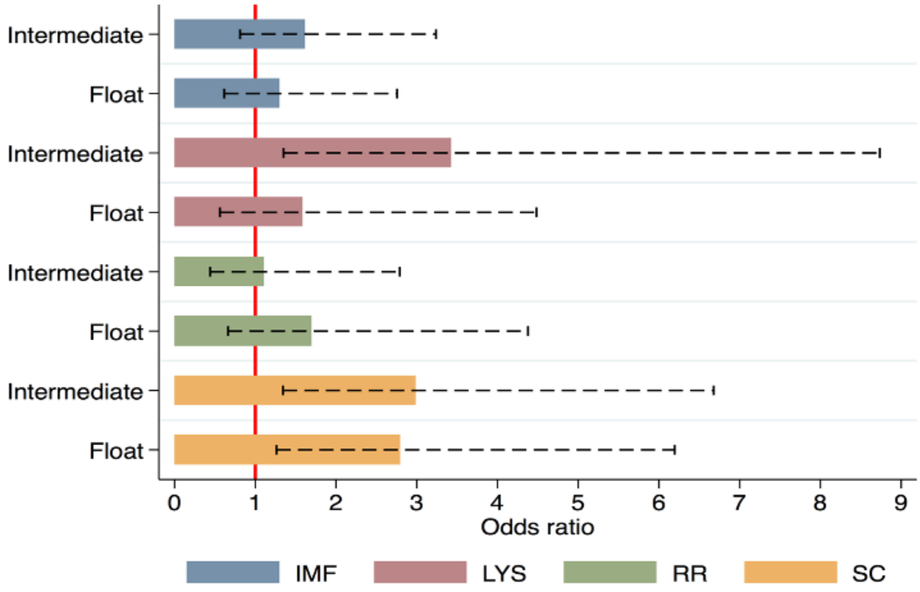

Figure 4 — Exchange rate regimes and currency crisis: odds ratios
Notes: The bars indicate the estimated odds ratios for the model in columns C.1.7 to C.1.10 in Table C.1. The dashed lines represent the 95% confidence intervals. An odds ratio not statistically different from 1 indicates a crisis risk identical to that of the *Fixed* regime.

The quality of the analysis provided by the *SC* can also be evaluated from the model's predictive power relying on ROC analyses. The bottom lines of Table C.1 present the share of crisis episodes rightly identified (true positives, also called *sensitivity* in the ROC terminology), the share of non-crisis episodes rightly identified (true negatives also called *specificity*), and the *AUROC* (Area Under the ROC curve) which gives the overall goodness of the predictions. Relying on these indicators, the *SC* appears to perform better than the other classifications. By exploring the complementary between the *RR* and *LYS* classification, the synthesis classification helps to improve the quality of the results, most notably in reducing the percentage of crises missed and increasing the percentage predicted correctly. This finding is explained by the improvement of the categorization of ERRs with the synthesis classification, particularly around crisis episodes, allowing a more accurate and comprehensive assessment of the susceptibility of ERRs to currency crises. The difference is particularly noticeable for countries with a turbulent history in terms of exchange rate policy. The Mexican stabilization experience during 1987-94 is an illustrative example of how the conclusions on the viability of ERRs may be influenced by their categorization. According to the synthesis classification, Mexico was under the intermediate category from 1989 to 1994, reflecting the policies pursued during the stabilization program. Under this program, the peso evolved according to a crawling peg from 1989 to 1991 and then moved to a limited flexibility regime (bands) until 1994. In contrast, the *RR* (1992 and 1993) and the *LYS* (1991 and 1994) classifications classify the Mexican regime as *Fixed*. As a result, they capture with some delay the stabilization program of the peso —against the US dollar— and associate it with a *Fixed* ERR. This feature explains why the peso crisis (that ended up with a devaluation) is linked to the *Intermediate* regime using the synthesis classification and wrongly correlated with the *Fixed* category using the *LYS* classification. Another notable illustration is the exchange rate policy followed by Thailand before the Asian crisis. Under the synthesis classification, the baht moved from a rigid peg towards a soft peg (currency basket peg with a preponderance of the US dollar) from 1986 to 1998. According to the *RR* classification, Thailand has moved away from its rigid peg following the 1997 crisis, hence mistakenly associating the crisis to the *Fixed* category. Such imprecisions in the characterization of the ERR on the eve of crisis episodes inevitably distort the analysis of the causality linking currency crises to exchange rate regimes.

Overall, the proposed enhanced *de facto* classification thus allows refining the analysis of ERR vulnerabilities. As evidenced above, for a comparable macroeconomic environment, we found that currency crises are less frequent in *Fixed* ERR. It should, however be noted, that our analysis, altogether, basic even if it meets the current standards, does not claim to exhaust this complex subject but rather to set out the empirical framework based on a broader coverage.

5. Conclusion

In this paper, we develop a classification aiming at synthesizing the different ERR conceptions of the *LYS* and *RR* classifications. This synthesis classification provides, on average, the highest agreement rates among the most popular *de facto* classifications, which means that it not only conveys more information but also allows for greater comprehensiveness than the existing classifications. As such, the synthesis classification constitutes an essential contribution to the literature. Furthermore, along with the derivation of this synthesis classification, we provide explanations for the disagreements between the *LYS* and *RR* classifications. Besides the evidenced complex nature of the disagreements calling for a unified framework, i.e., a synthesis classification, we underline that around 40% of the observations amount to arbitrary decisions hence exposing studies to artifacts.

Relying on our classification, we also propose a reinterpretation of the ERR history in the post-Bretton Woods era. Among other takeaways, we bring an interesting nuance to the so-called "hollowing-out" hypothesis or bipolar view since the synthesis classification indicates that the evolution of ERR —especially in EMEs since the late 1990s, has been mainly a switch towards more tightly "managed" intermediate regimes. Furthermore, we perform an empirical exercise on the relationship between ERR and currency crises to highlight the interest of our classification. Our results indicate that exchange rate flexibility is associated with a higher risk of currency crisis —in EMEs and DCs. Estimates suggest an odds ratio of around 3 for both the *Intermediate* and *Float* categories. Besides, the synthesis classification displays better statistical performances than the other classifications.

All in all, we believe that our contributions may pave the way for future empirical works. Specifically, the synthesis classification can ultimately contribute to a better understanding of the differences between *de facto* classifications and shed new light on the determinants and consequences of exchange rate regimes. Likewise, more upstream, advances towards the enrichment of the empirical frameworks surrounding the identification of the ERR effects would definitely constitute an area for future research more than necessary.

References

Anderson, H. (2012), "Fear of Floating Revisited", available at SSRN: https://ssrn.com/abstract=2178767

Amemiya T. (1981), "Qualitative Response Models: A Survey", *Journal of Economic Literature* 19(4): 481-536.

Bailliu, J., Lafrance, R., Perrault, J.F. (2003), "Does Exchange Rate Policy Matter for Growth?", *International Finance* 6(3): 381-414.

Bénassy-Quéré, A., Coeuré, B., Mignon, V. (2006), "On the Identification of De Facto Currency Pegs", *Journal of the Japanese and International Economies* 20(1): 112-27.

Bleaney, M., Tian, M. (2014), "Classifying exchange rate regimes by regression methods", University of Nottingham School of Economics Discussion Paper no. 14/02.

Bleaney, M., Tian, M., Yin, L. (2015), "De facto exchange rate regime classifications are better than you think", University of Nottingham School of Economics Discussion Paper no. 15/01.

Bubula, A., Ötker-Robe, I. (2002), "The Evolution of Exchange Rate Regimes Since 1990: Evidence From De Facto Policies", IMF Working Paper, no. 02/155.

Calvo, G., Reinhart, C. (2002), "Fear of Floating", *The Quarterly Journal of Economics* 117(2): 379-408.

Eichengreen B., Razo-Garcia R. (2013), "How Reliable Are De Facto Exchange Rate Regime Classifications?", *International Journal of Finance & Economics* 18(3): 216-239.

Ghosh, A., Gulde, A. M., Wolf, H. (2003), "*Exchange Rate Regimes: Choices and Consequences*", Cambridge, M.A.: MIT Press.

Ghosh, A., Ostry, J., Qureshi, M. (2015), "Exchange Rate Management and Crisis Susceptibility: A Reassessment", *IMF Economic Review* 63(1): 238-76.

Ghosh, A., Qureshi, M., Tsangarides, C. (2019), "Friedman Redux: External Adjustment and Exchange Rate Flexibility", *Economic Journal* 129(617): 408-438.

Habermeier, K., Kokeyne, A., Veyrune, R., Anderson, H. (2009), "Revised System for the Classification of Exchange Rate Arrangement", IMF Working Paper 09/211.

Ilzetzki E., Reinhart C., Rogoff. K. (2017), "The Country Chronologies to Exchange Rate Arrangements into the 21st Century: Will the Anchor Currency Hold?" NBER Working Paper No. 23135.

IMF (1999), "Exchange Rate Arrangements and Currency Convertibility: Development and Issues", World Economic and Financial Surveys, Washington: International Monetary Fund.

Kahneman D., Tversky A. (1974), "Judgment Under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking under Uncertainty", *Science* 185(4157):1121-1131.

Laeven, L., Valencia, F. (2013), "Systemic Banking Crises", *IMF Economic Review* 61(2): 225-70.

Levy-Yeyati E., Sturzenegger F. (2005), "Classifying Exchange Rate Regimes: Deeds vs. Words", *European Economic Review* 49(6), 1603–35.

Levy-Yeyati, E., Sturzenegger, F., Gluzmann, P. (2013), "Fear of appreciation", *Journal of Development Economics* 101: 233-247.

Levy-Yeyati E., Sturzenegger F. (2016), "Classifying Exchange Rate Regimes: 15 Years Later", CID Faculty Working Paper No. 319, Center for International Development (Harvard University).

Obstfeld, M., Shambaugh, J., Taylor, A. (2010), "Financial Stability, the Trilemma, and International Reserves", *American Economic Association Journal – Macroeconomics* 2(2): 57-94.

Plous S. (1993), "*The Psychology of Judgment and Decision Making*", (New York: McGraw Hill).

Reinhart C., Rogoff K. (2004), "The Modern History of Exchange Rate Arrangements: A Reinterpretation", *The Quarterly Journal of Economics* 119(1): 1-48.

Rencher, A.C., W.F., Christensen. (2012), "*Methods of Multivariate Analysis*", 3rd ed. Hoboken, NJ: Wiley.

Shambaugh, J. (2004), "The Effect of Fixed Exchange Rates on Monetary Policy", *The Quarterly Journal of Economics* 119(1): 301-52.

Tavlas, G., Dellas H. Stockman, A.C. (2008), "The classification and performance of alternative exchange-rate systems, *European Economic Review* 52: 941-963.

Appendices

A. Data

Table A.1 — Country list

Advanced economies (AEs):

Australia, Austria, Belgium, Canada, Cyprus, Denmark, Finland, France, Germany, Greece, Hong Kong, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Macao, Malta, Netherlands, New Zealand, Norway, Portugal, Singapore, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States.

Emerging economies (EMEs):

Algeria, Antigua & Barbuda, Argentina, Aruba, Belarus, Bosnia & Herzegovina, Brazil, Brunei Darussalam, Bulgaria, Chile, China, Colombia, Costa Rica, Croatia, Czech Rep., Dominican Rep., Ecuador, Egypt, El Salvador, Equatorial Guinea, Estonia, Fiji, Guatemala, Hungary, India, Indonesia, Iran, Iraq, Jamaica, Jordan, Kazakhstan, Kuwait, Latvia, Lebanon, Lithuania, Macedonia (FYR), Malaysia, Marshall Islands, Mexico, Morocco, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Qatar, Romania, Russia, Saudi Arabia, Serbia, Slovakia, South Africa, Sri Lanka, Thailand, Tunisia, Turkey, United Arab Emirates, Ukraine, Uruguay, Venezuela.

Developing countries (DCs):

Afghanistan, Albania, Angola, Armenia, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belize, Benin, Bhutan, Bolivia, Botswana, Burkina Faso, Burundi, Cabo Verde, Cambodia, Cameroon, Central African Rep., Chad, Comoros, Congo, Congo D.R., Côte d'Ivoire, Djibouti, Dominica, Eritrea, Ethiopia, Gabon, Gambia, Georgia, Ghana, Grenada, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Kenya, Kiribati, Kyrgyzstan, Lao P.D.R., Lesotho, Liberia, Libya, Madagascar, Malawi, Maldives, Mali, Mauritania, Mauritius, Micronesia, Moldova, Mongolia, Montenegro, Mozambique, Myanmar, Namibia, Nepal, Nicaragua, Niger, Nigeria, Oman, Papua New Guinea, Palau, Rwanda, Sao Tome & Principe, Senegal, Seychelles, Sierra Leone, Solomon Islands, St. Kitts & Nevis, St. Lucia, St. Vincent & Grenadines, Sudan, Suriname, Swaziland, Syria, Tajikistan, Tanzania, Togo, Tonga, Trinidad & Tobago, Uganda, Vanuatu, Vietnam, Yemen, Zambia, Zimbabwe.

Note : Country groups are based on the IMF categorization. http://www.ieo-imf.org/ieo/files/completedevaluations/L.%20Annex%201.%20Country%20Group%20Profiles.pdf

Table A.2 — Data description and sources

Variable		Source
Exchange rate		
Official bilateral nominal exchange rate		IFS (IMF)
Market-determined parallel exchange rate		Carmen Reinhart website
Exchange rate regime classifications		
IMF de facto [a]		
Until 2008	**After 2008**	
	1. No separate legal tender $^{Fix.}$	
1. No separate legal tender $^{Fix.}$	2. Currency board $^{Fix.}$	
2. Currency board $^{Fix.}$	3. Conventional peg (Single currency; basket) $^{Fix.}$	
3. Conventional peg (Single currency; basket) $^{Fix.}$	4. Stabilized arrangement $^{Fix.}$	Anderson (2012); IMF's AREAER (various issues)
4. Pegs within horizontal bands $^{Int.}$	5. Pegs within horizontal bands $^{Int.}$	
5. Crawling peg $^{Int.}$	6. Craw ling peg $^{Int.}$	
6. Crawling band $^{Int.}$	7. Crawl-like arrangement $^{Int.}$	
7. Managed floating $^{Flt.}$	8. Other managed arrangement $^{Int.}$	
8. Independently floating $^{Flt.}$	9. Floating $^{Flt.}$	
	10. Free floating $^{Flt.}$	
IMF de jure [b]		
1. Hard pegs		
1.1. No separate legal tender		
1.2. Currency board		
1.3. Monetary union		
2. Traditional pegs		
2.1. Single currency		Gosh et al. (2003); IMF's AREAER (various issues)
2.2. Basket peg		
3. Floats with rule-based interventions		
3.1. Cooperative regimes		
3.2. Crawling peg		
3.3. Target zones and bands		
4. Floats with discretionary intervention (Managed floating)		
5. Floats		
Levy-Yeyati & Sturzenegger		Levy-Yeyati and Sturzenegger (2016)
Reinhart & Rogoff		Ilzetzki *et al.* (2017) (Carmen Reinhart website)
Macroeconomic variables		
Currency crisis: Dummy variable (1 equals crisis; 0 otherwise)		Laeven and Valencia (2013)
Currency misalignments: In percent		EQCHANGE (CEPII)
Current account: in %GDP		WDI (WB) & WEO (IMF)
Inflation[c]: in percent		WDI (WB) & WEO (IMF)
Real GDP growth rate[c]: in percent		WDI (WB) & WEO (IMF)
Real GDP per capita: in logs		WDI (WB) & WEO (IMF)
Reserves: in %GDP		WDI (WB) & WEO (IMF)

Notes: "a": the subscripts indicate the aggregation. "Fix." (resp. "Int." and "Flt.") indicates that the category is recorded as *Fixed* (resp. *Intermediate* and *Float*) (see Habermeier et al., 2009).
"b": *Fixed* = 1 + 2; *Intermediate* = 3 + 4; *Float* = 5.
"c": Transformed as x/(100+x) if x≥0; and x/(100-x) if x<0.

Appendix B. Methodologies and results

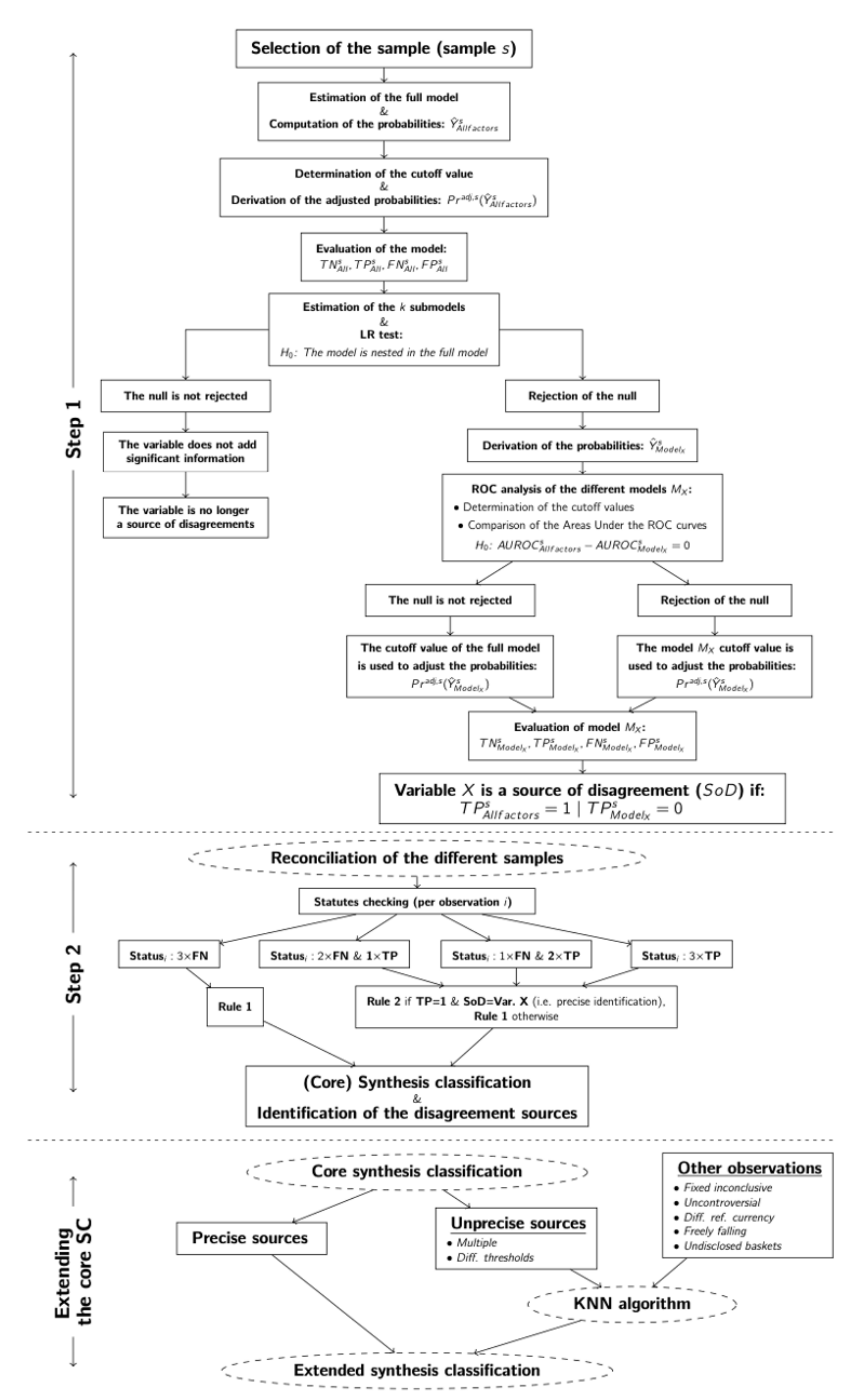

Figure A.2 — The methodology

Appendix B.1. Explaining the disagreements

B.1.1. The methodology

As mentioned before, it is often difficult to determine how the *RR* and *LYS* classifications relate to one another and for which application they are most suited. To obtain a clear picture of the exchange rate regime followed by a country in a particular year, we have to integrate these existing classifications into a unified framework. Consequently, we first need to understand the reasons behind the disagreements between the *RR* and *LYS* classifications to exploit them and extract the common conception of ERR categories. While one could advocate that we are "chasing two rabbits", it is worth noting that these likely two objectives are actually embedded. In reality, identifying the disagreement sources for each observed disagreement point —i.e., a given country i in a particular year t— raises the issue of a reference classification. Indeed, we cannot carry out such an exercise by implicitly considering, each time, either the *RR* or the *LYS* classification as the reference classification. Adopting an approach with two references will yield two potential sources of disagreement —one per classification as they do not share anything in common from a methodological point of view. We can illustrate this point with the following hypothetical example. Suppose a country A classified, in a particular year t, as *Fixed* by the *RR* classification but as *Float* by the *LYS* classification. If we consider the *RR* classification as the reference classification, the disagreement could be related, for instance, to the official reserves data used by the *LYS* classification. If, by contrast, we choose the *LYS* classification as the reference classification, the source of disagreement could originate from the *RR* classification that uses a longer time horizon over which the exchange rate volatility is measured and/or of information on parallel market exchange rates. Obviously, one would go around in circles, like a dog chasing its tail. Which of these sources should we retain? Capturing ERR under different prisms, none of the classifications should be privileged if one wants to get the common but also richest conception of the different ERR categories. The derivation of a synthesis classification, free of choosing a reference classification, allows overcoming the above problem. Formally, we derive the synthesis classification by inferring, for a given country in a particular year t, the closest ERR category, i.e., the ERR category *vis-à-vis* which the probability of disagreement is the lowest once unified the *LYS* and *RR* ERR conceptions. By integrating the classifications' distinctive characteristics into a consistent and complete framework, this synthesis classification will allow us to conclude not on an exclusive source of the disagreement between the two classifications —impossible given the existence of two reference frames—, but rather on the most important one.

We thus carry out our analysis in two steps. In the first step, we identify the disagreement source according to each classification. The results from this step are then used in the second step to derive the synthesis classification. Finally, we deduct the disagreement sources —that correspond to the ones identified in the sample coinciding with the synthesis classification.

Step 1: the sample-specific disagreement sources[18]

The first step of our general strategy consists in identifying the sources of disagreements according to each classification. The approach is relatively simple as it relies on stepwise disjunctive regressions to link each disagreement point to a specific source. More specifically, we compare the outcomes from a full model —including all candidate variables of disagreements (variables specific to each classification and reflecting their differences) as regressors— with the results from different k nested models —in which the k^{th} explanatory variable is dropped from the model. If eventually the removal of the variable k is associated with a fall in the sensitivity of the model —i.e. disagreement points are no longer detected in contrast with the full model, then the variable k is considered as the source of the disagreement. Hence, our backward stepwise approach appears as a parsimonious way of letting the data determine which variables are good (joint) predictors of disagreement between the two classifications.

As visible, our strategy rests on our ability to detect disagreement points. This, in turn, calls the issues of the estimation strategy and detection criteria. Regarding the estimation, we rely on probit models in which the dependent variable, Y, captures the disagreement between the two classifications and scores 0 in the absence of disagreements; 1 otherwise. As we are interested in identifying the sample-specific disagreement sources (i.e., the sources according to each classification), we consider several samples: a sample of *Fixed* ERR including observations recorded as *Fixed* at least by one classification, a sample of *Floating* ERR including observations reported as *Float* at least by one classification, and three samples of *Intermediate* ERR—(i) a *Lower-Intermediate* ERR sample composed of observations recorded as *Intermediate* at least by one classification and *Fixed* by the other one, *(ii)* an *Upper-Intermediate* ERR sample composed of observations recorded as *Intermediate* at least by one classification and *Float* by the other one, and *(iii)* a *Full Intermediate* ERR sample encompassing the two previous samples.[19] Each sample includes

[18]The term "sample-specific" corresponds to the different ERR categories to which a disagreement point belongs —that is the different classification viewpoints. Using again the country A example above, the *Fixed* (resp. *Float*) sample disagreement sources correspond to the sources identified considering the *RR* (resp. LYS) classification as the reference. This terminology is preferred as it suits well for the derivation of the synthesis classification (we provide further clarifications below).

[19]There are several important reasons for considering these different samples. Firstly, each of these samples

disagreement and agreement points, the latter constituting the reference group —see Figure B.1.1.1. We estimate all our probit models over all these samples and simulate each time the probabilities of disagreement.

		A	B	C
		LYS classification		
		Fixed	***Intermediate***	***Float***
1	***RR classif.*** ***Fixed***	**A1**	**B1**	**C1**
2	***Intermediate***	**A2**	**B2**	**C2**
3	***Float***	A3	**B3**	**C3**

Figure B.1.1.1 — Two-way contingency table and estimation samples

Note: The reading of the table is similar to that of the above contingency tables. The diagonal cells (A1 + B2 + C3) (resp. off-diagonal cells) correspond to the agreement (resp. disagreement) points between the two classifications. *Fixed* ERR sample = A1+A2+A3+B1+C1; *Lower-Intermediate* ERR sample = A2+B1+B2; *Upper-Intermediate* ERR sample = B2+B3+C2; *Float* ERR sample = C1+C2+C3+B3+A3.

Regarding the assessment of the model performances, i.e., their ability to detect disagreement points, we rely on Receiver Operating Characteristics (ROC) analyses. The ROC analyses provide us with ROC curves that are graphical plots illustrating the diagnostic ability. In the context of our analysis, a ROC curve plots the share of disagreements between the two classifications correctly identified by a given model (true positives "TP"; also called *sensitivity* in the ROC terminology) vs. the share of predicted disagreements not observed (false positives "FP"; "*1 - specificity*" in the terminology, *specificity* being

includes two alternative possibilities about the regime category, each referring to one classification. Secondly, considering a full sample (with $Y = 0$ for consensual points; 1 otherwise) would imply that the consensual points (consensual *Fixed*, Intermediate, and *Float*) are statistically identical regarding the variables and so form a homogenous group. Such an assumption can reasonably not be made. Thirdly, it is improbable that all the explanatory variables matter for all types of disagreements. For instance, the difference between the *Fixed* and the *Intermediate* ERR in the *LYS* classification involves only the exchange rate dynamics. On the contrary, the difference between the *Intermediate* and the *Float* ERR relies principally on reserves volatility. Hence, considering only one sample of all observations falling into the intermediate ERR could lead to biased coefficients and inaccurate simulated probabilities. Finally, the full *Intermediate* ERR sample makes it possible to assign disagreement points into the *Intermediate* regime. Indeed, one cannot exclude that "corner" observation (classified as *Fixed* in one classification and as *Float* in another one) corresponds neither to the *Fixed* category nor to the *Float* category, but instead to the *Intermediate* category of the synthesis classification. Further note that, for intermediate regimes, the distinction between downward and upward disagreements is crucial to facilitate the statistical discrimination of the observations. Indeed, while the *Fixed* and *Floating* ERR samples imply either lower or higher variability of the variables, facilitating one-way statistical discrimination of the observations, the *Intermediate* ERR full sample consists of a mix. Thus, the Lower (resp. Upper) *Intermediate* ERR sample consists of disagreements over the choice between the *Fixed* or *Intermediate* regime (resp. the *Intermediate* or *Float* regime).

the true-negative rate) along contiguous threshold settings.[20] We build a ROC function for each probit model and determine an optimal cut-off probability that corresponds to the highest true positive rate together with the lowest false positive rate.

Once the optimal cutoff value is selected, the probabilities of disagreement derived from each model are adjusted: probabilities higher than or equal to the cutoff value are considered equal to 1. In contrast, probabilities below the threshold are replaced by 0. This adjustment leads to four possibilities, as depicted in Table B.1.1.1.

Table B.1.1.1 — Observed disagreements and model outcomes

		Adjusted probabilities "$\hat{Y}$"	
		0 (agreements)	1 (disagreements)
Dep. variable "Y"	0 (actual agreements)	True negative "TN"	False Negative "FP"
	1 (actual disagreements)	False negative "FN"	True negative "TP"

Note: The cells in the table indicate the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), respectively. TPs and TNs are respectively disagreements and non-disagreements that are predicted correctly. FP is a predicted disagreement that does not occur and a FN is a predicted agreement that is actually a disagreement.

The identification strategy of the disagreement sources is based on these adjusted probabilities. It is worth noting that we simulate the disagreement probabilities for the different k submodels (the full model excluding the k^{th} variable) provided that the null of the likelihood ratio test is rejected. Also, before adjusting these probabilities, we compare the cut-off values of the submodels with that of the full model. If the difference is not significant, using the full model's cut-off value or the submodel's cut-off value leads to the same adjustment. However, if the difference is significant, we rely on the submodel's cut-off value. Finally, as stated above, we compare the changes in the number of disagreement points correctly identified (true positives). Hence, if when removing the explanatory variable k from the full model, we no longer detect a true positive, the variable k will be considered as the source of the disagreement —for the selected sample.

Step 2: derivation of the (core) synthesis classification and identification of the disagreement sources

The (core) synthesis classification aims to reclassify a disagreement point that has been classified, by definition, differently in the *RR* and *LYS* classifications. As a consequence, we only focus on disagreement points (predicted agreements that are actually disagree-

[20]Such an analysis is used in the *RR* classification to differentiate between the different types of pegs.

ments "FN" and disagreements that are correctly predicted "TP"), leaving unchanged the consensual points.

The (core) synthesis classification flows from the unified framework. We derive this unified framework by reconciling the information gathered from *Step 1*. Given that we estimate the probit model over different samples (which correspond to the confrontation of the classification viewpoints), analyses in *Step 1* yield at least two probabilities for each disagreement point: *(i)* the probability of disagreement *vis-à-vis* the ERR category i (when considering the sample of the ERR category i), and *(ii)* the probability of disagreement *vis-à-vis* the ERR category j (when considering the sample of the ERR category j). When the disagreement between the two classifications is related to a corner observation (i.e., *Fixed* in one classification and *Float* in the other), a third probability measuring the distance *vis-à-vis* the *Intermediate* category in the unified framework is also estimated.

In our view, the synthesis classification must combine both classifications' schemes into a unique and coherent framework. Therefore, a disagreement point should be classified in the most probable ERR category, that is, in the ERR category *vis-à-vis* which the probability of disagreement is the lowest.[21] By taking the previous example of country A's regime (classified in a particular year t as *Fixed* by the *RR* classification but *Float* by the *LYS* classification), the synthesis classification would consider country A's regime as *Fixed* if the probability of disagreement derived from the *Fixed* ERR sample is lower than that derived from the *Intermediate* and *Float* ERR samples. In other words, by applying such a rule (*Rule 1*), we consider that inferring to this disagreement point a *Fixed* ERR is "less false" than including it in another ERR category.

However, a disagreement cannot necessarily be detected in all estimation samples, as illustrated by the following situations: *(i)* the disagreement is not detected by the model in any of the estimation samples ("*3×FN*"), *(ii)* the disagreement is detected in only one of the estimation samples ("*2xFN & 1×TP*"), *(ii)* the disagreement is detected in two of the estimation samples ("*1×FN & 2×TP*"), and *(iv)* the disagreement is detected in all the estimation samples ("*3×TP*").[22]

When the disagreement point is not detected in any model —status "*2×FN*" or "*3×FN*" in the case of a corner observation, *Rule 1* shall apply. In the other situations, i.e., when the disagreement point is detected in at least one of the estimation samples, we introduce a refinement to *Rule 1*, conditional on the identification(s) of a single variable

[21]Note that we here focus on the probabilities of disagreement, not on the adjusted probabilities.

[22]The four configurations are relevant in the case of a corner observation (i.e., *Fixed* in one classification and *Float* in the other one) as this observation can also fall in the alternative of the *Intermediate* category in the synthesis classification. In other cases, there are two estimation samples and only three configurations are possible: *(i)* "*2×FN*", *(ii)* "*1×FN & 1×TP*", and *(iii)* "*2×TP*". Explanations about the failure of the models to detect some disagreement points are provided below.

as the disagreement source (i.e., precise identification(s)).[23] Precisely, the refinement rule —*Rule 2*— consists in comparing the different sample-specific probabilities, considering the probabilities derived from the submodel excluding the identified source(s) of disagreement in case of precise identification.[24] Thus, the refinement rule is a more specific rule describing a context in which the argument from *Rule 1* is not strong enough to select the most probable ERR category.

Once we have assigned each disagreement point to the most probable ERR category, it is possible to identify the —primary— sources of disagreement from the resulting (core) synthesis classification. By definition, they correspond to the sources identified in the sample coinciding with the unified framework.[25]

B.1.2. The data

The explanatory variables' selection is relatively straightforward, given that all the variables involved in the two classifications are known. Firstly, the two classifications differ regarding the time horizon over which the nominal exchange rate changes are calculated. The *LYS* classification focuses on the average of the absolute monthly percentage changes in the nominal exchange rate over a calendar year. In contrast, the *RR* classification focuses

[23]This refinement is a way to address the so-called confirmation bias —or my-side bias— which designates a tendency to search for, interpret, favor, and recall information in a way that confirms one's preexisting beliefs or hypotheses (Kahneman and Tversky, 1974; Plous, 1993). This type of cognitive bias leads to a systematic error inherent to inductive reasoning toward confirmation of the hypothesis under study. In short, it can be considered as a form of selection bias in collecting evidence. Let us illustrate this bias in our context by relying again, tirelessly, on our example of a given country A's regime classified as *Fixed* by the *RR* classification but *Float* by the *LYS* classification. Suppose that the disagreement on country A's regime is precisely identified in the *Fixed* ERR sample (i.e., status "TP") but not in the *Float* one (i.e., status "FN"). This provides prima facie evidence for considering the ERR of country A as closer to a *Float* ERR than a *Fixed* ERR (the disagreement probability being lower in the *Float* sample). Now, suppose further that the official reserves volatility is identified as the source of the disagreement in the *Fixed* ERR sample. In the face of this new information, one should challenge its beliefs/perceptions rather than sticking to *Rule 1* as another plausible hypothesis could be worth considering. Indeed, considering in this case that the major source of the disagreement is the use by the *LYS* classification of the reserves volatility is not meaningless. Underlying this is the proposition that the synthesis classification should record country A as a *Fixed* instead of a *Float* regime. Within our —inductive reasoning— framework, both conclusions have their place and are "equally" important. Indeed, it is not about being right, but rather being the more likely. It follows then that the comparisons of the sample-specific probabilities involving precise identification(s) cannot be performed using *Rule 1*.

[24]In our example, this new disagreement probability can be interpreted as the new distance between country A's regime and the *Fixed* ERR sample —after removing the effect of the identified disagreement source. By the way, note that precise identification excludes the case where a disagreement is associated with several variables simultaneously (multiple identifications). The sample-specific probability of disagreement is, in this case, that derived from the full model.

[25]While the existence of two decision rules may be perceived as ad hoc and consequently of nature to be accommodating, *Rule 2* allows for an update/a questioning of beliefs while preserving the general idea of the synthesis classification. Rather than dismissing or embracing new evidence as though nothing else matters, the refinement introduced by *Rule 2* helps to take into account, coherently, the various information hence ensuring a higher degree of consistency.

on the absolute percent change in the monthly nominal exchange rate averaged over five-year rolling windows —two-year in some cases. Secondly, the nature of the nominal exchange rate also matters. The *LYS* classification is based on official exchange rates, while the *RR* classification uses, in some cases, parallel market exchange rates. Thirdly, the *LYS* classification considers the official reserves' volatility to capture interventions on the exchange rate market.[26] Finally, the two classifications differ regarding the thresholds that define the different ERR categories' perimeters.

While these key variables should theoretically be incorporated in the model to be estimated, it is far more difficult to practically include such variables for the econometric analysis. In particular, while we can address the first three variables (the time horizon over which the exchange rate changes are calculated, the use of parallel market exchange rates, and the official reserves' volatility) in a consistent empirical framework, the inclusion of the last variable —the thresholds that determine the different ERR categories— is a whole lot trickier. However, if we assume that the model is perfectly specified —i.e., no omitted variables—, we can reasonably attribute observations misclassified by the model (i.e., false agreements or false disagreements) to the threshold differences between the two classifications. A second issue is accounting for the various sources that explain the differences between the two classifications in measuring the volatility of exchange rates. These differences can be related to the nature of the nominal exchange rate (official versus parallel market exchange rates) or the time horizon considered for assessing the exchange rate volatility (year-by-year approach versus five/two-year rolling window). Disentangling the effects of these two sources proves to be complicated. As a result, we simultaneously account for these two effects by computing the difference between the exchange rate volatility measures used by the two classifications:

$$Diff.H/P = \sigma_{Pk} - \sigma_e \tag{B.1.2.1}$$

where σ_{Pk} (resp. σ_e) stands for the measure of exchange rate volatility used in the *RR* (resp. *LYS*) classification.

A third issue arises from the two measures of exchange rate volatility used by the *LYS* classification: *(i)* the *exchange rate volatility* (σ_e), *(ii)* the *volatility of exchange rate changes* ($\sigma_{\Delta e}$). As shown in Table B.1.2.1, the two measures display very high correlations regardless of the considered sample. To remove the collinearity problem arising from the

[26] As noted above, the *RR* classification also considers the inflation rate in its procedure. Still, this variable is only intended to differentiate the "Freely falling" category —composed of countries whose twelve-month inflation rate is above 40 percent — from the others. As a reminder, we have dropped this category.

inclusion of both measures in the specification, we perform a principal component analysis (PCA) to obtain the latent variable, i.e., the unobservable variable which underlies the observed collinear variables. The first latent variable, the first component following the PCA terminology, concentrates 97.6% of the volatility measures variance (see the online appendix). We thus select this first component —which we will refer to as "*exchange rate (ER) volatility*"— as an explanatory variable instead of the two volatility measures.

Finally, we also control for the LYS's algorithm specificities by including *(i)* a dummy, Outlier, capturing whether the observation is labeled as an outlier, and *(ii)* a dummy, *Round2*, scoring 1 if the observation is classified in the second round, 0 otherwise.[27]

Table B.1.2.1 — Correlations: volatility of the exchange rate and of the exchange rate changes

Sample	Full sample	Fixed	Intermediate			Float
			Lower	*Upper*	*Full*	
$Corr(\sigma_e, \sigma_{\Delta e})$	0.9503	0.9806	0.9678	0.9113	0.9516	0.9119
[p.value]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

The equation (full probit model) to be estimated is thus as follows:[28]

$$Y_i = \alpha + \beta_1 Diff.H/P_i + \beta_2 ER.Volat._i + \beta_3 Reserves_i + \beta_4 Outlier_i + \beta_5 Round2_i + \varepsilon_i \tag{B.1.2.2}$$

The data on the parallel market exchange rates are from Carmen Reinhart.[29] We collect data on official nominal exchange rates from the *International Financial Statistics* database (IMF). To consider the effect of the *LYS* classification's official reserves, we have no other choice than to rely on a categorical variable derived from the *LYS* classification regimes. Indeed, despite the indications regarding the sources of the variables in Levy-Yeyati and Sturzenegger (2016), we could not collect the data needed to calculate the official reserves volatility for all countries. Nonetheless, the *LYS* classification, relying on cluster analysis, allows us to generate a variable regarding the reserves' volatility based on the different regimes. As noted above, in the *LYS* classification, the *Floating* ERR

[27]Again, in the *LYS* classification, outliers correspond to observations with very high variability —the two percent-upper tail of observations for each of the three classification variables. Similarly, the distinction between the first and second rounds mirrors observations with high and low variability.

[28]It is worth noting that while the different variables can have interactions, we do not include them in the model. The reason is that it is difficult to compute standard errors for interaction terms in probit regressions. Furthermore, while we can address this issue of the standard errors through a Bayesian probit model —i.e., the uncertainty of interaction terms, such analysis introduces bias regarding the estimated coefficients of the other variables due to its sequential inclusion and exclusion of the variables in the estimated models. Thus, the effects of interactions would be deduced ("True positive" points not associated with a unique variable).

[29]Website: https://carmenreinhart.com/exchange-rates-official-and-parallel/ (last accessed: November 2020)

category should exhibit relatively stable official reserves compared to the *Intermediate* and *Fixed* ERR. Hence, we compute a binary variable to control for the use and the importance of the reserves volatility. As the *Floating* category is the only one with a different level of reserves volatility —compared to the other *LYS* categories, our variable scores 1 for the *Floating* ERR and 0 otherwise.[30]

B.1.3. Results

B.1.3.1. The full model performances

Estimation results of the full model over the different samples are detailed in Table B.1.3.1.1. We report both the coefficients (standardized) and the average marginal effects. As expected, the coefficients vary significantly from a sample to another hence justifying the use of different estimation samples.[31]

The difference in the time horizon and/or the use of parallel market exchange rates display significant and positive coefficients for the *Fixed* and *Intermediate* ERR samples. This result suggests that a greater difference in the volatility measure increases the likelihood of disagreements between the *RR* and *LYS* classifications. The effect is positive and notably stronger in the *Intermediate* —lower— ERR sample —see the average marginal effect. In contrast, the coefficient is significant and negative when considering the *Floating* ERR sample. These opposed signs are in line with expectations. Indeed, the higher the difference —i.e. $\sigma_{P_k} > \sigma_e$—, the more likely the *RR* classification will assign the observations to the *Floating* category. Thus, the probability of observing a disagreement between the two classifications will be lower if the observations are also recorded in the *Floating* category by the *LYS* classification.

The *exchange rate volatility* is also associated with a positive sign in the *Fixed* ERR sample: an increase in the volatility leads to a higher predicted probability of disagreement, i.e., the probability of not being classified as a *Fixed* ERR. This result also holds in the Upper *Intermediate* ERR sample. However, in the *Floating* ERR sample, an increase in the exchange rate volatility reduces the likelihood of disagreements.

Reserves, when included, display the highest coefficients and average marginal effects. Except for the *Floating* ERR sample —where it is associated with a negative sign, the use of the reserves volatility in the *LYS* classification —or the distinction between the level

[30]Our variable capturing the reserves volatility is entirely in line with the country groupings in the *LYS* classification. It allows us to differentiate *Floating* ERR (here, more in the sense of countries that do not intervene in the Forex) from the other categories. Hence the coefficient on "Reserves" can be interpreted as the use of the reserves volatility in the classification process and the importance of the volatility.
[31]The unbalanced dependent variable led us to consider alternative estimation methods for the *Float* sample. Results indicate that the probit estimates do not suffer from bias. See the online appendix.

of volatility— significantly increases the likelihood of observing a difference between the two classifications. This effect is particularly marked in the *Upper-Intermediate* ERR sample since the observations are at the border of the low and high official reserves volatility —following the *LYS* clustering rationale.

Observations labeled as *Outlier* by the *LYS* classification are associated with a higher probability of disagreement for all the samples, except for the *Upper-Intermediate* ERR sample. The coefficient displays a negative sign for this sample, suggesting a relatively low probability of observing a disagreement. As noted above, observations labeled as *Outlier* correspond, in the *LYS* classification, to observations with very high variability —the two percent-upper tail of observations for each of the three classification variables. Consequently, these outliers are more present in the *Upper-Intermediate* ERR sample —specifically in the dirty float category in the *LYS* classification— since this category regroups observations with the highest volatility of exchange rate and reserves. Therefore, the obtained negative (resp. positive) sign in the *Upper-Intermediate* (resp. other) ERR sample(s) seems coherent.

The round of the classification has a different effect depending on the considered sample. In the *Fixed* and *Intermediate* ERR samples, observations classified in the second round by *LYS* are associated with a lower predicted probability of disagreement. On the contrary, in the *Floating* ERR sample, these observations are associated with a higher probability of disagreement. As the second round of the *LYS* classification focuses on observations with low variability, the observed coefficient signs also appear consistent.

Since we are interested in the models' performances in detecting disagreement points, we depart from the traditional indicator of the overall goodness-of-fit, the pseudo R-squared in our case, even if it provides some clues. As stated above, ROC analyses actually suit well the binary nature of the exercise. Figure B.1.3.1.1 plots the ROC curves for the estimations presented in Table B.1.3.1.1. As visible, the models appear to do well with AUROC (Area under the ROC curve) ranging between 0.72 (*Lower-Intermediate* ERR sample) to 0.99 (*Upper-Intermediate* ERR sample).[32] Hence, overall, these preliminary analyses give more than satisfactory results to proceed to the identification of the disagreement sources.

[32]The area under the ROC curve is the performance measurement associated with the ROC analysis. It ranges between 0 and 1, 1 indicating an excellent model.

Table B.1.3.1.1 — Probit model results

Sample	Fixed		Intermediate						Float	
			Lower		*Upper*		*Full*			
	Betas	*AME*	*Betas*	*AME*	*Betas*	*AME*	*Betas*	*AME*	*Betas*	*AME*
Diff. Horizon/Premium	0.241**	0.077 **	0.318***	0.107***	0.488***	0.029***	0.316***	0.054***	-0.346***	-0.059***
	(0.098)	(0.031)	(0.063)	(0.020)	(0.093)	(0.005)	(0.066)	(0.011)	(0.064)	(0.011)
E.R. volatility	0.475***	0.152***	-0.039	-0.013	0.607***	0.037***	-0.019	-0.003	-0.773***	-0.133***
	(0.119)	(0.037)	(0.034)	(0.011)	(0.115)	(0.006)	(0.022)	(0.004)	(0.183)	(0.031)
Reserves	6.326***	0.661***	(omitted)		7.849***	0.797***	6.099***	0.387***	-6.237***	-0.157***
	(0.063)	(0.012)			(0.149)	(0.022)	(0.067)	(0.013)	(0.519)	(0.012)
Outlier	0.239*	0.054	0.816***	0.267***	-1.652***	-0.058***	0.757***	0.122***	4.818***	0.121***
	(0.143)	(0.048)	(0.245)	(0.071)	(0.466)	(0.008)	(0.241)	(0.034)	(2.264)	(0.009)
Round 2	-0.516***	-0.172***	-1.219***	-0.438***	-1.437***	-0.115***	-1.348***	-0.249***	0.922***	0.116***
	(0.073)	(0.024)	(0.103)	(0.031)	(0.195)	(0.018)	(0.104)	(0.015)	(0.198)	(0.017)
Constant	-0.159***		0.678***		-0.555***		0.9146***		7.258***	
	(0.053)		(0.078)		(0.137)		(0.077)		(0.538)	
No. Obs.	1569		894		1351		1941		1193	
Pseudo R^2	0.1545		0.1471		0.8279		0.4485		0.1546	

Notes: "Betas" stand for standardized coefficients (except dummy variables). "***" (resp. "**" and "*") indicates statistical significance at 1% (resp. 5% and 10%). Robust standard errors are reported in parentheses. The columns "*AME*" indicate the average marginal effects (Delta-method standard errors). "omitted" indicates that the variables are dropped due to collinearity.

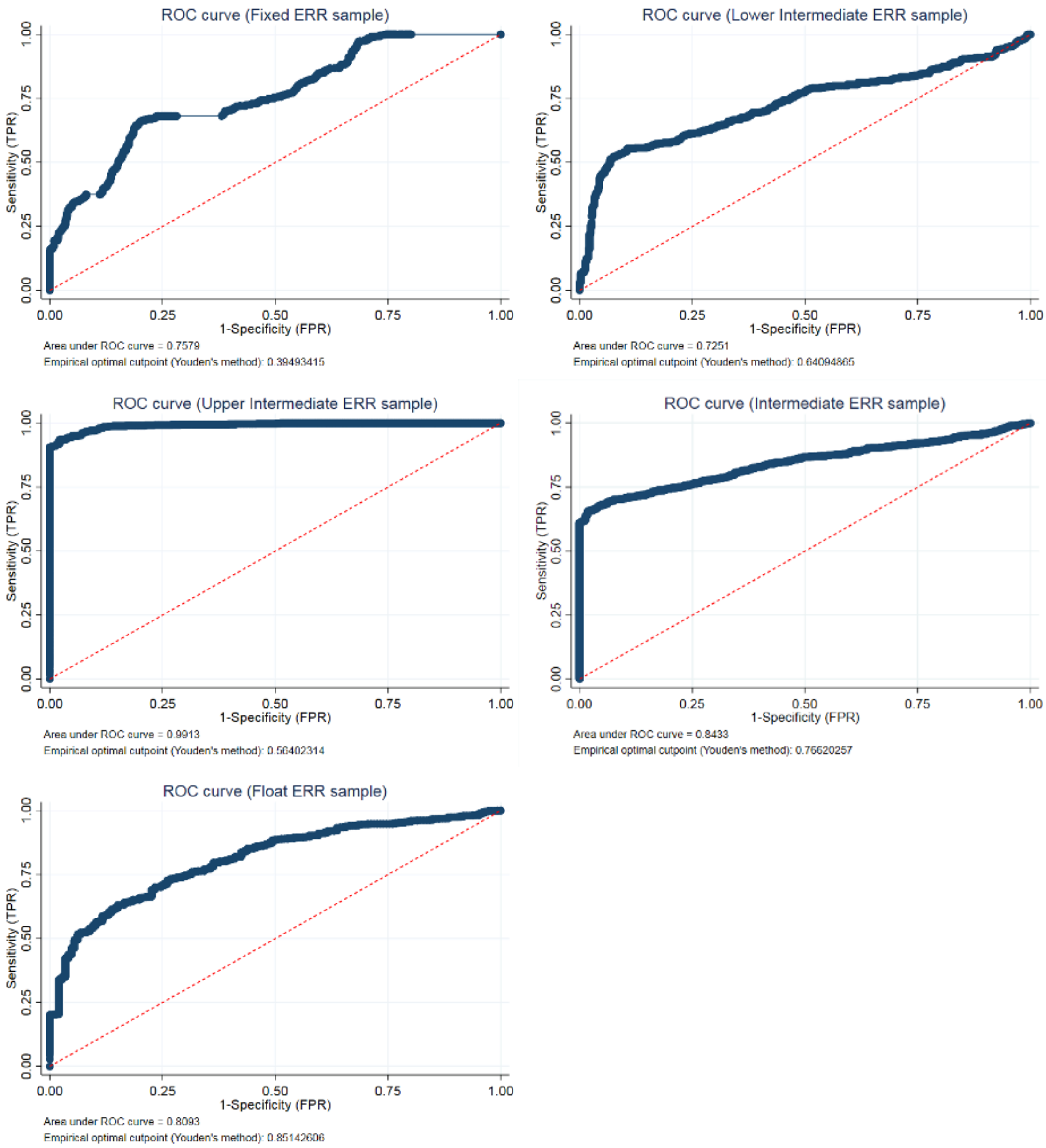

Figure B.1.3.1.1 — ROC curve (by estimation sample)

B.1.3.2. The disagreement sources

Figure B.1.3.2.1 schematically presents a summary of our findings.[33] The identified sources of disagreement are reported at the bottom of the figure. Table B.1.3.2.1 details the number and the percentage of disagreements by identified sources and by groups of countries.

For the whole sample, the primary vehicles responsible for the disagreements between the *LYS* and *RR* classifications are the differences in the thresholds delimiting the ERR categories and the entanglement of several sources ("*Multiple*").[34] The first explanation

[33]The intermediate analyses and results are reported in the online appendix to save space.
[34]It is worth noting the critical role of the *LYS* classification procedure based on a purely statistical method and its data-determined thresholds.

accounts for about 43% of the disagreements for EMEs and around a third for AEs and DCs (Table B.1.3.1) and holds particularly at the beginning of the period (Figure B.1.3.2.2). For Thailand, for instance, the differences in the thresholds between *RR* and *LYS* explain 18 —out of 29— disagreement points and, most specifically, the divergences noted from 1984 to 1996. For Colombia, where the disagreements —24 in total— imply the choice between a *Floating* regime (*LYS*) and an *Intermediate* regime (*RR*), the differences in the thresholds explain the divergences noted over the 1985-1999 period (excluding 1992 and 1993). For Brazil and Mexico, the differences in the thresholds explain respectively 75% and 78% of the disagreements.

The contribution of the second source —i.e., *Multiple*— varies more dramatically across regions: from 28.6% for EMEs to 34.9% for DCs and 47.2% for AEs. As shown in Figure B.1.3.2.2, the contributions of these multiple sources have increased since the mid-1980s for high-income countries. This latter source accounts for more than 85% of Ireland, Norway, and Denmark's disagreement points. For Canada (resp. Sweden), 14 of the 25 (resp. 13 out of 19) disagreement points have multiple sources.

The exchange rate volatility —measured by both *(i)* the changes in the exchange rate volatility and *(ii)* the volatility of these changes— is a relatively minor source of disagreements between the two classifications, accounting for 9.25% of disagreements in the whole sample. However, it is not very meaningful to separate this source of disagreements from the differences in the classifications' thresholds since the latter's definition is based on the exchange rate volatility. Similarly, the difference in the time horizon and/or the use of parallel market exchange rates explain(s) around 8.58% of the disagreements between the *RR* and *LYS* classifications. The proportion of disagreements associated with this explanation ranges from 1.3% in AEs to 13.12% in DCs. For AEs, this source of disagreements corresponds to years of financial turmoil or episodes of rapid currency movements (e.g., Singapore 1997 and 2008, Switzerland in 2010; Figure B.1.3.2.2) and is primarily related to the difference in the time horizon over which changes in the nominal exchange rate are assessed.

The use of the official reserves volatility is the fifth source of disagreements. It is associated with 7.83% of the disagreements —considering the whole sample. The proportion of disagreements explained by this variable varies from 6.35% in AEs to 8.51% in DCs. For instance, for Egypt, Greece, and Croatia, the use of the official reserves volatility accounts for 42.8%, 37.5%, and 35.3% of the disagreements, respectively.

The proportion of disagreements associated with observations classified in the second round by the *LYS* classification ("*Round 2*") is almost negligible. This latter reaches 1.35% when considering the whole sample; for AEs and EMEs, the proportions are 1.67% and

2.43%, respectively.

Finally, the disagreement point associated with *Outlier* corresponds to Nigeria in 1986.

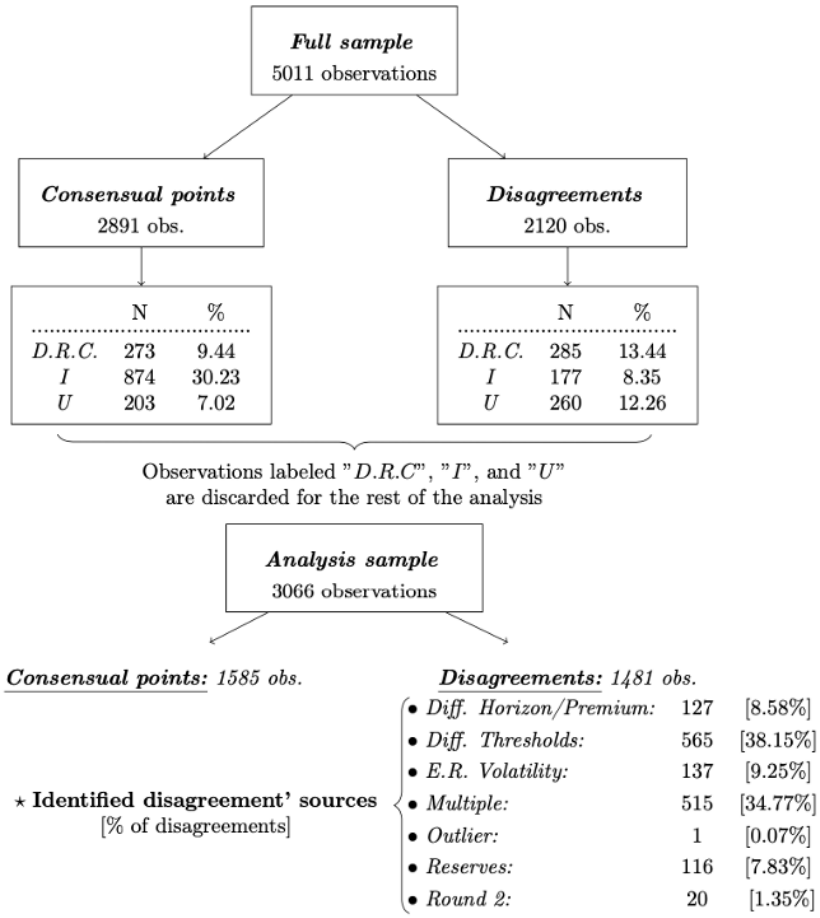

Figure B.1.3.2.1 — The disagreement sources (summary)

Note: "*U*" (resp. "*I*") indicates observations labeled as "*Uncontroversial*" (resp. "*Fixed inconclusive*") in the *LYS* classification. "*D.R.C.*" stands for "*Difference in the reference currency*".

Table B.1.3.2.1 — The disagreement sources by development level

	AEs	EMEs	DCs
Diff. Horizon/Premium	4 [1.34%]	49 [7.93%]	74 [13.12%]
Diff. Thresholds	94 [31.44%]	267 [43.20%]	204 [36.17%]
E.R. Volatility	36 [12.04%]	61 [9.87%]	40 [7.09%]
Multiple	141 [47.16%]	177 [28.64%]	197 [34.93%]
Outlier	—	—	1 [0.18%]
Reserves	19 [6.35%]	49 [7.93%]	48 [8.51%]
Round 2	5 [1.67%]	15 [2.43%]	—

Note: The figures correspond to the number of occurrences. Percentage (of the disagreements by sample) are reported in brackets.

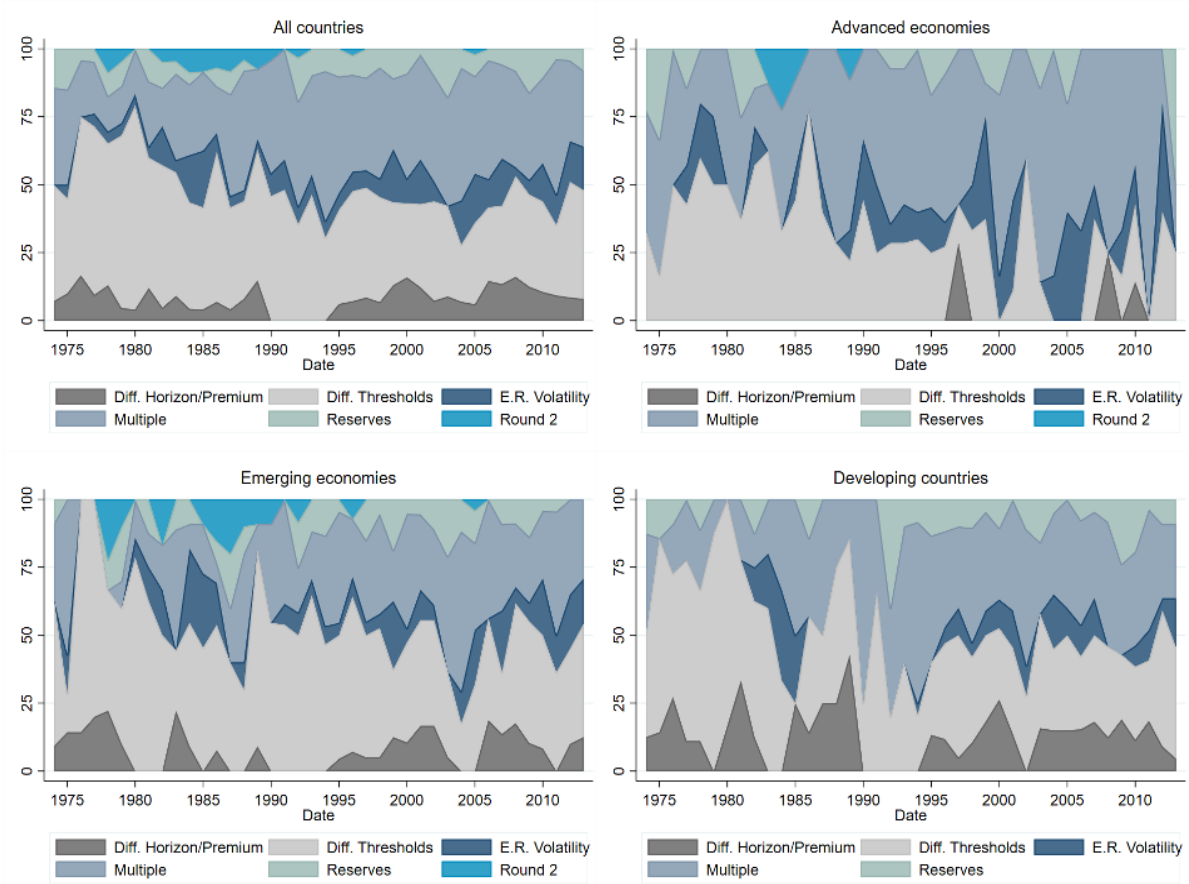

Figure B.1.3.2.2 — Evolution of the disagreement sources

Note: The charts present the evolution of the disagreement sources in percentage of the total disagreement points by year. We do not represent the observation associated with the source "*Outlier*" as this latter is imperceptible.

From our assessment, it is clear that relatively few disagreements observations —a bit less than 30% in total— are related to specific key variables underlying the *RR* and *LYS* classifications. Instead, our findings point out the complex nature of the disagreements since they largely originate from several variables' combinations and/or interactions. This feature is partly due to the *LYS* classification that makes joint use of several variables to classify the different ERR categories. Additionally, slightly more than a third of the disagreement points are due to the differences between the two classifications regarding the thresholds delimiting the three ERR categories. The complex nature of the disagreements between the two classifications provides additional support for a synthesis classification. Indeed, by providing a coherent framework, this latter overcomes the lack of confidence regarding ERR definitions that often plagues studies on the determinants and performances of ERR.

B.2. Standing on our own feet: towards an extended synthesis classification

This appendix is devoted to the presentation of the methodology underlying the extended synthesis classification. We first present the observations to be considered. We then discuss the ways to integrate them. Besides, we also address some blind spots in the *LYS* and *RR* classifications.

B.2.1. The observations

B.2.1.1. "Tricky" disagreement sources

As discussed previously, the identification of the disagreement sources between the two classifications reveals their complex nature. Indeed, only 27.08% of them have been precisely identified. The rest were associated with Multiple sources (515 observations, 34.77% of the disagreement points) or explained by the differences between the two classifications regarding the thresholds delimiting the categories (565 observations, 38.15%). As a result, we fell short regarding the definition of the ERR categories underlying these observations in the unified framework, i.e., in the synthesis classification (*SC*). It is worth mentioning that this is the result of a deliberate strategy, a fuse, to avoid a degeneration of the core *SC* —in these cases. Indeed, assigning these observations definitively to the ERR *vis-à-vis* which they displayed the lowest probability of disagreement (the above *Rule 1*) may leave room for the abovementioned confirmation bias. In case of unprecise identification —of the sources, for instance, one can no longer gauge the importance of the various sources and so take a position on the closest ERR category in the unified framework. The latter consequence follows from the fact that relying exclusively on *Rule 1* may be misleading. This is also true for the disagreement points attributed to the threshold differences identified indirectly through deduction. Thus, on balance, while our above methodology achieved its initial purpose —i.e., shed light on the disagreement sources, it did not permit us to disentangle the complexities of these disagreement types and so to **definitively** assign these latter to a category of the *SC*. Nevertheless, given the data availability and the definition of the core *SC* categories —thanks to the consensual observations and the precise identification cases, bypassing these complexities is fairly at hand. Reconsidering these "tricky" disagreement sources through the extension stage of the *SC* thereby serves as a reassessment step —of the category in the *SC*— based on a more elaborated approach.

B.2.1.2. Observations removed from the initial sample

Several conditions have guided the derivation of the core *SC* to ensure overall coherence. As a result, some observations have been removed from our initial sample since they blurred the perimeters of the different ERR categories defined by the classifications and so made the definition of a common conception of regime categories complicated at a minimum. More specifically, it concerned 1945 observations —38.81% of the original sample— either labeled as *"Fixed inconclusive"* and "*Uncontroversial*" or for which we noted a difference between the *RR* and the *LYS* classifications regarding the reference currency against which the exchange rate volatility is measured.

Observations labeled "*Fixed inconclusive*" in the *LYS* classification correspond to observations left unclassified after the second round. They were classified as *Fixed* ERR provided that they met one of the two following criteria: *(i)* zero volatility in the nominal exchange rate; *(ii) de jure* peg with average volatility in the nominal exchange rate smaller than 0.1%. Along the same lines, observations for which one variable was unavailable have been classified in "*Uncontroversial fix*" and "*Uncontroversial crawling peg*". The euro area countries have also been classified as "*Uncontroversial float*" on an *ad hoc* basis.

On the other hand, differences between the two classifications regarding the reference currency involve 558 observations, distributed between 273 agreements and 285 disagreements.[35]

Extending the synthesis classification (*SC*) with the first group of observations does not appear to be a big deal. However, the same exercise for observations with a difference in the reference currency is particularly tricky. Indeed, each classification surveyed a number of potential anchor currencies and selected the best anchor according to its methodology. Naturally, one cannot —*a priori*— discredit one anchor for the benefit of another. There are, however, few exceptions. Actually, for Australia, Germany, Japan, the United Kingdom, and the United States, the *RR* classification considers the national currency as the anchor, although the currencies are classified as floating. For these countries, we compute the volatility measures considering the anchor retained by the *LYS* classification.[36] For the

[35]Despite the differences in the reference currencies used by the classifications, most of the agreements are mainly the results of *(i)* the existence of "double pegs," i.e., a country is pegged to a currency which is itself pegged to another one (i.e., Luxembourg (1974-98) that had a pegged rate in the form of a monetary union with Belgium, and few countries pegged to the SDR); *(ii)* the *LYS* classification that classifies a country as a *Float vis-à-vis* an anchor currency while the *RR* classification considers the domestic currency as the anchor —pure float— (i.e., the Australian dollar, the Deutsche Mark, the Japanese Yen, and the US dollar). The rest of the consensual points are manifestly the result of an important correlation between the reference currencies (i.e., the US dollar and the SDR). As before, we do not reclassify these consensual observations.

[36]In most cases, both the *RR* and *LYS* classifications agree regarding these countries' ERR categories, so we did not undertake any reclassification for these points. Concerning the disagreement points, these countries were found to be floaters *vis-à-vis* the potential anchors against which their exchange rate displayed the lowest volatility. We instead retained the *LYS* anchors to ensure temporal coherence.

other observations, we look for the most appropriate anchor in our synthesis framework. We begin by computing the exchange rate's two volatility measures *vis-à-vis* each reference currency: *(i) à la LYS* and *(ii) à la RR*. We then derive the average volatility *vis-à-vis* each anchor and select the anchor currency against which the exchange rate exhibits the lowest volatility. This approach is particularly well suited for discriminating between pegs (or, to a lesser extent, soft pegs) and a more flexible ERR.[37]

B.2.1.3. Blind spots and shortfalls

The *SC*'s extension with the above-discussed observations would allow us to assign an ERR category to each observation from our initial sample. However, while one could argue that we have now completed our mission, the *SC* would still fall short in many respects. Indeed, the classification would only cover the 1974-2013 period and with significant gaps, limiting the classification's scope and usefulness. In fact, the *SC* is constrained by the samples' overlap between the *LYS* and *RR* classifications by construction. For the *SC* to considerably gain in scope, we need to overcome this constraint. This means filling the gaps and updating the classification to recent years.

Gaps in the *SC* have multiple sources. On the one hand, they stem from observations labeled as "freely falling" in the *RR* classification. In this latter classification, they form a specific category considered by Reinhart and Rogoff (2004) as dysfunctional arrangements —characterized by a twelve-month inflation rate greater or equal to 40%— for which the assessment of the ERR would be, according to the authors, at a minimum misleading.[38] On the other hand, gaps in the *SC* originate from observations not classified in the *LYS* classification, either due to a lack of data or because the observations correspond to *undisclosed basket pegs*. On this second point, the *LYS* classification suffers from inconsistency. *Undisclosed baskets* are not taken into account by the *LYS* classification because it is impossible in these cases to assess whether or not the countries are intervening to defend predetermined parities. However —and this is the crux, the authors do not make such checks either for "disclosed" baskets and simply consider the currency *vis-à-vis* which the exchange rate exhibits the lowest volatility as the anchor.

Since the various above sources of gaps are not concomitant, there is room to enrich the *SC*. First, *undisclosed baskets* can be classified following the same approach as for basket pegs with known central parity or disclosed weights. The *RR* classification adopts

[37]Indeed, the chosen reference currency is the one *vis-à-vis* which the domestic currency exhibits the lowest volatility. Being considered as a *Float* against this latter implies the same categorization if one had resorted to the other —not retained— reference currency. Hence, overall, the approach allows us to detect pegs (and soft pegs) for which the reference currency's issue makes sense.

[38]The first six months following an exchange rate crisis where this latter marked a transition from a (quasi) peg to a managed or independent float are also labeled as freely falling episodes.

the same approach for both "disclosed" and *undisclosed basket pegs*. We follow the same logic —for each observation labeled as an *undisclosed basket* by *LYS*— and compute the volatility measures —*à la LYS*— against the currency *vis-à-vis* which the exchange rate exhibits the lowest volatility. Second, observations considered as "freely falling" in the *RR* classification can also be incorporated in the *SC*, allowing the users to control the economic environment —i.e., inflation— in their studies. The problem with these observations in the *RR* classification is that the classification relies on the parallel market's exchange rates. In such a case, one may end up lumping together floaters with countries experiencing high and volatile inflation that triggered/fueled the change in the risk premia —and so the parallel market exchange rate changes— despite no changes in the official rate. Hence, in the *RR* classification context, distinguishing high inflation episodes is crucial. However, in our synthesis framework, the use of both the official and parallel exchange rates counteracts the risk premium changes. It is, therefore, possible to classify these observations given that, ultimately, we are interested in the exchange rate behavior. Since there is no anchor suggested for any of the countries classified as "freely falling", we follow the general strategy and retain as the anchor the currency against which the exchange rate exhibits the lowest volatility —*à la RR*. We consider the set of anchor currencies used in both the *LYS* and *RR* classifications for this exercise and the previous.

B.2.2. Methodology

Before discussing the empirical methodology, it is essential to note that the data availability issue —namely for the update to recent years— is no longer a matter of concern within our framework. The unified framework proposed by the *SC* allows us to get rid of several constraints, especially the availability of data on official reserve changes —as measured by Levy-Yeyati and Sturzenegger (2005). Similarly, the *Outlier* and *Round2* dummies included as controls for identifying the causes of the divergences no longer have any interest. Actually, from our initial set of variables, only *(i) Diff. Horizon/Premium* and *(ii) E.R. Volatility* not only appear relevant but also stand surety for the preservation of our unified framework.[39]

Regarding the empirical methodology, we rely on a discriminant analysis to extend the *SC*. We use the k^{th}-nearest-neighbor (KNN) algorithm to determine the ERR categories of the "new" observations. In a nutshell, the KNN algorithm, used namely in the growing field of Pattern Recognition, is a supervised machine learning non-parametric classification algorithm that seeks to find the best group to assign an observation by looking at

[39] Further note that our approach does not, in any sense, constitute a side step or a back door. Most observations considered here were classified in the *LYS* and *RR* classifications, relying exclusively on exchange rate data.

the immediate neighborhood of the observation. Thus, a data point, say x, is assigned to the predominant class (or group) of its k closest neighbor(s).[40] Say differently, the algorithm predicts the observation's group based on its "similarity" to observations within the group —while attempting to preserve group's statistics. Our choice in favor of the KNN methodology is also motivated by its nonlinear properties, namely its ability to distinguish irregular-shaped groups, including groups with multiple modes —see Rencher and Christensen (2012) for additional details. However, for the KNN methodology to deliver its promises, "k", i.e., the number of nearest neighbor(s) to be considered, is crucial. On the one hand, a small value of k can introduce noise and make the outcomes sensitive. On the other hand, larger values of k are associated with stable outcomes but increased bias. Since the whole methodology rests on distance calculations —which are even more of a concern in a multivariate framework, the method's choice is also essential. Besides, questions related to the selection of the priors and the best learning sample also arise.

We examined all the above issues and summarized the results in Figure B.2.2.1. As visible in the left chart, the *SC* sample appears to be the best learning sample. Indeed, it displays higher performances than the Consensual sample regarding the prediction of the ERR categories for the consensual observations. It also outperforms when predicting the ERR categories for disagreement points classified in the core *SC*. Furthermore, assuming proportional priors —i.e., group membership proportional to the group size— slightly improves the *SC* sample's performance —regardless of the distance measure. The right chart confirms this, indicating that the in-sample performance is maximized with $k = 2$ —for both the Euclidean and the Mahalanobis distances.

Based on the above, we select two values of k and predict the ERR category memberships and the associated probabilities. We consider $k = 2$ and $k = 8$ (see the vertical dashed lines in the right chart) and consider both the Euclidean and the Mahalanobis distance —assuming proportional priors. Regarding the priors, the relatively weak in-sample performances —of the simulations performed assuming $k = 2$ and equal priors— are worth noting. This ultimately explains our choices. Finally, it is worth noting that while one could have expected a single setting for the KNN algorithm, plurality is here probably the best way to ensure "robust" and "consistent" results, namely by allowing the k-based simulations to compensate each other shortcomings.

[40] Note, however, that the algorithm is flexible enough with the value of k. Actually, in the event of ties, the next largest value of k is selected. For instance, for a given observation, if there are more than k, say m, data points with equal distances from the observation, the computation will be based on all the m nearest points.

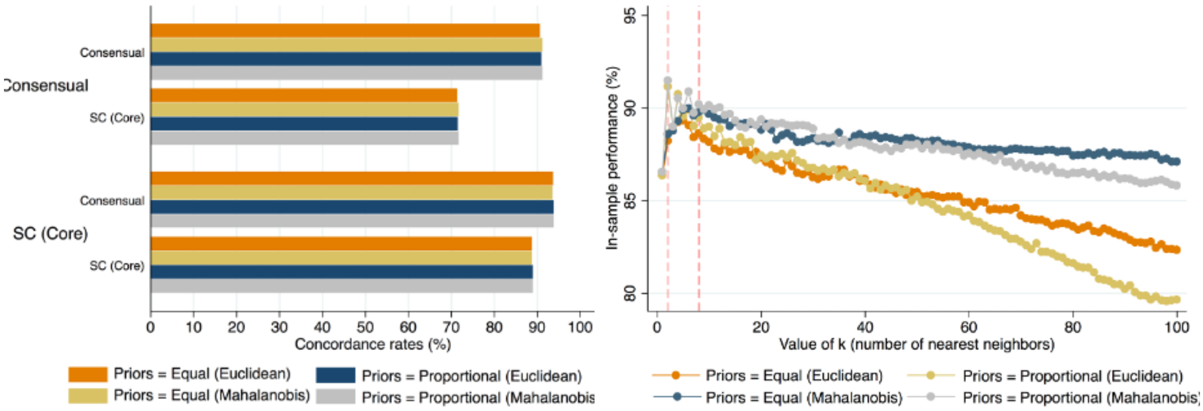

Figure B.2.2.1 — The KNN setting tests

Notes: Results in the left chart were obtained considering arbitrarily in an initial phase $k = 3$. "Priors" indicates the assumption regarding the prior probabilities for group membership (*equal* or *group-size-proportional*). We z-normalize the data. The "Consensual" sample is composed of all consensual observations in the original LYS and RR datasets.

B.2.3. The results

B.2.3.1. The outcomes

Applying our methodology led to 3912 consensual classifications. 86.6% of the observations to be classified were assigned to the same ERR category regardless of the KNN settings. For the remaining observations not unanimously classified, we relied on the group membership probabilities —predicted along with the class predictions by the KNN algorithm. More precisely, we assigned an observation to the ERR category *vis-à-vis* which it had the highest membership probability (average across the four predictions). 576 observations have been classified with this approach. Finally, note that we failed to assign an ERR category to 31 observations. Equal group membership probabilities explain this. We left these observations unclassified in the extended *SC*. One can, however, assign these latter to the *Intermediate* ERR category without substantial risk of error since they appear to be crossing data points. Inversely, we reclassified in the *Fixed* category 35 observations corresponding to single devaluation episodes (11 months without volatility) without any subsequent change in the exchange rate policy (i.e., *Fixed* category in the *SC* before and after the devaluation year). These observations, namely, include the CFA franc zone members that devalued in 1994 but kept their long-time standing anchor to the French franc.

All in all, the core synthesis classification has been supplemented with 4488 observations. The total number of observations —with an ERR category— in the extended *SC* reaches 7780. Table B.2.3.1.1 breaks down these observations by classification step.

Table B.2.3.1.1 — Summary of the different methodologies and outcomes

		Observations	
		Number	Percent
Core *SC*	Consensual observations ($RR = LYS$)	2891	37.16
	Diverging observations reclassified	401	5.15
	Consensual KNN-based classifications	3912	50.28
	Observations classified using the KNN-based probabilities	576	7.40
Extended *SC*		7780	100

B.2.3.2. Characterizing the exchange rate changes: Fixed and Intermediate categories

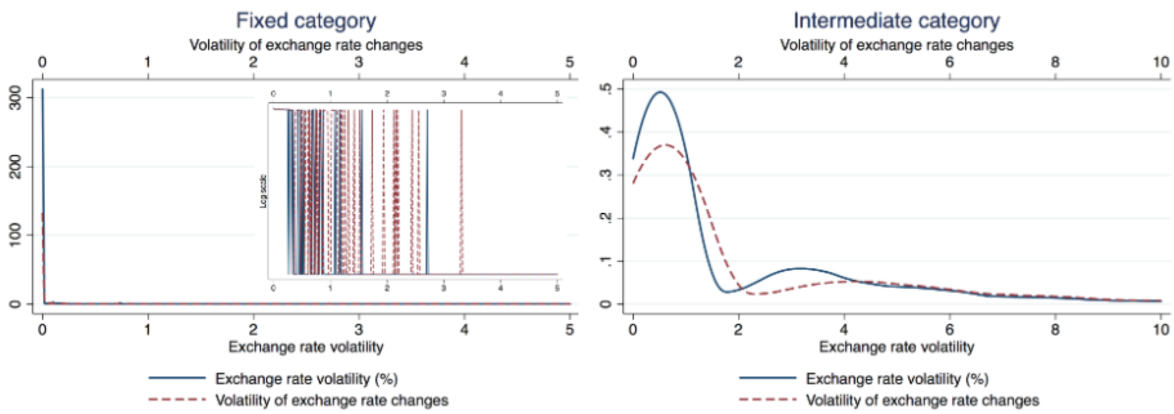

Figure B.2.3.2.1 — Distributions of the exchange rate changes (by *SC* ERR category)

Notes: kernel density estimates. The variables are computed *à la* Levy-Yeyati and Sturzenegger (2005). The inner chart in the left panel shows the volatilities on a logarithmic scale.

Appendix C. Empirical implications of the synthesis classification

Table C.1 — Exchange rate regimes and crisis susceptibility (probit estimates)

	Common sample					Respective samples				
	No ERR	IMF *de facto*	LYS	RR	SC	No ERR	IMF *de facto*	LYS	RR	SC
	(C.1.1)	(C.1.2)	(C.1.3)	(C.1.4)	(C.1.5)	(C.1.6)	(C.1.7)	(C.1.8)	(C.1.9)	(C.1.10)
Intermediate		0.336	**0.451****	0.079	**0.563****		0.151	**0.535*****	0.045	**0.430****
		(0.230)	(0.219)	(0.247)	(0.239)		(0.150)	(0.206)	(0.187)	(0.169)
Float		0.078	0.037	0.130	0.432		0.064	0.178	0.212	**0.395****
		(0.268)	(0.281)	(0.280)	(0.268)		(0.153)	(0.222)	(0.201)	(0.162)
Current account balance/GDP	-0.188	-0.103	-0.143	-0.219	-0.059	-0.243	-0.230	0.002	-0.286	-0.234
	(0.590)	(0.601)	(0.575)	(0.596)	(0.627)	(0.400)	(0.394)	(0.561)	(0.381)	(0.498)
Real GDP growth	-2.069	**-2.157***	-2.087	-2.143	**-2.444***	**-2.534*****	**-2.554*****	**-1.998***	**-2.700****	**-2.538*****
	(1.356)	(1.299)	(1.280)	(1.307)	(1.310)	(0.947)	(0.937)	(1.121)	(1.123)	(0.926)
Inflation	0.196	-0.055	-0.280	0.071	-0.584	0.644	0.553	-0.403	0.660	0.235
	(1.061)	(1.027)	(1.129)	(1.042)	(1.074)	(0.403)	(0.421)	(0.527)	(0.745)	(0.449)
REER misalignments	**3.813*****	**3.874*****	**3.866*****	**3.777*****	**4.026*****	**3.253*****	**3.266*****	**3.879*****	**3.140*****	**3.400*****
	(0.920)	(0.958)	(0.936)	(0.918)	(0.978)	(0.734)	(0.741)	(0.806)	(0.811)	(0.751)
Reserves/GDP	**-5.269*****	**-5.363*****	**-5.322*****	**-5.314*****	**-5.937*****	**-4.345*****	**-4.376*****	**-6.144*****	**-3.804*****	**-4.544*****
	(1.220)	(1.185)	(1.161)	(1.203)	(1.236)	(1.086)	(1.137)	(1.150)	(1.117)	(1.239)
Real GDP per capita	-0.068	-0.082	-0.066	-0.072	-0.088	-0.078	-0.087	**-0.129***	-0.054	**-0.104***
	(0.092)	(0.090)	(0.089)	(0.090)	(0.089)	(0.062)	(0.062)	(0.076)	(0.067)	(0.063)
Constant	-1.172	-1.057	-1.247	-1.161	-1.101	**-0.999***	**-0.932***	-0.546	**-1.362****	-0.903
	(0.918)	(0.903)	(0.917)	(0.920)	(0.907)	(0.544)	(0.545)	(0.732)	(0.649)	(0.553)
Observations	2,208	2,208	2,208	2,208	2,208	3,508	3,505	2,453	3,004	3,491
Pseudo R-squared	0.214	0.219	0.226	0.215	0.230	0.191	0.192	0.240	0.182	0.203
Number of crises	54	54	54	54	54	91	91	70	73	91
Intermediate = Float (p.value)		0.2325	0.1067	0.8032	0.5436		0.5048	0.0754	0.2437	0.7977
Sensitivity	0.83	0.85	0.87	0.81	0.87	0.78	0.78	0.86	0.81	0.85
Specificity	0.79	0.79	0.79	0.82	0.80	0.81	0.80	0.77	0.79	0.79
AUROC	0.81	0.82	0.83	0.82	0.84	0.79	0.79	0.81	0.80	0.82

Notes: The *Fixed* category is taken as the base category. " *** " (resp. " ** " and " * ") indicates statistical significance at 1% (resp. 5% and 10%). The common sample consists in observations available across all the classifications. All the models include regional fixed-effects and the standard errors (reported in parentheses) are clustered at the country level. All regressors are lagged one period. The line "Intermediate = Float (p.value)" reports the p.values associated to the coefficients equality tests between the *Intermediate* and the *Float* regimes. The Sensitivity, Specificity, and AUROC are calculated for each model optimal cut-off value.