



UNIVERSITÉ D'ORLÉANS
ET
MAASTRICHT UNIVERSITY



**ÉCOLE DOCTORALE SCIENCES DE L'HOMME ET
DE LA SOCIÉTÉ**

Laboratoire d'Economie d'Orléans / Department of Economics

THÈSE EN COTUTELLE INTERNATIONALE

présentée par :

Elena-Ivona DUMITRESCU

soutenue le : **31 Mai 2012**

pour obtenir le grade de :

Docteur de l'Université d'Orléans et de Maastricht University

Discipline : **Sciences Economiques**

Econometric Methods for Financial Crises

THÈSE dirigée par :

Mr. Bertrand CANDELON

Professeur, Maastricht University

Mr. Christophe HURLIN

Professeur, Université d'Orléans

RAPPORTEURS :

Mr. Massimiliano MARCELLINO

Professeur, European University Institute

Mme. Valérie MIGNON

Professeur, Université Paris X Nanterre

Mr. Franz PALM

Professeur, Maastricht University

JURY :

Mr. Bertrand CANDELON

Professeur, Maastricht University

Mr. Christophe HURLIN

Professeur, Université d'Orléans

Mr. Gilbert COLLETAZ

Professeur, Université d'Orléans, Président

Mr. Massimiliano MARCELLINO

Professeur, European University Institute

Mme. Valérie MIGNON

Professeur, Université Paris X Nanterre

Mr. Joan MUYSKEN

Professeur, Maastricht University

Mr. Franz PALM

Professeur, Maastricht University

To all those who encouraged me to pursue my dreams,

Acknowledgements

Looking back, I feel that the last few years as a PhD student were a stimulating and pleasant journey filled with constant challenges and plentiful rewarding moments. Taking a trip down the memory lane, I recall the beginning of the second year of master in econometrics at the University of Orléans, when Christophe Hurlin asked me if I would be interested in pursuing a PhD. The idea of a joint PhD with Maastricht University immediately came forth. Bertrand Candelon hence suggested the topic of my research master thesis: *Financial Crises Early Warning Systems*. In the fall 2009 when I started the joint PhD program, we decided that this subject is worth being more profoundly explored by adding a strong econometric flavor. Over the past three years, I therefore had the chance to learn a lot, both in econometrics and international finance, by working in different environments (Orléans, Maastricht, Florence) and by discussing with interesting people that marked the way I now address problems and formulate my ideas. I would like to take the opportunity here to thank some of them.

First, I would like to express heartfelt thanks to my PhD supervisors, Christophe Hurlin and Bertrand Candelon. I have learned so much from you, there are no words to describe my gratitude. Thank you for sharing your experiences with me, guiding me all the way, giving me precious career advice, and having confidence in me; it has made me grow as a person and researcher. I really appreciate your always open door, your availability on skype and e-mail when working at distance, as well as the time and effort you invested to make this joint PhD possible. Thank you also for including me in the PICS project: Econometric Approaches to Risk Modelling which provided most of the financial support for my stays in Maastricht and for attending top international conferences. Besides, as a research assistant for the Runmycode project I had the opportunity to further improve my knowledge and skills in econometrics. It was a privilege for me to work under your supervision.

I am very grateful to the members of my assessment committee: Massimiliano Marcellino, Valérie Mignon, Joan Muysken and Franz Palm for carefully reading and providing helpful comments on my manuscript. A special word of thanks to Franz Palm for his advices and fruitful discussions on the fourth chapter of this dissertation and to Peter

Reinhard Hansen from the European University Institute for collaboration during my visit in the fall 2011 and for accepting to be my mentor next year during the Max Weber Postdoctoral Fellowship.

Next, I would like to thank my friends and colleagues from both universities for the friendly working environment. A word of gratitude goes to Renée-Hélène, Caroline, Silvana and Fleur for their help in all administrative matters, especially since these duties become more complicated when it comes to joint PhDs. Alexandra, Camelia, Cristina and Denisa, thank you for being my friends! I truly appreciate all the good times we had during and after office hours and I hope time and space constraints will not interfere in our friendship. Audrey and Sylvain, thanks for motivating discussions about research and everyday life. Andreea, Lavinia, Lenart and Norbert, thank you for making my stay in Maastricht more pleasant. Warm thoughts to Andreea S., Andreea E., Mia, and Vanda, as well as to Lena and Claudia for helping me to quickly adjust to life in Florence.

Finally, I am extremely grateful to my parents. Thank you for your unconditional love and support.

Elena-Ivona Dumitrescu
Orléans, April 2012

Contents

Acknowledgements	i
List of Figures	vii
List of Tables	ix
1 Introduction	1
1.1 Are Financial Crises Predictable?	3
1.2 Early Warning Systems	5
1.3 Contribution	9
2 Testing Interval Forecasts: a GMM-Based Approach	15
2.1 Introduction	15
2.2 General Framework	17
2.3 A GMM-Based Test	19
2.3.1 Environment Testing	19
2.3.2 Orthonormal Polynomials and Moment Conditions	20
2.3.3 Testing Procedure	21
2.4 Monte-Carlo Experiments	23
2.4.1 Empirical Size Analysis	23
2.4.2 Empirical Power Analysis	26
2.5 An Empirical Application	30
2.6 Conclusion	31
2.7 Appendix	33
2.7.1 Appendix: J statistics	33
2.7.2 Appendix: Dufour (2006) Monte-Carlo Corrected Method	34
3 How to Evaluate an Early Warning System?	35
3.1 Introduction	35
3.2 Optimal Cut-off	38

3.2.1	How important is the cut-off choice?	38
3.2.2	A credit-scoring approach	40
3.2.3	Accuracy Measures	41
3.3	Evaluation Criteria	42
3.3.1	Cut-off based criteria	43
3.3.2	Example	44
3.4	Comparison Tests	47
3.5	Empirical Application	49
3.5.1	EWS Specification	49
3.5.2	Data and Estimation	50
3.5.3	EWS Evaluation	52
3.5.4	Optimal cut-off	55
3.5.5	Robustness Check	61
3.6	Conclusion	63
3.7	Appendix	65
3.7.1	Appendix: Comparison of <i>ROC</i> Curves Test	65
3.7.2	Appendix: Dataset	66
3.7.3	Appendix: A Robust Estimator of the Variance of the Parameters	68
4	Currency Crises Early Warning Systems: why they should be Dynamic	71
4.1	Introduction	71
4.2	A Dynamic Specification of EWS	74
4.2.1	Specification and Estimation	74
4.2.2	Panel Framework	76
4.3	Dynamic EWS Estimation	78
4.3.1	Dataset	78
4.3.2	Dating Currency Crises	79
4.3.3	Optimal Country Clusters	79
4.3.4	Estimation Results	80
4.4	Forecasts Evaluation	87
4.4.1	In-Sample Analysis	87
4.4.2	Out-of-sample analysis	95
4.5	Conclusion	101
4.6	Appendix	102
4.6.1	Appendix: Constrained Maximum Likelihood Estimation (Kauppi and Saikkonen, 2008)	102
4.6.2	Appendix: Modified Maximum Likelihood Estimation (Carro, 2007)	103

4.6.3	Appendix: Evaluation Methodology	107
5	Modeling Financial Crises Mutation	109
5.1	Introduction	109
5.2	A Multivariate Dynamic probit Model	111
5.3	Exact Maximum Likelihood Estimation	114
5.3.1	The Maximum Likelihood	115
5.3.2	The Empirical Procedure	116
5.4	Empirical Application	118
5.4.1	Dating the crises	118
5.4.1.1	The Database	118
5.4.1.2	Dating the Crisis Periods	119
5.4.1.3	Remarks	121
5.4.2	Bivariate Analysis	122
5.4.3	Trivariate Analysis	126
5.4.4	Further results	127
5.5	Conclusion	133
5.6	Appendix	135
5.6.1	Appendix: Proof of lemma 1	135
5.6.2	Appendix: The Gauss-Legendre Quadrature rule	135
5.6.3	Appendix: The EML score vector for a trivariate dynamic probit model	136
6	Conclusions	143
	References	146
	Résumé en Français	156
	Nederlandse Samenvatting	170
	Curriculum Vitae	173

List of Figures

2.1	Partial sums y_h and block size N	19
2.2	Figure 2. Corrected power of the $J_{CC}(2)$ test statistic as function of the sample size T (coverage rate $\alpha = 5\%$)	30
2.3	Figure 3. Corrected power of the $J_{CC}(2)$ test statistic as function of the block size N (coverage rate $\alpha = 5\%$)	31
3.1	Figure 1: Optimal Cut-off determination	41
3.2	Figure 2: The ROC curve	44
3.3	Figure 1: QPS - Graphical Approach	45
3.4	Figure 1: Brazil - Crisis probabilities	54
3.5	Figure 1: Cut-off - Regional Panel Models	57
3.6	Figure 1: Crisis Probabilities - Time-Series Models	58
3.7	Figure 1: Crisis Probabilities - Regional Panel Models (continued)	59
3.8	Philippines - Resilience	61
4.1	Figure 1: Predicted probability of crisis - in sample	93
4.2	Figure 1: Predicted probability of crisis -in sample (continued)	94
4.3	Figure 1: Predicted probability of crisis ($C1$) - out-of-sample	97
4.4	Figure 1: Predicted probability of crisis ($C1$) - out-of-sample (continued) .	98
4.5	Figure 1: Predicted probability of crisis ($C24$) - out-of-sample	99
4.6	Figure 1: Predicted probability of crisis ($C24$) - out-of-sample (continued)	100
5.1	Figure 1: Conditional crisis probabilities - Ecuador	128
5.2	Figure 1: IRF after a banking crisis shock - Ecuador 3 months	130
5.3	Figure 1: IRF after a debt crisis shock - Ecuador 3 months	131
5.4	Figure 1: IRF after a currency crisis shock - Ecuador 3 months	132

List of Tables

2.1	Empirical size (block size $N = 100$, nominal size 5%)	24
2.2	Empirical size (block size $N = 25$, nominal size 5%)	26
2.3	Feasibility ratios (coverage rate $\alpha = 1\%$	27
2.4	Empirical Power (block size $N = 100$)	28
2.5	Empirical Power (block size $N = 25$)	29
2.6	Interval Forecast Evaluation (SP500)	32
2.7	Interval Forecast Evaluation (Nikkei)	32
3.1	Example: Evaluation Criteria	46
3.2	EWS Estimation	52
3.3	EWS Evaluation: Regional Panel Model	53
3.4	EWS Optimal Cut-off: Descriptive Statistics	55
3.5	EWS Forecasting abilities	56
3.6	EWS Evaluation: Regional Panel Models (Robustness check)	62
3.7	EWS Optimal Cut-off: Descriptive Statistics (Robustness check)	63
3.8	Database	69
4.1	SBC information criterion (time-series logit models)	81
4.2	Estimation results (time-series logit models)	83
4.3	Estimation results (time-series logit models) - continued	84
4.4	Estimation results (time-series logit models) - continued	85
4.5	Estimation results (time-series logit models) - continued	86
4.6	Estimation results (panel logit models)	87
4.7	Evaluation criteria	88
4.8	Comparison tests	89
4.9	Optimal cut-off identification	91
4.10	Comparison tests (Out of sample exercise)	95
4.11	Optimal cut-off identification (Out-of-sample exercise)	96
5.1	Database	122

5.2	Percentage of crisis periods	123
5.3	Bivariate Analysis	124
5.4	Trivariate Analysis	127

Chapter 1

Introduction

Is there any hope for a world free of financial crises?

From a historical perspective, financial crises seem to be more like the rule rather than the exception (Reinhart et al., 2010; Bordo et al., 2001b). Proposing an exhaustive inventory of the numerous speculative episodes that have shaken various countries, starting from the famous “tulip mania” episode (around 1636) to the South Sea Bubble (1720), the crises following the wars between 1713 and 1820, the 1929 stock market crash, as well as the banking, exchange rate and sovereign debt crises since the ’80s (see Kindleberger, 2000, for a detailed list) thus seems both tedious and deceptive.

This dissertation has a different, narrower objective. Econometrically speaking, it tackles the technical feasibility of Early Warning Systems for different types of financial crises. A diversity of turmoils and economic mechanisms can be grouped under the name of “financial crisis” according to the markets or institutions they hit. In this introduction we consider the three main types of financial crises, *i.e.* currency crises, banking crises and sovereign-debt crises (Reinhart and Rogoff, 2008) and focus especially on currency crises, which are most often the object of the empirical applications developed in the following chapters. In a fixed exchange regime, a currency crisis is determined as a forced abandonment of the currency peg, resulting either in a realignment of the currency or even in a complete abandonment of the fixed exchange regime, whereas in a floating regime it consists in a strong depreciation in a short period of time as a result of a speculative attack (Bordo et al., 2001a). Caprio and Klingebiel (1996) define banking crises as a situation “when a significant fraction of the banking sector is insolvent but remains open”. A debt crisis involves outright default on payment of debt obligations, repudiation, or the restructuring of debt into terms less favorable to the lender than in the original (Reinhart and Rogoff, 2011).

Two main research directions can be identified in the financial crises literature since the ’90s. The first line of research endeavoured to understanding the economic mechanisms

that explain the emergence and the unfolding of past crises in general. The different generations of currency crises models proposed since the '80s represent the leading example of this research direction that consists mainly in theoretic advancements. In the meantime, a second direction has been concretized in a large number of both theoretical and empirical models for forecasting these turmoils, *i.e.* Early Warning Systems (EWS). And after each severe crisis the call for forecasting models by governing bodies increases, generally resulting in the development of new EWS models.

Historical evidence shows that financial crises continue to pass unobserved and to hit profoundly numerous countries in spite of the various EWS models proposed. Informed audience can actually notice a parallel evolution of the number of observed turmoils and the number of EWS models proposed in the literature and used by the authorities. These facts lead some academic researchers to doubt the usefulness of such models (*e.g.* Rose and Spiegel, 2010, 2011). Other researchers (*e.g.* Frankel and Saravelos, 2011), ignoring this positive correlation, subsequently show that a set of leading indicators for financial crises can be identified and hence research work in this field should be encouraged.

However, this academic debate on the existence of EWS does not call into question their usefulness to international organizations and other economic parties. Different institutions such as the International Monetary Fund, Federal Reserve, Credit Suisse, Deutsche Bank, and French Banking Commission, need such models (for monitoring economic health and in their decision-making process) in spite of their imperfectness when it comes to forecasting. The main reason behind this is that apart from expert's advice EWS are the only objective criteria justifying some policy measures and decision taking. Furthermore, EWS implementation sometimes turns out to be the result of political mandates. For example, in 2011 the G20 nations agreed that the International Monetary Fund should constantly monitor the levels of debt, budget deficits and trade balances of the countries representing more than 5% of the combined output of the G20, with a view to prevent another global financial crisis.

EWS hence play a key role in defining economic policies at microeconomic, macroeconomic, and international level. From an extreme viewpoint, disbelief in crises predictability is not incompatible with improving the results of EWS. We can draw an analogy with theoretical econometric research on non-linear forecasting models: some recent tests allow us to compare forecasts issued from potentially misspecified models and to identify the least misspecified one (Corradi and Svanson, 2006b). However, in the EWS literature there is no clear methodology to compare different specifications and to identify the best forecasting model (or least misspecified one), although it would provide additional support to the quest for the optimal forecasting model for financial crises.

Two questions arise in this context: how to improve the specification of EWS models? and how to choose the best forecasting model? In this applied econometrics dissertation we address (and answer) both questions by relying on recent methodological improvements in econometric forecasting, panel data econometrics and the econometrics of discrete-choice models. Still, before detailing the main insights of this thesis, a short (and hence incomplete) overview of the main theoretical and empirical advancements in explaining and in forecasting financial crises is presented.

1.1 Are Financial Crises Predictable?

One way to address the predictability of financial crises consists in assessing the causes of past turmoils and drawing inference about crises occurrence so as to construct EWS. Models that attempt to understand the underlying causes of financial crises and to offer appropriate policies to correct their effects are hence constructed. This approach, comparable to that of Cowles Commission, is used in the case of currency crises forecasting models for example.

The “first generation” models tried to explain the occurrence of currency crises that hit Latin-American countries in the 80’s. It considered the eruption of crises as the result of a fundamental inconsistency between domestic policies (expansionary monetary policy, mainly the persistence of budget deficits) and the attempt to maintain a fixed exchange rate. When the amount of reserves has sufficiently decreased, the speculators launch an attack, accelerating the consumption process of the reserves. Consequently, the policymaker is compelled to let the exchange rate go by devaluating the currency or by letting it float. In this model, unsustainable macroeconomic policies hence lead to vulnerability and the collapse of a fixed exchange rate regime (Krugman, 1979; Flood and Garber, 1984). The corresponding EWS hence rely on a set of macroeconomic leading indicators.

Despite the virtues of the canonical model, a number of economists have argued that it is an inadequate representation of the forces at work in most real crises (Obstfeld, 1994; Krugman, 1997). A “second-generation” crisis model which exhibits multiple equilibria evolved largely as a response to the two-stage crisis which wrecked the European Exchange Rate Mechanism in 1991 and 1992. In this case, speculative attacks are possible because of self-fulfilling expectations. It follows that a crisis can occur even without any visible macroeconomic weakness and even though the currency peg appears sustainable.

Still, these theories could not explain the episodes of speculative pressure and eventual turbulence that happened to countries affected by the Tequila crisis in 1994-95 and the Asian Financial Crisis in 1997-98. In this context, a “third generation” crisis models that explores the linkages between currency crises and the banking as well as the financial

sectors was born. In this type of crisis, a country's currency can be affected by a crisis from another country that could not be explained by economic fundamentals (Masson, 1998). This phenomenon is known as cross-country contagion (Bruinshoofd, Candelon and Raabe, 2008). In view of the large number of financial and banking instability mechanisms, numerous factors can be considered at the basis of crisis outset, from asset prices booms and balance-sheet crises (Schneider and Tornell, 2000) to excessive external debt (Corsetti, Pesenti and Roubini, 1999). It is a vicious circle of deleveraging that leads to severe costs for the real economy (Krugman, 2002). Note that joint banking and currency crises, known as twin crises (Kaminski, 1999), can be explained by these models.

This alternation between tumultuous periods and theoretical explication of their causes seems to indicate that models “run after” crises and hence puts into question the predictability of currency crises. Looking at the three generations of currency crises, it appears that the models trying to explain the origins of crises do not exploit past information efficiently, as most of the currency crises (and of financial crises in a more general context) have not been successfully predicted. Rose and Spiegel, (2010, 2011) hence refute the idea of an EWS itself. However, the enormous social and economic costs engendered by the recent crises emphasize the importance of developing EWS models regardless of their deficiencies.

Numerous studies have shown that the effects of financial crises are quantifiable in terms of very high bailout cost, enormous output losses, and high level of unemployment. According to Caprio and Klingebiel (1996), bailout costs represent on average 10% of GDP, some crises proving much more costly than others, *e.g.* the Mexican Tequila Crisis (1994) cost 20% of GDP while the Jamaican crisis (1996) cost 37% of GDP. Additional costs of foregone economic output (mainly reduced investment and consumption) correspond to credit rationing and uncertainty, among others. An IMF report (1998) estimates that emerging markets suffer a cumulative loss in real output of around 8% during a severe currency crisis. Hoggarth et al., 2002 find that, on average, banking crises increase the cumulative output gaps by 13% of GDP while fiscal costs are 18% of annual GDP higher when associated with a currency crisis. With respect to the global financial crisis, the IMF noted in 2009: “Global GDP is estimated to have fallen by an unprecedented 5 per cent in the fourth quarter (annualized), led by advanced economies, which contracted by around 7 per cent”. These enormous costs more than offset Rose and Spiegel's arguments. The cost of failing to detect a crisis largely outweighs the cost of developing a comprehensive EWS that may draw a signal of alarm prior to the arrival of a crisis. It follows that correctly forecasting a part of the financial crises across the

globe, would imply large gains (savings), as opposed to the losses incurred by such an event. EWS models are hence an essential tool for the regulatory authorities.

Beyond this argument of high costs and regulatory decisions motivating the need for EWS, the question of crises predictability and identification of an “optimal” EWS persists. Note that the object studied by an EWS (currency crises, for example) is non-invariant to changes in economic policy and microeconomic environment.¹ The sequence of crisis and calm periods can actually be modified by economic policy measures, in contrast with the microeconomic literature, *e.g.* credit-scoring, where the crisis event (individual default) is invariant to the actions undertaken by banks. The outcome of the EWS (crisis or calm, in a binary setting), which relies on past information, will, therefore, not correspond anymore to the ex-post realization in view of the new economic context (the new policy rules implemented). To put it differently, an “optimal” EWS must be a false model from the perspective of type I error, as it is supposed, by definition, to lead to crises prevention. Policy measures relying on a timely early warning signal should lead to the prevention or at least to the attenuation of the impact of a crisis. Consequently, in the presence of efficient policy intervention forecast error (taking the form of false alarms) will arise. Careful ex-post comparison of the probabilities issued from an EWS (hence not taking into account the policy measures) with the observation of crisis / calm periods is hence required.

Another issue related to crises forecasting consists in the choice of the “optimal” forecast horizon. An EWS is useful if and only if it gives the authorities enough time to take the necessary steps leading to the prevention or at least to the attenuation of the impact of such events. Signalling an impending crisis 1/2 hours before its arrival is worthless, in view of the authorities’ response time. It follows that in this literature longer forecast horizons (6 months to 2 years) are needed than in the classic econometric forecasting literature.

In view of the numerous interlinkages between macroeconomic fundamentals and the implementation of new economic policies, especially in an context of increasing interconnectedness of financial institutions and markets, it seems that new perspectives must be taken into account so as to succeed in better forecasting these events. The role of EWS is even more crucial, as financial fragility easily spills-over from one country to another leading to a global financial crisis, as the one recently experienced.

1.2 Early Warning Systems

The aim of EWS models is to forecast crises, *i.e.* the crisis periods, the outset of crises or even their duration as correctly as possible. They are hence vital for crisis prevention,

¹This goes along the lines of Lucas’ critique.

as discussed in the previous section. Two research issues should be scrutinized in this context, namely the specification of EWS models and their evaluation. Econometrically speaking, the *EWS specification* defines a link function that relates different leading indicators to the occurrence of financial crises, while *EWS evaluation* consists in procedures intended to assess the validity of EWS models.

A. EWS Specification

The technical specification of an EWS requires three elements: (i) a crisis dating method, (ii) a set of explanatory variables, and (iii) a link function between these indicators and crisis probabilities. First, no official dating method, similar to the one of economic cycles (in particular the one developed by the *National Bureau of Economic Research* for US business cycles), exists for financial crises. Several methods have hence been proposed for each type of crisis. The *market pressure index* and the *event-based* methods are the most generally used ones (see Jacobs et al., 2004 for a survey). However, this leads to major consequences in terms of EWS modelling, as the dependent variable of the forecasting model is actually the output of another model (the dating method) that is independent of the EWS. In view of its constructed nature, the crisis indicator is characterized by a Markov-type temporal dependence. Failing to take this into account may hence lead to improper statistic inference (Harding and Pagan, 2011).

Second, the choice of the explanatory variables for financial distress should be guided by economic theory (see section 1.1) or by a data-mining approach. Jacobs et al. (2004) examine a vast number of studies with respect to the choice of economic indicators for currency, banking and sovereign-debt crises and their significance in empirical applications. More recently, Frankel and Saravelos (2010) investigate the leading indicators relevant to the 2008-2009 crisis, whilst Alessi and Detken (2011) look into 'real-time' indicators for asset price cycles.

Third, the specification of an EWS requires to define a link function relating leading indicators to the occurrence of financial crises. One of the first EWS was proposed by Kaminski et al. (1998) and relies on a signalling approach. Their study links currency and banking crises, and searches for the origins of twin crises. Using a large set of macroeconomic indicators, and relying on the noise-to-signal ratio (NSR) criterion they find the threshold beyond which a joint crisis is signaled. The forecasting abilities of the EWS thus depend on the cut-off, a crisis being forecasted each time the cut-off is exceeded. This is a pioneer paper both for the study of the determinants of banking and currency crises and for the literature on Early Warning Systems for financial crises.

The discrete-choice EWS (logit / probit) are, nevertheless, the most used ones not only in EWS literature but also in practice (*e.g.* IMF, Federal Reserve, Deutsche Bank, French Banking Commission, Asian Development Bank), despite numerous attempts to propose other parametric and non-parametric models. These models were first introduced in currency crises forecasting literature by Berg and Patilo (1999), who compare the signalling approach with panel probit EWS and conclude in favor of the latter method. These studies paved the way for a large number of papers scrutinizing the occurrence of currency crises. If Kumar et al. (2002) consider logit models, Bussiere and Fratzscher (2006) propose a multinomial logit EWS to account for *post-crisis bias* and Kamin et al. (2007) rely on a probit approach. Tudela (2004) scrutinize the determinants of currency crises in a duration-based framework, while other studies use Markov-Switching models (see Martinez-Peria, 2002; Abiad, 2003; and Fratzcher, 2003), and even neural networks (Peltonen, 2006).

The EWS based on a discrete-choice methodology are very popular in forecasting other types of financial crises too. Demirgüç and Detragiache (1998), Eichengreen and Rose (1998) and Davis and Karim (2008), *inter alii*, use them for banking crises, whilst Detragiache and Spilimbergo (2001), Ciarlone and Trebeschi (2005), Fuertes and Kalotychou (2007), *inter alii*, consider such models when forecasting sovereign-debt crises. Note also that in the aftermath of the global financial crisis, the analysis of the predictability of financial crises soared. Among others, Davis and Karim (2008) analyze logit and binary-tree EWS in the sub-prime crisis context, Phillips and Yu (2011) propose an early warning diagnostic test of bubble activity, while Jorda et al. (2011) study the usefulness of external imbalances indicators to predict financial crises in a logit framework.

EWS Evaluation

EWS evaluation is a vital component of crises prediction, although so far it has not received the attention it deserved. To our knowledge, no EWS evaluation methodology has actually been proposed in financial crises literature to date. Kaminski et al. (1998)'s NSR criterion, *i.e.* threshold discriminating between crisis and calm periods, is one of few exceptions. But this is not enough for an adequate evaluation of EWS. First, the NSR criterion does not arbitrate between type I and type II errors. It cannot hence lead to an optimal discrimination of crisis and calm periods. Second, the relative accuracy of an EWS should be assessed by relying on comparison tests. Simple Quadratic Probability Score (QPS)-type evaluation criteria, sometimes considered in the literature, do not allow us to directly gauge the statistical significance of the difference between two models.

These findings are puzzling, as a large number of papers on forecast evaluation have been proposed in the forecasting literature. The econometric developments in non-linear modelling over the last decade have led to an increasing volume of research dedicated to the construction and evaluation of point forecasts, interval forecasts and density forecasts issued from non-linear models (see Terasvirta, 2006; West, 2006; Clark and McCracken, 2011, for some surveys). In particular, we distinguish between absolute evaluation and relative evaluation (comparison) methods. Point-forecasts issued from alternative (nested or non-nested) models are generally compared by using evaluation tests that rely on a loss-function associated with the sequences of forecasts (*e.g.* Diebold and Mariano, 1995; Harvey et al., 1997; Clark and McCracken, 2001; Clark and West, 2007). By contrast, absolute evaluation is based on MSFE, MAFE, etc.,- type criteria. More recently another research direction, focused on the absolute and relative evaluation of interval and density forecasts, has emerged in this literature. Christoffersen (1998) defines the hypotheses allowing to assess the validity of an interval forecast obtained by using any type of model (linear or nonlinear) and proposes LR-type testing strategies. Besides, various correct specification tests and density comparison tests have been developed (*e.g.* Bao et al. 2004, 2007; Corradi and Swanson, 2006b). An appealing feature of these tests is that they allow us to compare density forecasts issued from potentially misspecified models.

But is there a way to explain this gap between the lack of interest in forecast evaluation noticed in EWS literature and the rich econometric literature on forecasting with non-linear models? Why is EWS evaluation so particular? Several differences between EWS and the models generally considered in forecasting literature can be identified. First, the event to be forecasted, *i.e.* crises occurrence, is not directly observable. Indeed, it is the output of a crisis-dating model, as previously discussed. Second, the output of these models consists in a series of probabilities, whereas the usual output of a forecasting model takes the form of a continuous predictor. Although these differences seem important, they cannot entirely explain the lack of interest in EWS evaluation. Indeed, EWS are not the only forecasting models handling an unobserved dependent variable. It is also, for example, the case of Value-at-Risk and volatility forecasts in financial econometrics, where adequate, robust evaluation methods have been proposed (Patton, 2011). Similarly, Markov-switching models output probabilities without disrupting forecast evaluation.

This paradox is even more astounding as evaluation methods specifically adapted to probability forecasts exist (*e.g.* credit-scoring analysis). Considering such methods in EWS evaluation would hence be important both from the point of view of absolute accuracy analysis (as measure of how good an approximation the predictions are for their realizations) and relative accuracy analysis (as comparative assessment of several alternative models). A rigorous selection of the best EWS and an optimal identification

method of future crises and calm periods, based on the probabilities issued from the outperforming EWS, is hence necessary to successfully transpose these methods to crises predictability.

1.3 Contribution

In a similar vein to the Harvard barometer (1919, 1924), this research focuses on EWS models forecasting the arrival of financial turmoils. The broad goal of this dissertation is hence (i) to propose a systematic evaluation methodology of the forecasting abilities of EWS as well as (ii) to introduce new EWS models with improved forecasting abilities. This work has been concretized so far in four chapters (articles) that can be studied independently of one another.

The first two chapters fill in the gap existing in forecast evaluation, in particular for EWS, which is a new area of research. Chapter 2 proposes a new and general test to validate interval forecasts, particularly useful for non-linear models. In chapter 3, a toolbox specifically designed for the evaluation of EWS is developed. Note the importance of this toolbox, since it is useful not only for financial crises EWS, but also to assess the validity of business cycles forecasting models and, more generally, that of any model returning a sequence of forecasted probabilities.

The other two chapters propose improvements of EWS specifications for the classic logit/probit methods by considering the case of currency crises. In this context, crises predictability signifies the correct identification (in- and out-of-sample) of historical crisis and calm periods. Chapter 4 introduces dynamic EWS specifications, whereas chapter 5 proposes a multivariate EWS, jointly modelling several types of financial crises. In the following, we detail the contents of each chapter.

Chapter 2: Testing Interval Forecasts: a GMM-Based Approach

Chapter 2, “Testing Interval Forecasts: a GMM-Based Approach”², forthcoming in *Journal of Forecasting*, proposes an original test to evaluate the forecasting performance of non-linear models by assessing the validity of interval forecasts and High Density Regions (HDR). Even though interval forecast is the most generally used method by applied economists to account for forecast uncertainty, there are only few studies proposing validation methods adapted to these kind of forecasts (*e.g.* Christoffersen, 1998).

To test the interval forecasts and HDR validity, we hence develop an original test based on a simple J-statistic (Hansen, 1982) using particular moments defined by orthonormal polynomials associated with the binomial distribution (Bontemps, 2006 and Bontemps

²Based on Dumitrescu, Hurlin and Madkour (2011), “Testing Interval Forecasts: a GMM-Based Approach”, forthcoming in *Journal of Forecasting*.

and Meddahi, 2005, 2011). This model free test can be applied to interval forecasts or HDR obtained from any type of model (linear or non-linear).

The test relies on the concept of violation (Christoffersen, 1998) defined as a situation where the ex-post realization does not belong to the confidence interval or HDR defined ex-ante. An original approach that transforms the violation series (a violation is said to occur if the *ex-post* realization of the variable does not lie in the *ex-ante* forecast interval) into a series of sums of violations defined for H blocks of size N is introduced. Under the null hypothesis of interval forecasts validity, these sums are binomial distributed. Assessing the validity of interval forecasts hence comes down to testing the distributional (binomial) assumption for the violation process.

Our GMM-test performs well in finite samples of relatively small size since the finite-sample distribution of the J -statistic is close to its chi-squared distribution regardless of the block size chosen. If the block size is small, the number of blocks is large, and the statistic is close to its asymptotic distribution. For large block sizes, the number of blocks is small and the convergence in distribution does not hold. However, in this case, given the properties of the binomial distribution, each sum of violations follows a normal distribution. The test-statistic, which is a sum of squared normal variates, has a chi-squared distribution, as in the asymptotic case.

Several advantages of this approach can be identified. First, the three hypotheses of conditional coverage, independence and unconditional coverage can be tested independently in a unified framework. Second, no restrictions are imposed under the alternative hypothesis. Third, this evaluation test is simple to implement and always feasible. Most importantly, Monte-Carlo experiments show that for realistic sample sizes our GMM test has very good power properties, better than those of Christoffersen (1998)'s test. An empirical application on stock market indexes confirms our findings.

Chapter 3: How to Evaluate an Early Warning System?

Chapter 3, “How to Evaluate an Early Warning System? Towards a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods”³, forthcoming in *IMF Economic Review*, proposes an original and unified toolbox precisely designed for the evaluation of EWS. As aforementioned, the validation of such models has not been tackled in EWS literature to date even though it is essential for crisis forecast. Previous studies used QPS as evaluation criterion and relied on Kaminski et al. (1998)'s NSR cut-off to discriminate between the crisis and calm periods. Furthermore, no statistical inference was provided to identify the outperforming model.

³Based on Candelon, Dumitrescu and Hurlin (2012), “How to Evaluate an Early Warning System? Towards a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods”, *IMF Economic Review* 60(1).

We fill in this gap by proposing a model-free original evaluation framework. It can be applied to any type of EWS model, for any type of crisis, both in-sample and out-of-sample. It is a two-step procedure. First, we identify of the optimal cut-off, *i.e.* threshold that best discriminates between crisis and calm periods. For this, we propose an original method relying on the concepts of *sensitivity* and *specificity*. We show that this method, which balances type I and type II errors performs better than existing ones, *i.e.* NSR and arbitrarily chosen cut-off.

Second, we compare alternative models by using different validation criteria and tests. We hence propose several validation criteria (ROC curve) and different tests (the Area under the ROC curve) by elaborating on the credit-scoring literature (*i.e.* individual risk forecasting). We provide both theoretical and empirical evidence that comparison tests are essential to identify the optimal EWS. In this chapter we thus argue that an adequate EWS evaluation should take into account the cut-off both in the optimal crisis forecast step (by relying on *sensitivity* and *specificity* concepts to identify the optimal cut-off) and in the model comparison step (by using the Area Under the ROC Curve test, AUC). More insights about the usefulness of different EWS specifications can be hence obtained.

We illustrate the importance of this methodology in an empirical application on 12 emerging markets gauging the importance of the yield spread in forecasting currency crises. Our main finding is that the yield spread is an important indicator of currency crises in half of the countries when relying on the AUC test, whereas it first seems to be essential in all countries. Besides, the optimal cut-off correctly identifies more than 2/3 of the crisis and calm periods, in contrast with the NSR threshold that does not succeed to identify most of the crisis periods.

Chapter 4: Currency Crises Early Warning Systems: Why They Should Be Dynamic?

Chapter 4, “Currency Crises Early Warning Systems: Why They Should Be Dynamic”⁴, provides evidence that dynamics, *i.e.* persistence, is important in forecasting crises. As Berg and Coke (2004) showed, EWS are per nature autoregressive, property that is hardly reproduced by static models. This chapter hence proposes a new generation of EWS that combines the discrete-choice character of the crisis indicator and the dynamic dimension of this phenomenon. The endogenous dynamics of crises can be pictured in different ways. It can be included through the lagged binary crisis variable. In this case, there is a non-linear transmission of the crisis from one period to the next, as the index must go beyond a threshold to set of a crisis. Another way, consisting in

⁴Based on Candelon, Dumitrescu and Hurlin (2010), “Currency Crises Early Warning Systems: Why They Should Be Dynamic?”, METEOR Research memorandum RM/10/047.

an autoregressive model for the crisis index can be considered. Finally, both types of dynamics can be jointly considered in a dynamic binary model.

In this context, we propose the first EWS based on a dynamic discrete-choice model estimated by an exact maximum-likelihood method (proposed by Kauppi and Saikonen, 2008) and perform specification tests in a unified framework. Subsequently, we adapt this methodology to panel by drawing on the works of Carro (2007). Easy to implement for any type of crisis, our EWS also takes into account macroeconomic leading indicators, source of exogenous crisis persistence.

An empirical application for currency crises on 15 emerging countries scrutinizes the forecasting abilities of this new EWS with respect to those of its main competitors, *i.e.* Markov-Switching and static logit models. For this, we rely on the evaluation toolbox introduced in chapter 3. Our first finding is that dynamic logit models consistently outperform static ones as well as Markov-Switching both in-sample and out-of-sample. Second, dynamic EWS deliver good out-of-sample forecasting probabilities, correctly identifying most of the out-of-sample crisis and calm periods for most of the countries. We hence argue that dynamics constitutes a key characteristic that delivers more adequate signals to prevent financial turmoils. It should hence be taken into account more often in the quest for optimal EWS.

Chapter 5: Modelling Financial Crises Mutation

Chapter 5, “Modelling Financial Crises Mutation”⁵, proposes a multivariate dynamic EWS that encompasses the main three types of financial crises, *i.e.* banking, currency and sovereign debt, and allows us to investigate the potential causality between them. In view of the development of recent crises in Latin America and Europe, it appears that one or two perspectives do not provide an exhaustive picture of the debacle. This article is hence a multivariate extension of papers like Glick and Hutchinson (1999) that investigate twin crises. For this, a methodological novelty which consists in an exact maximum-likelihood approach to estimate our multivariate dynamic probit model is proposed. We hence extend Huguenin et al. (2009) to dynamic models and Kauppi and Saikonen (2008) to multivariate models.

The empirical application shows that while bivariate causality from banking to currency (and vice-versa) are common, the trivariate model outperforms the bivariate one for countries that know the three types of crises. Most importantly, this method allows us to disentangle the different reasons why a specific crisis mutates to another one: this can be due to common shocks (as in South Africa) or to a strong causal structure (as in Ecuador). The conditional probability plots and the impulse-response function analysis

⁵Based on Candelon, Dumitrescu, Hurlin and Palm (2011), “Modelling Financial Crises Mutation”, DR LEO 2011-17.

support these findings by emphasizing the diffusion mechanisms of the three types of crises. The possible mutation of a crisis, should hence be taken into account to develop an efficient EWS, in particular through the implementation of trivariate models whenever it is feasible.

Finally, chapter 6 summarizes the main findings of this thesis and puts forward several objectives for future research.

Chapter 2

Testing Interval Forecasts: a GMM-Based Approach¹

This paper proposes a new evaluation framework for interval forecasts. Our model free test can be used to evaluate intervals forecasts and high density regions, potentially discontinuous and/or asymmetric. Using a simple J-statistic based on the moments defined by the orthonormal polynomials associated with the binomial distribution, this new approach presents many advantages. First, its implementation is extremely easy. Second, it allows for a separate test for unconditional coverage, independence and conditional coverage hypotheses. Third, Monte-Carlo simulations show that for realistic sample sizes our GMM test has good small-sample properties. These results are corroborated by an empirical application on SP500 and Nikkei stock market indexes. It confirms that using this GMM test leads to major consequences for the ex-post evaluation of interval forecasts produced by linear versus nonlinear models.

2.1 Introduction

In recent years, the contribution of nonlinear models to forecasting macroeconomic and financial series has been intensively debated (see Teräsvirta, 2006; and Colletaz and Hurlin, 2005, for a survey). As suggested by Teräsvirta, there are relatively numerous studies in which the forecasting performance of nonlinear models is compared with that of linear models using actual series. In general, no dominant nonlinear (or linear) model has emerged. However, the use of nonlinear models has actually led to the renewal of the forecasting approach, especially through the emergence of concepts like high density regions (Hyndman, 1995) or density forecasts as opposed to point forecasts. Consequently, this debate on non-linearity and forecasting involves new forecast validation criteria. In

¹This chapter is based on Dumitrescu, Hurlin and Madkour (2011), forthcoming in *Journal of Forecasting*.

the case of density forecasts, many specific evaluation tests have been developed (*e.g.* Bao *et al.* 2004, 2007; Corradi and Swanson, 2006a).

Conversely, if there are numerous methods to calculate HDR and interval forecasts (Chatfield, 1993), only a few studies propose validation methods adapted to these kinds of forecasts. This paradox is even more astounding if we take into consideration the fact that interval forecast is the most generally used method by applied economists to account for forecast uncertainty.

One of the main exceptions is the seminal paper of Christoffersen (1998), which introduces general definitions of hypotheses allowing to assess the validity of an interval forecast obtained by using any type of model (linear or nonlinear). His model-free approach is based on the concept of violation: a violation is said to occur if the *ex-post* realization of the variable does not lie in the *ex-ante* forecast interval. Three validity hypothesis are then distinguished. The *unconditional coverage* hypothesis means that the expected frequency of violations is precisely equal to the coverage rate of the interval forecast. The *independence hypothesis* means that if the interval forecast is valid then violations must be distributed independently. In other words, there must not be any cluster in the violations sequence. Finally, under the *conditional coverage hypothesis* the violation process satisfies the assumptions of a martingale difference. Based on these definitions, Christoffersen proposes a Likelihood Ratio (LR) test for each of these hypotheses, by considering a binary first-order Markov chain representation under the alternative hypothesis.

More recently, Clements and Taylor (2002) applied a simple logistic regression with periodic dummies and modified the first-order Markov chain approach in order to detect dependence at a periodic lag. In 2003, Wallis recast Christoffersen's tests (1998) in the framework of contingency tables increasing users' accessibility to these interval forecast evaluation methods. Owing to his innovative approach, it became possible to calculate exact *p*-values for the LR statistics in small-sample cases.

Beyond their specificities, the main common characteristic of these tests is that assessing the validity of interval forecasts comes down to testing a distributional assumption for the violation process. If we define a binary indicator variable that takes the value one in case of violation and zero otherwise, it is obvious that under the null of conditional coverage the sum of the indicators associated to a sequence of interval forecasts follows a Binomial distribution.

On these grounds, we propose in this paper a generalized method of moments (GMM) approach to test the interval forecasts and HDR validity. To be more precise, we propose to test interval forecasts using discrete polynomials. The series of violations, I_t , (a violation indicates whether the forecast belongs to the $1 - \alpha$ confidence interval or not) is split

into blocks of size N . The sum of I_t within each block follows a binomial distribution $B(N, \alpha)$. The test involves testing that the series of sums is indeed a *i.i.d.* sequence of random variables which are binomially distributed. Relying on the GMM framework of Bontemps and Meddahi (2005), we propose simple J -statistics based on particular moments defined by the orthonormal polynomials associated with the Binomial distribution. A similar approach has been used by Candelon *et al.* (2011) in the context of the Value-at-Risk (VaR)² backtesting. The authors test the VaR forecasts validity by testing the geometric distribution assumption for the durations observed between two consecutive VaR violations. Here, we propose a general approach for all kinds of intervals and HDR forecasts that directly exploits the properties of the violation process (and not the durations between violations). We adapt the GMM framework to the case of discrete distributions and more exactly to a binomial distribution.

Our approach has several advantages. First, we develop an unified framework in which the three hypotheses of unconditional coverage, independence and conditional coverage are tested independently. Second, this approach imposes no restrictions under the alternative hypothesis. Third, this GMM-based test is easy to implement and does not generate computational problems regardless of the sample size. Finally, some Monte-Carlo simulations indicate that for realistic sample sizes, our GMM test has good power properties.

The paper is structured as follows. Section 2.2 presents the general framework of interval forecast evaluation, while section 2.3 introduces our new GMM-based evaluation tests. In section 2.4 we scrutinize the finite-sample properties of the tests through Monte-Carlo simulations and in section 2.5 we propose an empirical application. Section 2.6 concludes.

2.2 General Framework

Formally, let $x_t, t \in \{1, \dots, T\}$ be a sample path of a time series x_t . Let us denote by $\{C_{t|t-1}(\alpha)\}_{t=1}^T$ the sequence of *out-of-sample* interval forecasts for the coverage probability α , so that

$$\Pr[x_t \in C_{t|t-1}(\alpha)] = 1 - \alpha. \quad (2.1)$$

Hyndman (1995) identifies three methods to construct a $100(1-\alpha)\%$ forecast region: (i) a symmetrical interval around the point forecast, (ii) an interval defined by the $\alpha/2$ and $(1 - \alpha/2)$ quantiles of the forecast distribution, (i) and a High Density Region (*HDR*). These three forecast regions are identical (symmetric and continuous) in the case of symmetric and unimodal distribution. By contrast, HDR_α is the smallest forecast region for asymmetric or multimodal distributions. When the interval forecast is continuous,

²Note that the Value-at-Risk can be interpreted as a one-sided and open interval.

$C_{t|t-1}(\alpha)$ can be defined as in Christoffersen (1998), by $C_{t|t-1}(\alpha) = [L_{t|t-1}(\alpha), U_{t|t-1}(\alpha)]$, where $L_{t|t-1}(\alpha)$ and $U_{t|t-1}(\alpha)$ are the limits of the *ex-ante* confidence interval for the coverage rate α .

Whatever the form of the HDR or the interval forecasts (symmetric or asymmetric, continuous or discontinuous), we define an indicator variable $I_t(\alpha)$, also called violation, as a binary variable that takes the value one if the realization of x_t does not belong to this region:

$$I_t(\alpha) = \begin{cases} 1, & x_t \notin C_{t|t-1}(\alpha) \\ 0, & x_t \in C_{t|t-1}(\alpha) \end{cases}. \quad (2.2)$$

Based on the definition of the violations process, a general testing criterion for interval forecasts can be established. Indeed, as stressed by Christoffersen (1998), the interval forecasts are valid if and only if the conditional coverage (CC hereafter) hypothesis is fulfilled, implying that both the independence (IND hereafter) and unconditional coverage (UC hereafter) hypotheses are satisfied. Under the UC assumption, the probability to have a violation must be equal to the α coverage rate:

$$H_{0,UC} : \Pr[I_t(\alpha) = 1] = \mathbb{E}[I_t(\alpha)] = \alpha. \quad (2.3)$$

Under the IND hypothesis, violations observed at different moments in time for the same coverage rate ($\alpha\%$) must be independent. In other words, we do not observe any clusters of violations and past violations should not be informative about the present or future violations. The UC property places a restriction on how often violations may occur, whereas the IND assumption restricts the order in which these violations may appear.

Christoffersen (1998) pointed out that in the presence of higher-order dynamics it is important to go beyond the UC assumption and test the CC hypothesis. Under the CC assumption, the conditional (on a past information set Ω_{t-1}) probability to observe a violation must be equal to the α coverage rate, *i.e.* the I_t process satisfies the properties of a martingale difference:

$$H_{0,CC} : \mathbb{E}[I_t(\alpha) \mid \Omega_{t-1}] = \alpha. \quad (2.4)$$

Christoffersen considers an information set Ω_{t-1} that consists of past realizations of the indicator sequence $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, \dots, I_1\}$. In this case, testing $\mathbb{E}[I_t(\alpha) \mid \Omega_{t-1}] = \alpha$ for all t is equivalent to testing that the sequence $\{I_t(\alpha)\}_{t=1}^T$ is identically and independently distributed Bernoulli with parameter α . So, a sequence of interval/HDR forecasts $\{C_{t|t-1}(\alpha)\}_{t=1}^T$ has correct conditional coverage, if

$$I_t \stackrel{i.i.d}{\sim} \text{Bernoulli}(\alpha), \quad \forall t. \quad (2.5)$$

This feature of the violation process is actually at the core of most of the interval forecast evaluation tests (Christoffersen, 1998, Clements and Taylor, 2002, etc.) and so it is for our GMM-based test.

2.3 A GMM-Based Test

In this paper we propose a unified GMM framework for evaluating interval forecasts and HDR by testing the Bernoulli distributional assumption of the violation series $I_t(\alpha)$. Our analysis is based on the recent GMM distributional testing framework developed by Bontemps and Meddahi (2005) and Bontemps (2006). We first present the environment of the test, then we define the moment conditions used to test the interval forecasts efficiency, and finally we propose simple J -statistics corresponding to the three hypotheses of UC, IND and CC.

2.3.1 Environment Testing

Given the result in eq. 2.5, it is obvious that if the interval forecast has a correct conditional coverage, the sum of violations follows a Binomial distribution:

$$H_{0,CC} : \sum_{t=1}^T I_t(\alpha) \sim B(T, \alpha). \quad (2.6)$$

A natural way to test CC involves testing this distributional assumption. However, this property cannot be directly used to develop an implementable testing procedure, since, for a given sequence $\{C_{t|t-1}(\alpha)\}_{t=1}^T$, we have only one observation for the sum of violations.

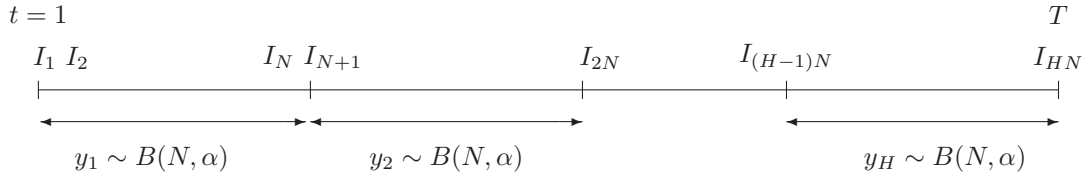


Figure 2.1: Partial sums y_h and block size N

Therefore, we propose to divide the sample of violations into blocks. Since under the null hypothesis the violations $\{I_t(\alpha)\}_{t=1}^T$ are independent, it is possible to split the initial series of violations into H blocks of size N , where $H = \lceil T/N \rceil$ (see Figure 2.1). The sum of I_t within each block follows a binomial distribution $B(N, \alpha)$. More formally, for each block, we define y_h , $h \in \{1, \dots, H\}$ as the sum of the corresponding N violations:

$$y_h = \sum_{t=(h-1)N+1}^{hN} I_t(\alpha). \quad (2.7)$$

As a result, under the null hypothesis, the constructed processes y_h are *i.i.d.* $B(N, \alpha)$, and thus the null of CC that the interval forecasts are well specified can simply be expressed

as follows:

$$H_{0,CC} : y_h \sim B(N, \alpha), \quad \forall h \in \{1, \dots, H\}. \quad (2.8)$$

This approach can be compared to the sub-sampling methodology proposed by Politis, Romano and Wolf, (1999). However, the objective here is entirely different. In our case, we do not aim to obtain the finite sample distribution of a particular test statistic. We only divide the initial sample of T violations into H blocks of size N in order to compute our CC test (which is a simple distributional test). In other words, we choose the distributional assumption that we want to test. In order to test the CC assumption, we propose to test the $B(N, \alpha)$ distribution rather than the $B(T, \alpha)$ one, even if theoretically both approaches are possible. The advantages of this approach will be presented in the next sections, in the specific context of the test that we propose.

2.3.2 Orthonormal Polynomials and Moment Conditions

There are many ways to test conditional coverage hypothesis through the distributional assumption in eq. 2.8. Following Bontemps and Meddahi (2005) and Bontemps (2006), we propose here to use a GMM-based framework. The general idea is that for many continuous and discrete distributions it is possible to associate some particular orthonormal polynomials whose expectation is equal to zero. These orthonormal polynomials can be used as moment conditions in a GMM framework to test for a specific distributional assumption. For instance, the Hermite polynomials associated with the normal distribution can be employed to build a test for normality (Bontemps and Meddahi, 2005). Other particular polynomials are used by Candelon et al. to test for a geometric distribution hypothesis.

In the particular case of a Binomial distribution, the corresponding orthonormal polynomials are called Krawtchouk polynomials. These polynomials are defined as follows:

Definition 1. *Let us consider a discrete random variable y_h such that $y_h \sim B(N, \alpha)$. The corresponding orthonormal Krawtchouk polynomials are defined by the following recursive relationship:*

$$P_{j+1}^{(N,\alpha)}(y_h) = \frac{\alpha(N-j) + (1-\alpha)j - y_h}{\sqrt{\alpha(1-\alpha)(N-j)(j+1)}} P_j^{(N,\alpha)}(y_h) - \sqrt{\frac{j(N-j+1)}{(j+1)(N-j)}} P_{j-1}^{(N,\alpha)}(y_h),$$

where $j < N$ and $P_{-1}^{(N,\alpha)}(y_h) = 0$, $P_0^{(N,\alpha)}(y_h) = 1$ verify

$$\mathbb{E} [P_j^{(N,\alpha)}(y_h)] = 0, \quad \forall j < N. \quad (2.9)$$

Our test exploits these moment conditions. More precisely, let us define $\{y_1; \dots; y_H\}$ a sequence of sums defined by eq. 2.7 and computed from the sequence of violations

$\{I_t(\alpha)\}_{t=1}^T$. Under the null of conditional coverage, variables y_h are *i.i.d.* and have a Binomial distribution $B(N, \alpha)$, where N is the block size. Hence, the null of CC can be expressed as follows:

$$H_{0,CC} : \mathbb{E} [P_j^{(N,\alpha)}(y_h)] = 0, \quad j = \{1, \dots, m\}, \quad (2.10)$$

with the number of moment conditions $m < N$. The expressions of the first two polynomials are the following:

$$P_1^{(N,\alpha)}(y_h) = \frac{\alpha N - y_h}{\sqrt{\alpha(1-\alpha)N}}, \quad (2.11)$$

$$P_2^{(N,\alpha)}(y_h) = \left(\frac{\alpha(N-1) + (1-\alpha) - y_h}{\sqrt{\alpha(1-\alpha)2(N-1)}} \right) \left(\frac{\alpha N - y_h}{\sqrt{\alpha(1-\alpha)N}} \right) - \sqrt{\frac{N}{2(N-1)}}. \quad (2.12)$$

An appealing property of the test is that it allows one to test separately for the UC and IND hypotheses. Let us remind that under the UC assumption the unconditional probability to have a violation is equal to the coverage rate α . Consequently, under UC the expectation of the sum y_h is then equal to αN , since:

$$\mathbb{E}(y_h) = \sum_{t=(h-1)N+1}^{hN} \mathbb{E}[I_t(\alpha)] = \alpha N, \quad \forall h \in \{1, \dots, H\}. \quad (2.13)$$

Given the properties of the Krawtchouk polynomials, the null UC hypothesis can be expressed as:

$$H_{0,UC} : \mathbb{E} [P_1^{(N,\alpha)}(y_h)] = 0. \quad (2.14)$$

In this case, we need to use only the first moment condition defined by $P_1^{(N,\alpha)}(y_h)$, since the condition $\mathbb{E} [P_1^{(N,\alpha)}(y_h)] = 0$ is equivalent to the UC condition $\mathbb{E}(y_h) = \alpha N$ or $\mathbb{E}(I_t) = \alpha$.

Under the IND hypothesis, the violations are independently and identically distributed, but their probability is not necessarily equal to the coverage rate α . Let us denote β the violation probability. If the violations are independent, the sum y_h follows a $B(N, \beta)$ distribution, where β may be different from α . Thus, the IND hypothesis can simply be expressed as:

$$H_{0,IND} : \mathbb{E} [P_j^{(N,\beta)}(y_h)] = 0 \quad j = \{1, \dots, m\}, \quad (2.15)$$

with $m < N$.

2.3.3 Testing Procedure

Let $P^{(N,\alpha)}$ denote a $(m, 1)$ vector whose components are the orthonormal polynomials $P_j^{(N,\alpha)}(y_h)$, for $j = 1, \dots, m$, associated with the Binomial distribution $B(N, \alpha)$. Under

the CC assumption and some regularity conditions (Hansen, 1982), it can be shown that:

$$\left(\frac{1}{\sqrt{H}} \sum_{h=1}^H P^{(N,\alpha)}(y_h) \right)' \Sigma^{-1} \left(\frac{1}{\sqrt{H}} \sum_{h=1}^H P^{(N,\alpha)}(y_h) \right) \xrightarrow{H \rightarrow \infty} \chi^2(m), \quad (2.16)$$

where Σ is the long-run covariance matrix of $P^{(N,\alpha)}(y_h)$. By the definition of orthonormal polynomials, this long-run covariance matrix corresponds to the identity matrix.³ Therefore, the corresponding J -statistic is very easy to implement. Let us denote by $J_{CC}(m)$ the CC test-statistic associated with the $(m, 1)$ vector of orthonormal polynomials $P^{(N,\alpha)}(y_h)$.

Definition 2. *Under the null hypothesis of conditional coverage, the CC test statistic verifies:*

$$J_{CC}(m) = \frac{1}{H} \sum_{j=1}^m \left(\sum_{h=1}^H P_j^{(N,\alpha)}(y_h) \right)^2 \xrightarrow{H \rightarrow \infty} \chi^2(m), \quad (2.17)$$

where $P_j^{(N,\alpha)}(y_h)$ denotes the Krawtchouk polynomial corresponding to a Binomial distribution $B(N, \alpha)$ of order j , for $j \leq m$.

Proof. see appendix 2.7.1. Since the $J_{UC}(m)$ statistic corresponding to the UC hypothesis is a special case of the $J_{CC}(m)$ test statistic, it can be immediately computed by taking into account only the first moment condition $\mathbb{E} [P_1^{(N,\alpha)}(y_h)] = 0$, and can be expressed as follows:

$$J_{UC} = J_{CC}(1) = \frac{1}{H} \left(\sum_{h=1}^H P_1^{(N,\alpha)}(y_h) \right)^2 \xrightarrow{H \rightarrow \infty} \chi^2(1), \quad (2.18)$$

Finally, the independence hypothesis statistic, denoted by $J_{IND}(m)$ takes the form of:

$$J_{IND}(m) = \frac{1}{H} \sum_{j=1}^m \left(\sum_{h=1}^H P_j^{(N,\beta)}(y_h) \right)^2 \xrightarrow{H \rightarrow \infty} \chi^2(m), \quad (2.19)$$

where $P_j^{(N,\beta)}(y_h)$ is the orthonormal polynomial of order $j \leq m$ associated with a Binomial distribution $B(N, \beta)$, where β can be different from α . The coverage rate β is generally unknown, and thus it has to be estimated. When using a consistent estimator $\hat{\beta} = (1/T) \sum_t I_t(\alpha)$ instead of β , the degree of freedom of the GMM-statistic $J_{IND}(m)$ has to be adjusted accordingly:

$$J_{IND}(m) = \frac{1}{H} \sum_{j=1}^m \left(\sum_{h=1}^H P_j^{(N,\hat{\beta})}(y_h) \right)^2 \xrightarrow{H \rightarrow \infty} \chi^2(m-1), \quad (2.20)$$

³If we neglect this property, it is also possible to use the Kernel estimate of the long-run logit matrix, as it is usually done in the GMM literature.

where $P_j^{(N,\hat{\beta})}(y_h)$ is the orthonormal polynomial of order j associated to a Binomial distribution $B(N, \hat{\beta})$ and $\hat{\beta}$ is the estimated coverage rate.

Our block-based approach has many advantages for testing the validity of interval forecasts, especially in finite samples with relatively small size (as it will be shown in the Monte Carlo simulation section). First, let us consider without loss of generality the case of two moment conditions. The test statistic $J_{CC}(2)$ based on $P_1^{(N,\alpha)}(y_h)$ and $P_2^{(N,\alpha)}(y_h)$ can be viewed as a function of both y_h and y_h^2 which, once expanded, involves the cross product $I_t(\alpha)I_s(\alpha)$ for two periods t and s within a given block. When the block size N is equal to 2, $J_{CC}(2)$ is close to the joint test of Christoffersen. When $N = 3$, the test statistic involves the product (*i.e.* correlation) between $I_{t-2}(\alpha)$, $I_{t-1}(\alpha)$ and $I_t(\alpha)$ and more generally, for any N , it includes the correlations between $I_{t-h}(\alpha)$ for $h = 1, \dots, N$ and $I_t(\alpha)$. Consequently, we expect that our approach will reveal some dependencies that cannot be identified by Christoffersen's approach.

Second, when the block size N is small, H is relatively important, and many observations of the sums y_h are available. The finite sample distribution of the J -statistic is then close to its asymptotic chi-squared distribution. On the contrary, when N is large compared to T , the binomial distribution $B(N, \alpha)$ can be approximated by a normal distribution. Then, each sum y_h has also a normal distribution and their sum of squares has a chi-squared distribution. Consequently, as we will show in the next section, the J -statistics have a finite chi-squared distribution.

2.4 Monte-Carlo Experiments

In this section we gauge the finite sample properties of our GMM-based test using Monte-Carlo experiments. We first analyze the size performance of the test and then investigate its empirical power in the same framework as in Berkowitz et al. (2010). A comparison with Christoffersen (1998)'s LR tests is provided for both analyses. In order to control for size distortions, we use Dufour (2006)'s Monte-Carlo method.

2.4.1 Empirical Size Analysis

To illustrate the size performance of our UC and CC tests in finite samples, we generate a sequence of T violations by taking independent draws from a Bernoulli distribution, considering successively a coverage rate $\alpha = 1\%$ and $\alpha = 5\%$. Several sample sizes T ranging from 250 (which roughly corresponds to one year of daily forecasts) to 1,500 are considered. The size of the blocks (used to compute the H sums y_h) is fixed to $N = 100$ or $N = 25$ observations. Additionally, we consider several moment conditions m from 1 (for the UC test statistic J_{UC}) to 5. Based on a sequence $\{y_h\}_{h=1}^H$, with $H = [T/N]$, we compute both statistics J_{UC} and $J_{CC}(m)$. The reported empirical sizes correspond to the rejection rates calculated over 10,000 simulations for a nominal size equal to 5%.

Table 2.1: Empirical size (block size $N = 100$, nominal size 5%)

Coverage rate 5%							
T	H	J_{UC}	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	LR_{UC}	LR_{CC}
250	2	0.0316	0.0643	0.0499	0.0442	0.0587	0.0404
500	5	0.0521	0.0556	0.0615	0.0662	0.0544	0.0443
750	7	0.0409	0.0513	0.0595	0.0734	0.0503	0.0462
1000	10	0.0487	0.0535	0.0614	0.0655	0.0503	0.0565
1250	12	0.0522	0.0490	0.0543	0.0577	0.0417	0.0781
1500	15	0.0489	0.0479	0.0596	0.0577	0.0489	0.0656
Coverage rate 1%							
T	H	J_{UC}	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	LR_{UC}	LR_{CC}
250	2	0.0516	0.0397	0.0397	0.0397	0.0132	0.0112
500	5	0.0314	0.0397	0.0360	0.0383	0.0640	0.0113
750	7	0.0330	0.0543	0.0456	0.0425	0.0384	0.0220
1000	10	0.0361	0.0482	0.0548	0.0473	0.0572	0.0251
1250	12	0.0575	0.0517	0.0592	0.0575	0.0627	0.0286
1500	15	0.0487	0.0518	0.0489	0.0414	0.0541	0.0312

Note: Under the null hypothesis, the violations are i.i.d. and follows a Bernoulli distribution. The results are based on 10,000 replications. For each sample, we provide the percentage of rejection at a 5% level. $J_{CC}(m)$ denotes the GMM based conditional coverage test with m moment conditions. J_{UC} denotes the unconditional coverage test obtained for $m=1$. LR_{CC} (resp. LR_{UC}) denotes the Christoffersen's conditional (resp. unconditional) coverage test. T denotes the sample size of the sequence of interval forecasts violations I_t , while $H=[T/N]$ denotes the number of block (size $N=100$) used to define the sums (y_h) of violations.

In Table 2.1, the rejection frequencies for the $J_{CC}(m)$ statistic and a block size N equal to 100 are presented. For comparison reasons, the rejection frequencies for the Christoffersen (1998)'s LR_{UC} and LR_{CC} test statistics are also reported. For a 5% coverage rate and whatever the choice of m , the empirical size of the J_{CC} test is close to the nominal size, even for small sample sizes. For a 1% coverage, the J_{CC} test is also well sized, whereas the LR_{CC} test seems to be undersized in small samples (especially for $\alpha = 1\%$), although its size converges to the nominal one as T increases.⁴ On the contrary, the performance of our J_{UC} statistic and the LR_{UC} are quite comparable (especially for $T \geq 500$). It can be proved that J_{UC} is a local expansion of the unconditional test of Christoffersen. Indeed, our statistic can be expressed as a simple function of the sample size and the total number of hits $\sum_{h=1}^H y_h$, since:

$$J_{UC} = \frac{1}{H} \left(\sum_{h=1}^H \frac{y_h - N\alpha}{\sqrt{\alpha(1-\alpha)N}} \right)^2 = \frac{T}{\alpha(1-\alpha)} \left(\alpha - \frac{1}{T} \sum_{h=1}^H y_h \right)^2. \quad (2.21)$$

⁴Berkowitz et al. (2010) and Candelon et al. (2011) found that the LR_{CC} is oversized in small sample. The difference comes from the computation of the LR statistic. Under H_1 , the computation of the LR_{CC} statistic requires calculating the sum of joint violations $I_t(\alpha)$ and $I_{t-1}(\alpha)$. Consequently, the size of the available sample is equal to $T-1$. On the contrary, under H_0 , the likelihood depends on the sample size and the coverage rate α . Contrary to previous studies, we compute the likelihood under H_0 by adjusting the sample size to $T-1$. This slight difference explains the differences in the results. By considering a sample size T under H_0 , we get exactly the same empirical size as in Berkowitz and al. (2010) or Candelon et al. (2011).

The performance of our test is quite remarkable, since under the null, in a sample with $T = 250$ and a coverage rate equal to 1%, the expected number of violations lies between two and three. It is worth noting that even if our asymptotic result requires that the number of blocks H tends to infinity, our testing procedure works even with very small H values. Indeed, when the block size is substantial there is also an asymptotic normality that explains these results. For instance, let us consider the UC statistic J_{UC} , defined by the first orthonormal polynomial $P_1^{(N,\alpha)}$. For $N = 100$ and $\alpha = 0.05$, the binomial distribution can be approximated by a normal distribution (since $N\alpha \geq 5$, $N(1 - \alpha) \geq 5$ and $N > 30$), so that under UC:

$$P_1^{(N,\alpha)}(y_h) = \frac{y_h - N\alpha}{\sqrt{\alpha(1 - \alpha)N}} \sim N(0, 1), \quad \forall h \in \{1, \dots, H\}. \quad (2.22)$$

Consequently, for $H = 1$ ($N = T$), it is obvious that our J_{UC} statistic (equation 2.18) has a chi-squared distribution:

$$J_{UC} = \left[P_1^{(N,\alpha)}(y_1) \right]^2 \sim \chi^2(1). \quad (2.23)$$

For values of $H > 1$, we have the same result. For instance, let us consider the case where $H = 2$, *i.e.* where the block size N is equal to $T/2$. Then, the J_{UC} statistic is defined as follows (equation 2.18):

$$J_{UC} = \frac{1}{2} \left[P_1^{(N,\alpha)}(y_1) + P_1^{(N,\alpha)}(y_2) \right]^2, \quad (2.24)$$

or equivalently by

$$J_{UC} = \left(\frac{P_1^{(N,\alpha)}(y_1) + P_1^{(N,\alpha)}(y_2)}{\sqrt{2}} \right)^2, \quad (2.25)$$

where $P_1^{(N,\alpha)}(y_1) + P_1^{(N,\alpha)}(y_2)$ is the sum of two independent standard normal variables provided that the blocks are independent. So, under the UC assumption, we have:

$$\frac{P_1^{(N,\alpha)}(y_1) + P_1^{(N,\alpha)}(y_2)}{\sqrt{2}} \sim N(0, 1), \quad (2.26)$$

and consequently $J_{UC} \sim \chi^2(1)$.

The same type of results can be observed when the block size N is decreased. The rejection frequencies of the Monte-Carlo experiments for the $J_{CC}(m)$ GMM-based test statistic, both for a coverage rate of 5% and of 1% and for a block of size 25 are reported in Table 2.2. In that case, the normal approximation of the binomial distribution is not valid (since $N\alpha = 1.25$ or 0.25 given the values of α) and cannot be invoked to explain the quality of the results of our test. However, the number of observations H increases

for a given size T (relatively to the previous case $N = 100$), so the J statistics converge more quickly to a chi-squared distribution.

 Table 2.2: Empirical size (block size $N = 25$, nominal size 5%)

Coverage rate 5%							
T	H	J_{UC}	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	LR_{UC}	LR_{CC}
250	10	0.0386	0.0481	0.0417	0.0345	0.0558	0.0417
500	20	0.0547	0.0546	0.0550	0.0469	0.0573	0.0425
750	30	0.0461	0.0520	0.0583	0.0533	0.0572	0.0496
1000	40	0.0545	0.0567	0.0607	0.0510	0.0573	0.0592
1250	50	0.0489	0.0472	0.0555	0.0476	0.0423	0.0745
1500	60	0.0503	0.0515	0.0546	0.0472	0.0532	0.0685
Coverage rate 1%							
T	H	J_{UC}	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	LR_{UC}	LR_{CC}
250	10	0.0456	0.0551	0.0551	0.0462	0.0157	0.0128
500	20	0.0309	0.0673	0.0632	0.0537	0.0651	0.0114
750	30	0.0592	0.0588	0.0645	0.0624	0.0390	0.0196
1000	40	0.0345	0.0498	0.0508	0.0849	0.0534	0.0231
1250	50	0.0423	0.0546	0.0448	0.0438	0.0582	0.0244
1500	60	0.0461	0.0540	0.0449	0.0289	0.0513	0.0286

Note: Under the null hypothesis, the violations are i.i.d. and follows a Bernoulli distribution. The results are based on 10,000 replications. For each sample, we provide the percentage of rejection at a 5% level. $J_{CC}(m)$ denotes the GMM based conditional coverage test with m moment conditions. J_{UC} denotes the unconditional coverage test obtained for $m=1$. LR_{CC} (resp. LR_{UC}) denotes the Christoffersen's conditional (resp. unconditional) coverage test. T denotes the sample size of the sequence of interval forecasts violations I_t , while $H=[T/N]$ denotes the number of block (size $N=25$) used to define the sums (y_h) of violations.

It is important to note that these rejection frequencies are only calculated for the simulations providing a LR test statistic. Indeed, for realistic sample size (for instance $T = 250$) and a coverage rate of 1%, some simulations do not deliver a LR statistic. The LR_{CC} test statistic is computable only if there is at least one violation in the sample. Thus, at a 1% coverage rate for which the scarcity of violations is more obvious, a large sample size is required in order to compute this test statistic. The fraction of samples for which a test is feasible is reported in Table 2.3 for each sample size, both for the size and power experiments (at 5% and 1% coverage rate). By contrast, our GMM-based test can always be computed as long as the number of moment conditions m is inferior or equal to the block size N . This is one of the advantages of our approach.

2.4.2 Empirical Power Analysis

We now investigate the empirical power of our GMM test, especially in the context of risk management. As previously mentioned, Value-at-Risk (VaR) forecasts can be interpreted as one-sided and open forecast intervals. More formally, let us consider an interval $CI_{t|t-1}(\alpha) = [-\infty, VaR_{t|t-1}(\alpha)]$, where $VaR_{t|t-1}(\alpha)$ denotes the conditional VaR obtained for a coverage (or risk) equal to $\alpha\%$. As usual in the backtesting literature, our power experiment is based on a particular DGP for financial returns and a method to

Table 2.3: Feasibility ratios (coverage rate $\alpha = 1\%$)

Size simulations				
	$T = 250$	$T = 500$	$T = 750$	$T = 1000$
LR_{UC}	0.9185	0.9939	0.9991	0.9999
LR_{CC}	0.9179	0.9936	0.9991	0.9999
Power simulations				
	$T = 250$	$T = 500$	$T = 750$	$T = 1000$
LR_{UC}	0.9023	0.9966	1.0000	1.0000
LR_{CC}	0.9010	0.9966	1.0000	1.0000

Note: the fraction of samples for which a test is feasible is reported for each sample size, both for the size and power tests for a coverage rate equal to 1%. LR_{UC} and LR_{CC} are Christoffersen's (1998) unconditional and conditional coverage LR tests. Note that for J_{CC} the feasibility ratios are independent of the number of moment conditions m and are equal to 1. All results are based on 10,000 simulations. Note also that at 5% the LR tests can always be computed.

compute VaR out-of-sample forecasts. This method has to be chosen to produce invalid VaR forecasts according to Christoffersen's hypotheses.

Following Berkowitz et al. (2010), we assume that returns r_t are issued from a simple t -GARCH model with an asymmetric leverage effect:

$$r_t = \sigma_t z_t \sqrt{\frac{\nu - 2}{\nu}}, \quad (2.27)$$

where z_t is an *i.i.d.* series from Student's t -distribution with ν degrees of freedom, and where the conditional variance σ_t^2 is given:

$$\sigma_t^2 = \omega + \gamma \sigma_{t-1}^2 \left(\sqrt{\frac{\nu - 2}{\nu}} z_{t-1} - \theta \right)^2 + \beta \sigma_{t-1}^2. \quad (2.28)$$

Once the returns series has been generated, a method of VaR out-of-sample forecasting must be selected.⁵ Obviously, this choice has deep implications in terms of power performance for the interval forecast evaluation tests. We consider the same method as in Berkowitz et al. (2010), *i.e.* the historical simulation (HS), with a rolling window size T_e equal to 250. This unconditional forecasting method generally produces clusters of violations (violation of the independence assumption), and some slight deviations from the unconditional coverage assumption when we consider out-of-sample forecasts (these deviations depend on the size of the rolling window). Formally, we define the HS-VaR as following:

$$VaR_{t|t-1}(\alpha) = \text{Percentile} \left(\{r_i\}_{i=t-T_e}^{t-1}, 100\alpha \right). \quad (2.29)$$

For each simulation, a violation sequence $\{I\}_{t=1}^T$ is then constructed, by comparing the *ex-ante* $VaR_{t|t-1}(\alpha)$ forecasts to the *ex post* returns r_t . Next, the sequence $\{y_h\}_{h=1}^H$

⁵The coefficients of the model are parametrized as in Berkowitz et al. (2010): $\gamma = 0.1$, $\theta = 0.5$, $\beta = 0.85$, $\omega = 3.9683e^{-6}$ and $d = 8$. At the same time, ω has been chosen so as to be consistent with a 0.2 annual standard deviation. Additionally, the global parametrization corresponds to a daily volatility persistence of 0.975.

is computed for a given block size N by summing the corresponding I_t observations (see section 2.3.1). Based on this sequence, the J_{CC} test statistics are then implemented for different number of moment conditions and sample sizes T ranging from 250 to 1500. For comparison, both LR_{UC} and LR_{CC} statistics are also computed for each simulation. The rejection frequencies, at a 5% nominal size, are based on 10,000 simulations. In order to control for size distortions between LR and J_{CC} tests and to get a fair power comparison, we use Dufour's (2006) Monte-Carlo method (see appendix 2.7.2).

Tables 2.4 and 2.5 report the corrected power of the J_{UC} , $J_{CC}(m)$, LR_{UC} and LR_{CC} tests for different sample sizes T , in the case of a 5% and a 1% coverage rate, both for a block size $N = 100$ and $N = 25$. We can observe that the two GMM-based tests

 Table 2.4: Empirical Power (block size $N = 100$)

Coverage rate 5%							
T	H	J_{UC}	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	LR_{UC}	LR_{CC}
250	2	0.2776	0.3991	0.4274	0.4203	0.2268	0.3333
500	5	0.1586	0.6151	0.6379	0.6221	0.1464	0.3298
750	7	0.1457	0.7197	0.7280	0.7099	0.1209	0.3632
1000	10	0.1302	0.8164	0.8209	0.8116	0.1152	0.4212
1250	12	0.1266	0.8703	0.8774	0.8639	0.1179	0.4874
1500	15	0.1367	0.9122	0.9118	0.9079	0.1322	0.5207
Coverage rate 1%							
T	H	J_{UC}	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	LR_{UC}	LR_{CC}
250	2	0.1828	0.2709	0.2709	0.2820	0.1662	0.2730
500	5	0.2348	0.4525	0.4601	0.4403	0.1498	0.2361
750	7	0.2604	0.5410	0.5458	0.5516	0.2175	0.3073
1000	10	0.2980	0.6495	0.6596	0.6518	0.2116	0.3786
1250	12	0.3422	0.7051	0.6999	0.7058	0.2771	0.4407
1500	15	0.3663	0.7795	0.7738	0.7686	0.3330	0.4899

Note: Power simulation results are provided for different sample sizes T and number of blocks H , both at a 5% and 1% coverage rate. $J_{CC}(m)$ denotes the conditional coverage test with m moment conditions, J_{UC} represents the unconditional coverage test for the particular case when $m = 1$, and LR_{UC} and LR_{CC} are the unconditional and respectively conditional coverage tests of Christoffersen (1998). The results are obtained after 10,000 simulations by using Dufour's (2006) Monte-Carlo procedure with $M=9999$. The rejection frequencies are based on a 5% nominal size.

(J_{UC} and J_{CC}) have good small sample power properties, whatever the sample size T and the block size N considered. Additionally, our test is proven to be quite robust to the choice of the number of moment conditions m . Nevertheless, in our simulations it appears that the optimal power of our GMM-based test is reached when considering two or three moment conditions. For a 5% coverage rate, a sample size $T = 250$ and a block size $N = 25$, the power of our $J_{CC}(2)$ test statistic is approximately twice the power of the corresponding LR test for that experiment. For a coverage rate $\alpha = 1\%$, the power of our $J_{CC}(2)$ test remains by 30% higher than the one of the LR test. On the contrary, our unconditional coverage J_{UC} test does not outperform the LR test. This result is logical,

Table 2.5: Empirical Power (block size $N = 25$)

Coverage rate 5%							
T	H	J_{UC}	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	LR_{UC}	LR_{CC}
250	10	0.2656	0.5229	0.5314	0.4864	0.2285	0.3355
500	20	0.1842	0.7116	0.7022	0.6815	0.1482	0.3334
750	30	0.1509	0.8333	0.8277	0.8098	0.1155	0.3605
1000	40	0.1441	0.9091	0.9073	0.8919	0.1154	0.4374
1250	50	0.1444	0.9492	0.9439	0.9358	0.1218	0.4881
1500	60	0.1529	0.9717	0.9674	0.9637	0.1287	0.4981
Coverage rate 1%							
T	H	J_{UC}	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	LR_{UC}	LR_{CC}
250	10	0.2447	0.3697	0.3825	0.3866	0.1835	0.3355
500	20	0.2423	0.5163	0.5368	0.5410	0.1455	0.3334
750	30	0.2721	0.6436	0.6569	0.6232	0.2112	0.3605
1000	40	0.3253	0.7176	0.7428	0.7226	0.2044	0.4374
1250	50	0.3753	0.7926	0.7911	0.7896	0.2741	0.4881
1500	60	0.4373	0.8499	0.8456	0.8352	0.3368	0.4981

Note: Power simulation results are provided for different sample sizes T and number of blocks H , both at a 5% and 1% coverage rate. $J_{CC}(m)$ denotes the conditional coverage test with m moment conditions, J_{UC} represents the unconditional coverage test for the particular case when $m = 1$, and LR_{UC} and LR_{CC} are the unconditional and respectively conditional coverage tests of Christoffersen (1998). The results are obtained after 10,000 simulations by using Dufour (2006)'s Monte-Carlo procedure with $M=9999$. The rejection frequencies are based on a 5% nominal size.

since both exploit approximately the same information, *i.e.* the frequency of violations. Note that for UC tests (J and LR tests), the empirical power is decreasing, contrary to the CC tests. This result is specific to that experiment and comes from the use of the historical simulation to produce out-of-sample VaR forecasts. For large T sample, the deviation from the CC mainly comes from the clusters of violations. The empirical frequencies of hits is then close to the nominal coverage rate α .

The choice of the block size N has two opposite effects on the empirical power. A decrease in the block size N leads to an increase in the length of the sequence $\{y_h\}_{h=1}^H$ used to compute the J -statistic, and then leads to an increase in its empirical power. On the contrary, when the block size N increases, the normal approximation of the binomial distribution is more accurate. Thus, the finite sample distribution of our J statistics is close to the chi-squared distribution. This result is not due to the number of observations H , but to the normal approximation of the binomial. Figure 2.2 displays the Dufour's corrected empirical power of the $J_{CC}(2)$ statistic as a function of the sample size T for three values (2, 25 and 100) of the block size N . We note that whatever the sample size, the power for a block size $N = 100$ is always lesser than that obtained for a block size equal to 25. At the same time, the power with $N = 100$ is always larger than that with $N = 2$. In order to get a more precise idea of the link between the power and the block size N , Figure 2.3 displays the Dufour's corrected empirical power of the $J_{CC}(2)$ statistic

as a function of the block size N , for three values (250, 750 and 1500) of the sample size T . The highest corrected power corresponds to block sizes between 20 and 40, that is why we recommend a value of $N = 25$ for applications. Other simulations based on Bernoulli trials with a false coverage rate (available upon request) confirm this choice.

Thus, our new GMM-based interval forecasts evaluation test seems to perform better both in terms of size and power than the traditional LR ones.

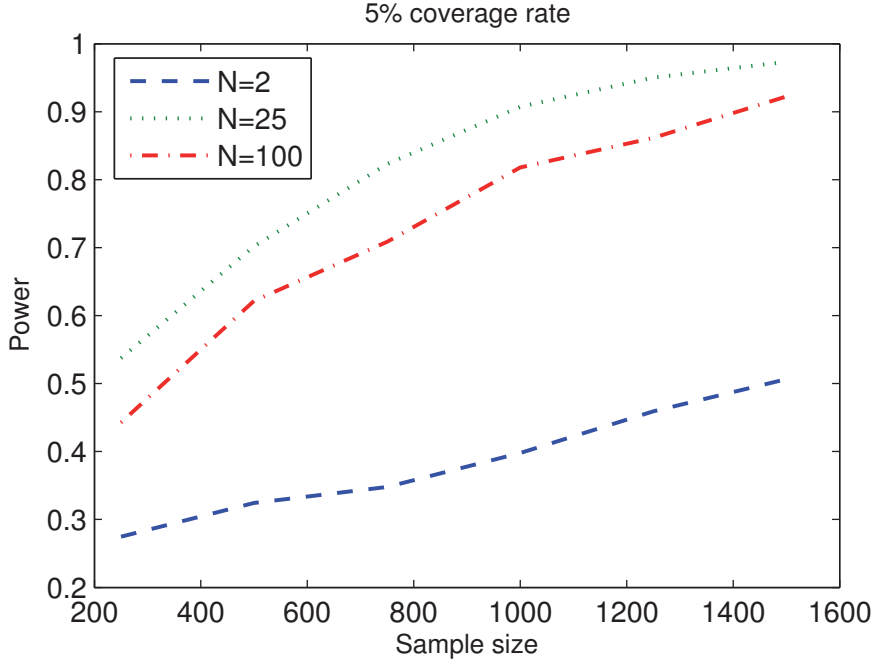


Figure 2.2: Corrected power of the $J_{CC}(2)$ test statistic as function of the sample size T (coverage rate $\alpha = 5\%$)

2.5 An Empirical Application

Now, we propose an empirical application based on two series of daily returns, namely the SP500 (from 05 January 1953 to 19 December 1957) and the Nikkei (from 27 January 1987 to 21 February 1992). The baseline idea is to select some periods and assets for which the linearity assumption is strongly rejected by standard specification tests. Then, we use (at wrong) a linear model to produce a sequence of invalid interval forecasts. The issue is then to check if our evaluation tests are able to reject the nulls of UC, IND and/or CC.

Here we use the nonlinearity test recently proposed by Harvey and Laybourne (2007). This takes into account both an ESTAR or LSTAR alternative hypothesis, and has very good small sample properties. For the considered periods, the conclusion of the test are clear: the linearity assumption is strongly rejected for both assets. For the SP500 (respectively Nikkei), the statistic is equal to 24.509 (respectively 89.496) with a p -value less than 0.001. As previously mentioned, we use simple autoregressive linear models

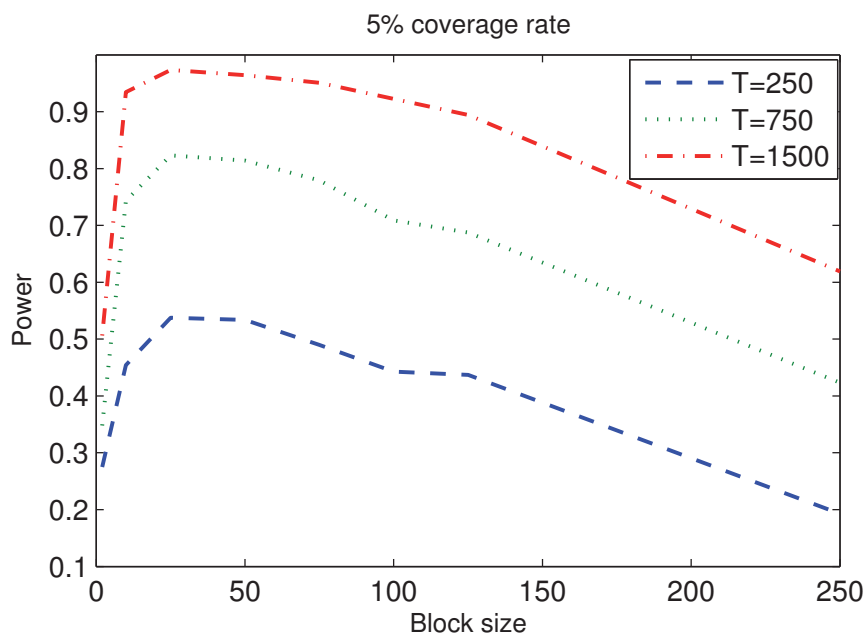


Figure 2.3: Corrected power of the $J_{CC}(2)$ test statistic as function of the block size N (coverage rate $\alpha = 5\%$)

AR(1) to produce forecasts and interval forecasts at an horizon $h = 1, 5$ or 10 days. More precisely, each model is estimated on the first 1,000 in sample observations, while continuous and symmetrical confidence intervals are computed for each sequence of 250 out-of-sample observations both at a 5% and 1% coverage rate.

Tables 2.6 and 2.7 report the main results of the interval forecast tests, based on a block size N equal to 25. It appears that for the SP500 index (see Table 2.6) our GMM-based test always rejects the CC hypothesis and thus the validity of the forecasts. In this case, the LR_{CC} test does not reject this hypothesis for a 5% coverage rate. When considering a 1% coverage rate, both CC tests succeed in rejecting the null hypothesis. Still, further clarifications are required. Both the UC and IND hypotheses are rejected when using GMM-based tests, whereas the only assumption rejected by the LR tests is the UC one. Similar results are obtained for the Nikkei series (see Table 2.6). Thus, the two series of interval forecasts are characterized by clusters of violations detected only by our GMM-based test. On the contrary, the LR_{IND} test appears not to be powerful enough to reject the independence assumption. This analysis proves that our evaluation tests for interval forecasts have interesting properties for applied econometricians, especially when they have to evaluate the validity of interval forecasts on short samples.

2.6 Conclusion

This paper proposes a new evaluation framework of interval and HDR forecasts based on simple J -statistics. Our test is model free and can be applied to intervals and/or HDR forecasts, potentially discontinuous and/or asymmetric. The underlying idea is that if the

Table 2.6: Interval Forecast Evaluation (SP500)

Coverage rate 5%						
Horizon	GMM-based tests			LR tests		
	J_{UC}	$J_{IND}(2)$	$J_{CC}(2)$	LR_{UC}	LR_{IND}	LR_{CC}
1	2.5263 (0.1120)	11.612 (0.0006)	29.493 (<0.0001)	2.4217 (0.1197)	3.3138 (0.0687)	5.8816 (0.0528)
5	4.4912 (0.0341)	10.615 (0.0011)	37.604 (<0.0001)	4.0607 (0.0439)	7.5661 (0.0059)	11.787 (0.0028)
10	2.5263 (0.1120)	19.605 (<0.0001)	46.040 (<0.0001)	2.4217 (0.1197)	3.3138 (0.0687)	5.8816 (0.0528)
Coverage rate 1%						
Horizon	GMM-based tests			LR tests		
	J_{UC}	$J_{IND}(2)$	$J_{CC}(2)$	LR_{UC}	LR_{IND}	LR_{CC}
1	109.09 (<0.0001)	11.612 (0.0006)	2072.4 (<0.0001)	49.234 (<0.0001)	3.3138 (0.0687)	52.693 (<0.0001)
5	134.68 (<0.0001)	10.615 (0.0011)	2658.3 (<0.0001)	57.475 (<0.0001)	7.5661 (0.0059)	65.201 (<0.0001)
10	109.09 (<0.0001)	19.605 (<0.0001)	2714.6 (<0.0001)	49.234 (<0.0001)	3.3138 (0.0687)	52.693 (<0.0001)

Note: 250 out of sample forecasts of the SP500 index (from 20/12/1956 to 19/12/1957) are computed for three different horizons (2, 5 and 10) both at a 5% and 1% coverage rate. The evaluation results of the corresponding interval forecasts are reported both for our GMM-based tests and Christoffersen's (1998) LR tests. For this objective, a block size $N=25$ was used. For all tests, the numbers in the parentheses denote the corresponding p-values.

Table 2.7: Interval Forecast Evaluation (Nikkei)

Coverage rate 5%						
Horizon	GMM-based tests			LR tests		
	J_{UC}	$J_{IND}(2)$	$J_{CC}(2)$	LR_{UC}	LR_{IND}	LR_{CC}
1	2.5263 (0.1120)	3.9132 (0.0479)	12.060 (0.0024)	1.7470 (0.1863)	0.2521 (0.6156)	2.1382 (0.3433)
5	1.7544 (0.1853)	3.8728 (0.0491)	9.6337 (0.0081)	1.1744 (0.2785)	0.4005 (0.5268)	1.7072 (0.4259)
10	1.7544 (0.1853)	3.8728 (0.0491)	9.6337 (0.0081)	1.1744 (0.2785)	0.4005 (0.5268)	1.7072 (0.4259)
Coverage rate 1%						
Horizon	GMM-based tests			LR tests		
	J_{UC}	$J_{IND}(2)$	$J_{CC}(2)$	LR_{UC}	LR_{IND}	LR_{CC}
1	109.09 (0.0000)	3.9132 (0.0479)	1279.3 (0.0000)	45.258 (<0.0001)	0.2521 (0.6100)	45.649 (<0.0001)
5	97.306 (0.0000)	3.8728 (0.0491)	1073.6 (0.0000)	41.384 (<0.0001)	0.4005 (0.5268)	41.916 (<0.0001)
10	97.306 (0.0000)	3.8728 (0.0491)	1073.6 (0.0000)	41.384 (<0.0001)	0.4005 (0.5268)	41.916 (<0.0001)

Note: 250 out of sample forecasts of the Nikkei index (from 27 January 1987 to 21 February 1992) are computed for three different horizons (2, 5 and 10) both at a 5% and 1% coverage rate. The evaluation results of the corresponding interval forecasts are reported both for our GMM-based tests and Christoffersen's (1998) LR tests. For this objective, a block size $N=25$ was used. For all tests, the numbers in the parentheses denote the corresponding p-values.

interval forecast is correctly specified, then the sum of violations should be distributed according to Binomial distribution with a success probability equal to the coverage rate. So, we adapt the GMM framework proposed by Bontemps (2006) in order to test for this distributional assumption that corresponds to the null of interval forecast validity.

More precisely, we propose an original approach that transforms the violation series into a series of sums of violations defined for H blocks of size N . Under the null of validity, these sums are distributed according to a Binomial distribution.

Our approach has several advantages. First, all three hypotheses of unconditional coverage, independence and conditional coverage can be tested independently. Second, these tests are easy to implement. Third, Monte-Carlo simulations show that all our GMM-based tests have good properties in terms of power, especially in small samples and for a 5% coverage rate (95% interval forecasts), which are the most interesting cases from a practical viewpoint.

Assessing the impact of the estimation risk for the parameters of the model that generated the HDR or the interval forecasts (and not for the distributional parameters) on the distribution of the GMM test-statistic by using a subsampling approach or a parametric correction is left for future research.

2.7 Appendix

2.7.1 Appendix: J statistics

Let us denote by $P^{(N,\alpha)} = (P_1^{(N,\alpha)}, \dots, P_m^{(N,\alpha)})$ a $(m, 1)$ vector whose components are the orthonormal polynomials $P_j^{(N,\beta)}(y_h)$ associated with the Binomial distribution $B(N, \alpha)$. Under the CC assumptions, the J statistic is simply defined by

$$J_{CC}(m) = \left(\frac{1}{\sqrt{H}} \sum_{h=1}^H P^{(N,\alpha)}(y_h) \right)' \Sigma^{-1} \left(\frac{1}{\sqrt{H}} \sum_{h=1}^H P^{(N,\alpha)}(y_h) \right), \quad (2.30)$$

where Σ denotes the long-run covariance matrix of $P^{(N,\alpha)}(y_h)$. Since Σ is by definition equal to the identity matrix, we have

$$J_{CC}(m) = \frac{1}{H} \sum_{j=1}^m \left(\sum_{h=1}^H P_j^{(N,\alpha)}(y_h) \right)^2. \quad (2.31)$$

Similarly, the independence hypothesis statistic, denoted $J_{IND}(m)$ takes the form of:

$$\begin{aligned} J_{IND}(m) &= \left(\frac{1}{\sqrt{H}} \sum_{h=1}^H P^{(N,\beta)}(y_h) \right)' \left(\frac{1}{\sqrt{H}} \sum_{h=1}^H P^{(N,\beta)}(y_h) \right) \\ &= \frac{1}{H} \sum_{j=1}^m \left(\sum_{h=1}^H P_j^{(N,\beta)}(y_h) \right)^2, \end{aligned} \quad (2.32)$$

where $P^{(N,\beta)}(y_h)$ is the $(m, 1)$ vector of orthonormal polynomials $P_j^{(N,\beta)}(y_h)$ defined for a coverage rate β that can be different from α .

2.7.2 Appendix: Dufour (2006) Monte-Carlo Corrected Method

To implement MC tests, we first generate M independent realizations of the test statistic, say S_i , $i = 1, \dots, M$, under the null hypothesis. Denote by S_0 the value of the test statistic obtained for the original sample. As shown by Dufour (2006) in a general case, the MC critical region is obtained as $\hat{p}_M(S_0) \leq \eta$ with $1 - \eta$ the confidence level and $\hat{p}_M(S_0)$ defined as

$$\hat{p}_M(S_0) = \frac{M \hat{G}_M(S_0) + 1}{M + 1}, \quad (2.33)$$

where

$$\hat{G}_M(S_0) = \frac{1}{M} \sum_{i=1}^M \mathbb{I}(S_i \geq S_0), \quad (2.34)$$

when the ties have zero probability, *i.e.* $\Pr(S_i = S_j) \neq 0$, and otherwise,

$$\hat{G}_M(S_0) = 1 - \frac{1}{M} \sum_{i=1}^M \mathbb{I}(S_i \leq S_0) + \frac{1}{M} \sum_{i=1}^M \mathbb{I}(S_i = S_0) \times \mathbb{I}(U_i \geq U_0). \quad (2.35)$$

Variables U_0 and U_i are uniform draws from the interval $[0, 1]$ and $\mathbb{I}(\cdot)$ is the indicator function. As an example, for MC tests procedure applied to the test statistic $S_0 = J_{CC}(m)$, we just need to simulate under H_0 , M independent realizations of the test statistic (*i.e.*, using durations constructed from independent Bernoulli hit sequences with parameter α) and then apply formulas (2.33 to 2.35) to make inference at the confidence level $1 - \eta$. Throughout the paper, we set M at 9,999.

Chapter 3

How to Evaluate an Early Warning System

Towards a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods¹

This paper proposes an original and unified toolbox to evaluate financial crisis Early Warning Systems (EWS). It presents four main advantages. First, it is a model free method which can be used to assess the forecasts issued from different EWS (probit, logit, Markov switching models, or combinations of models). Second, this toolbox can be applied to any type of crisis EWS (currency, banking, sovereign debt, etc.). Third, it does not only provide various criteria to evaluate the (absolute) validity of EWS forecasts but also proposes some tests to compare the relative performance of alternative EWS. Fourth, our toolbox can be used to evaluate both in-sample and out-of-sample forecasts. Applying it to a logit model for twelve emerging countries, we show that the yield spread is a key variable for predicting currency crises exclusively for South-Asian countries. Besides, the optimal cut-off correctly allows us to identify now on average more than 2/3 of the crisis and calm periods.

3.1 Introduction

Early Warning Systems (EWS) constitute a crucial tool for authorities to implement optimal policies to prevent or at least attenuate the impact of a financial turmoil. The first EWS was proposed by Kaminsky, Lizondo and Reinhart (1998) (KLR hereafter) relying on a signaling approach. They use a large database of 15 indicators covering the external position, the financial sector, the real sector, the institutional structure and

¹This chapter is based on Candelon, Dumitrescu and Hurlin (2012), forthcoming in *IMF Economic Review*.

the fiscal policy of a particular country. An indicator signals a crisis when it exceeds a particular cut-off. The estimation of this threshold is at the core of such an analysis. KLR determine it so as to minimize the noise-to-signal ratio (NSR), such that the probability of occurrence of a crisis is at its maximum after exceeding the cut-off. The EWS for country j is then built as the weighted-sum of the individual indicators, the weights being given by the inverse of the NSR . Berg and Patillo (1999) use panel probit models as EWS and show that their forecasting ability outperforms the one obtained using a signaling based model. This analysis hence paved the way for several other studies (Kumar et al., 2003; Fuertes and Kalotychou, 2007; Berg et al., 2008). These EWS do not exploit the fact that financial turmoils refer to specific regimes structurally different from the ones observed during tranquil periods. Hence, Bussiere and Fratzscher (2006) propose a multinomial logit EWS, whereas other studies use Markov-Switching models (see Abiad, 2003; Martinez-Peria, 2002 and Fratzcher, 2003).

Nevertheless, even if these approaches seem to be different, they suffer from similar drawbacks in their evaluation strategies. First, they all use the NSR measure (or sometimes similar measures of correct identification of crisis and calm periods) based on *ad hoc* cut-offs as *in fine* comparison criterion. Yet, as noticed by Bussiere and Fratzscher (2006 p.957) the choice of the cut-off is crucial: if it is low, crises will be more accurately detected (*i.e.*, the type I error will decrease), but at the same time, the number of false alarms will increase (*i.e.*, the type II error) leading to an efficiency cost in terms of economic policy. Second, no statistical inference is provided to test for the forecasting superiority of an EWS compared to another one. This absence represents an important issue, in particular when one has to choose between an EWS model exhibiting low type I and high type II errors and another one with different features.

Therefore, we argue that the evaluation of the forecasting abilities of EWS has not been sufficiently exploited, even though it is essential for crisis forecast. This paper aims at filling this gap by proposing an original evaluation methodology. First, our toolbox is model free, *i.e.* it can be applied to any EWS, whatever the model considered. This characteristic is essential given the great diversity of econometric approaches used in the EWS literature. Second, it can be applied to any type of crisis (currency, banking, debt, etc.) forecasting models. Third, we not only provide various criteria to evaluate the (absolute) validity of EWS forecasts but also propose some tests to compare the relative performance of alternative EWS. Fourth, our toolbox can be used to evaluate both in-sample and out-of-sample forecasts.

Our evaluation methodology is based on two steps. In a first step, for a given EWS model, we determine optimal cut-off points, *i.e.* thresholds, that best discriminate between crisis and calm periods. Elaborating on the traditional credit-scoring measure

(Basel Committee on Banking Supervision, 2005 and Lambert and Lipkovich, 2008 *inter alii*), it goes beyond a simple analysis of the *NSR* by determining the optimal threshold for each country as the value of the cut-off that maximizes (minimizes) different measures balancing type I and type II errors, *i.e.* *sensitivity-specificity* and *accuracy measures*. In a second step, various criteria and tests are proposed to compare alternative models.

The main finding of our paper is that an adequate EWS evaluation requires to take into account the cut-off in the model comparison step and then to determine an optimal crisis forecast. We show that traditional *Quadratic Probability Score (QPS)* - type criteria tend to conclude to the superiority of a model, even though the two alternative EWS considered have identical forecasting abilities. On the contrary, the criteria integrating the cut-off, *i.e.* *The Area Under the ROC Curve (AUC)*, behave correctly in this case. Furthermore, we argue that inference for nested and non-nested hypotheses is essential to identify the optimal specification. The choice of the outperforming model should thus rely on proper statistical tests which check the significance of the difference between the evaluation criteria associated with two alternative models. To this aim, the classic Diebold-Mariano test (1995) (and its nested version, Clark-West, 2007), as well as an *AUC* comparison test that takes into account the cut-off are proposed.

To empirically illustrate the utility of such an evaluation toolbox, we propose an application which aims at assessing the relevance of the yield spread in currency crises EWS. For economic theory, yield spreads are usually associated with credit growth sustained by excessive monetary expansion as well as investors' anticipations which can result in capital flight. Hence they may contain information on a potential future distress in the balance of payment. This economic reasoning can be considered as a special case of an EWS specification issue that gages the importance of a leading indicator for correctly forecasting crises. To this aim, we consider two EWS models, *one including the spread, the other without the spread*, in a fixed-effects panel framework. We assess their forecasting performances for six Latin-American and six South-Asian countries. We show that the criteria and tests including the cut-off should be favored as they allow us to refine the forecasting abilities of EWS. Indeed, the yield spread appears to be an important indicator of currency crises in half of the countries when we rely on tests including the cut-off such as the *Area under the ROC test*, whereas it first appears to be essential in almost all the countries in the sample if we consider the general *Clark-West test* based on standard QPS. The outperforming model (with or without spread) for each country is then used to forecast crises by relying on the optimal cut-off. It turns out that the optimal cut-off is quite different from the *NSR* one and, more importantly, it leads to a better trade-off between the two types of errors. In particular, the optimal cut-off correctly identifies on average more than 2/3 of the crisis and calm periods, in contrast with the *NSR* one, that

correctly forecasts all the calm periods at the expense of most of the crisis ones. Our findings seem robust to changes in the crises dating method.

The paper is organized as follows. Sections 3.2 to 3.4 present our new evaluation framework. More exactly, we tackle the determination of the optimal cut-off in section 3.2. Section 3.3 introduces the evaluation criteria whereas section 3.4 presents the comparison tests. Section 3.5 is devoted to the empirical application, which reveals the role played by the yield spread in currency crises EWS. Section 3.6 concludes.

3.2 Optimal Cut-off

The aim of any EWS is to forecast crisis and calm periods as correctly as possible, so that the appropriate policy measures can be taken in both tranquil and tumultuous situations. In this section we thus propose to quantify how well an EWS discriminates between the two types of periods by identifying the optimal cut-off.

3.2.1 How important is the cut-off choice?

EWS deliver probabilities indicating the chance for a specific crisis to occur in a certain period. Therefore, the in-sample (or out of-sample) evaluation of an EWS relies on the direct comparison of these crisis probabilities with an original crisis dating, which constitutes the benchmark.² This comparison implies two inputs of different nature: a sequence of probabilities and a crisis dating that takes the form of a dichotomic variable, labelled y_t . By convention, we assume that y_t takes a value equal to one if a crisis is identified at time t and zero otherwise.³

The forecasted probabilities are thus transformed into a dichotomic variable, known as crisis forecast. Formally, if we denote $\hat{p}_t$ the estimated (or forecasted) crisis probability at time t issued from an EWS model, the crisis forecast variable $\hat{y}_t$ is computed as follows:

$$\hat{y}_t(c) = \begin{cases} 1, & \text{if } \hat{p}_t > c \\ 0, & \text{otherwise} \end{cases}, \quad (3.1)$$

where $c \in [0, 1]$ represents the cut-off. In this perspective, the first step of any EWS evaluation consists in determining an optimal cut-off c that discriminates between predicted crisis periods ($\hat{y}_t(c) = 1$) and predicted calm periods ($\hat{y}_t(c) = 0$). The choice of the cut-off has strong implications for both forecast evaluation and economic analysis. Obviously, the cut-off determines type I and type II errors, *i.e.* the errors associated to

²We do not tackle here the pertinence of the crisis dating. We assume that economic experts are able (ex-post) to precisely date the periods of crisis. Nevertheless, a robustness analysis with respect to the potential inaccuracy of the crisis dating will be performed in the last section.

³It can also be assumed that y_t equals one if a crisis occurs in a certain time horizon (6, 12, 24 months, etc.), so as to forecast the approximate timing of a crisis some periods before it actually occurs (see KLR; Berg et al., 1999). This approach presents the advantage of giving the authorities the time necessary to implement appropriate policies to avoid an economic crash.

a misidentified crisis or to a false alarm. The type I error (or false negative) corresponds to a case in which the estimated (or forecasted) probability of crisis is smaller than the cut-off, but a crisis occurs. On the contrary, the type II error (also known as false alarm) corresponds to a situation in which the estimated (or forecasted) probability of crisis is larger than the cut-off whereas no crisis occurs. *Ceteris paribus*, the higher the cut-off is, the more type I (respectively type II) errors are frequent (respectively infrequent).

The optimal cut-off also contains economic interpretation in terms of vulnerability. The higher the probabilities during observed calm periods, the larger the optimal cut-off is and the more vulnerable the country is. However, the vulnerability concept is not model free, *i.e.* it depends on the underlying model and the decision-maker's risk aversion reflected in the method chosen to compute the cut-off. These vulnerability results should hence be interpreted with caution.

Given the cut-off's importance, it is surprising that the methods used to determine it are so over-arbitrary. At the same time, there is a very rich literature devoted to the EWS specification topic, stressing the choice of the most pertinent explanatory variables, the consequences of particular events on the models, etc. (see Jacobs et al., 2004 for a survey on financial crises EWS). And yet, to the best of our knowledge, no paper has been devoted to EWS evaluation and, more specifically, to the choice of an optimal cut-off.

Thus far, two types of cut-offs have been used in this literature. In most papers, the cut-off is arbitrarily fixed, generally to 0.5 or 0.25. This approach is economically nonsensical, since it means to arbitrarily determine type I and type II nominal risks. In other papers, it relies on the "Noise to Signal Ratio" (*NSR*) criterion proposed by KLR. In order to define the *NSR* cut-off, let us consider a sequence $\{y_t, \hat{y}_t(c)\}_{t=1}^T$.

Definition 1. *The optimal cut-off according to KLR minimizes the NSR criterion and is defined as:*

$$c_{NSR} = \arg \min_{c \in [0,1]} NSR(c), \quad (3.2)$$

where $NSR(c)$ represents the ratio of the false alarms (type II error or false positive) to the number of crises correctly identified (true positive) by the EWS for a given cut-off.

$$NSR(c) = \frac{\sum_{t=1}^T \mathbb{I}_{(\hat{y}_t(c)=1)} \times \mathbb{I}_{(y_t(c)=0)}}{\sum_{t=1}^T \mathbb{I}_{(\hat{y}_t(c)=1)} \times \mathbb{I}_{(y_t(c)=1)}}, \quad (3.3)$$

where $\mathbb{I}_{(z)}$ denotes an indicator function that takes the value 1 if z is true and 0 otherwise.

However, this criterion omits type I error and assumes that the costs related to the occurrence of a false alarm (type II error) overweight the ones inflicted by a misidentified crisis (type I error). This clearly constitutes an important constraint one may want to rule out, since type I error is usually the main element we try to control, as generally

done in other statistical literatures. We thus propose two methods which identify the optimal cut-off by taking into account both types of errors.

3.2.2 A credit-scoring approach

The first method is based on the traditional credit-scoring notions of *sensitivity* and *specificity* (Basel Committee on Banking Supervision, 2005). We thus label it the *credit-scoring* approach.

Definition 2. *The optimal credit-scoring cut-off is the threshold that minimizes the absolute value of the difference between sensitivity and specificity:*

$$c_{CSA}^* : \arg \min_{c \in [0,1]} |Se(c_{CSA}) - Sp(c_{CSA})|, \quad (3.4)$$

where *sensitivity* (Se), also known as *hit rate*, represents the proportion of crisis periods correctly identified by the EWS, while *specificity* (Sp) is the proportion of calm periods correctly identified by the model:

$$Se(c) = \frac{\sum_{t=1}^T \mathbb{I}_{(\hat{y}_t(c)=1)} \times \mathbb{I}_{(y_t(c)=1)}}{\sum_{t=1}^T \mathbb{I}_{(y_t(c)=1)}}, \quad (3.5)$$

$$Sp(c) = \frac{\sum_{t=1}^T \mathbb{I}_{(\hat{y}_t(c)=0)} \times \mathbb{I}_{(y_t(c)=0)}}{\sum_{t=1}^T \mathbb{I}_{(y_t(c)=0)}}. \quad (3.6)$$

The underlying idea is that variation in the cut-off leads to higher values of *sensitivity* corresponding to lower values of *specificity*. Figure 3.1 displays the *specificity* and *sensitivity* of an hypothetical EWS as functions of the cut-off c . The *sensitivity* is a decreasing function of c , since a rise in c results in decreasing the number of crisis signals, $\hat{y}_t(c) = 1$, and thus in the percentage of crises correctly predicted. On the contrary, the *specificity* is an increasing function of c . The higher c is, the higher the number of calm signals, $\hat{y}_t(c) = 0$, and hence the larger the proportion of calm periods correctly identified. The general form of both curves depends on the specification of the EWS. Even so, an optimal cut-off can be found at the intersection of both curves, as shown in Figure 3.1.

The main advantage of this *credit-scoring* identification method is that, in contrast to the *NSR* criterion, it relies on both type I and type II errors (see Engelmann et al., 2003; Renault et al., 2004 and Stein, 2005). It actually assigns equal weight to both types of errors. However, this assumption should be relaxed if we assume that the identification of a crisis is more costly for an economy than a false alarm (or vice versa). A possible extension of our method can be envisaged so as to take into account the costs c_1 and c_2 associated with the non-predicted crises and with a false alarm respectively. It simply consists in determining the optimal cut-off as the threshold that minimizes the

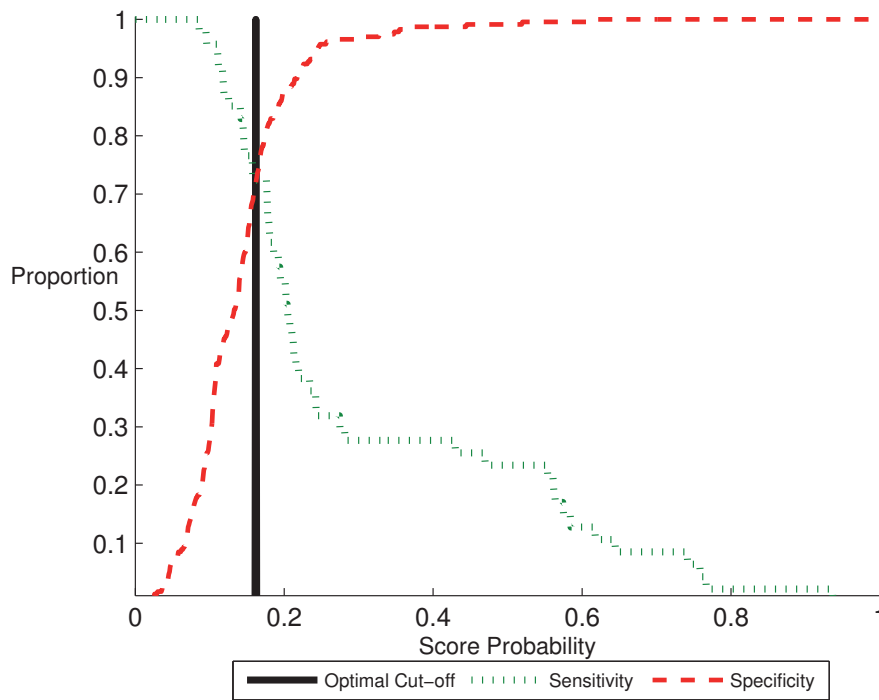


Figure 3.1: Optimal Cut-off determination

difference between the weighted *sensitivity* and the weighted *specificity*, where the weights are defined by c_1 and c_2 respectively. For example, the costs of the misidentification of a crisis -in terms of GDP- (c_1) can be approximated by an econometric evaluation of GDP gap during crisis periods. By contrast, the costs linked to false alarms (c_2) cannot be assessed easily, because they consist of costs incurred as a result of the reaction of monetary and/or banking authorities to an unfounded crisis announcement. If the policymaker can estimate these costs, however, this weighted method of identification of the optimal cut-off should be privileged.

Alternatively, instead of directly arbitrating between type I and type II errors, the optimal cut-off can be determined as the one that maximizes some accuracy measures or the one that minimizes the misclassification error measures respectively (Lambert and Lipkovich, 2008).

3.2.3 Accuracy Measures

The second approach consists in aggregating the number of crisis and calm periods correctly identified by the EWS in an accuracy measure. c is thus obtained by the maximization of the corresponding accuracy measure. The simplest measure, named *Total Accuracy* (TA), is defined as the ratio of cases correctly predicted and the total number of periods. Maximizing the TA measure is thus equivalent to maximizing the number of

correctly identified periods, whatever their type (crisis or calm). This measure does not arbitrate between type I and type II errors as the two types of periods are not considered separately (the denominator represents the total number of periods in the sample). We can thus be confronted with an undesirable situation in which the optimal cut-off correctly identifies all calm periods, but only a few or none of the crisis periods. We hence propose another measure, which arbitrates between the two types of errors.

Definition 3. *According to this accuracy measure, the optimal cut-off satisfies:*

$$c_{AM}^* = \arg \max_{c \in [0,1]} J(c), \quad (3.7)$$

where $J(c)$ denotes the Youden Index, defined as $J = Se(c) + Sp(c) - 1$.

The *Youden Index* ranges between 0 and 1; the higher the proportion of calm and crisis periods correctly identified by the model (relatively to the number of crisis and calm periods), the greater the J -measure. This optimal cut-off c_{AM}^* also corresponds to the cut-off that minimizes the misclassification error measure (also called *Total Error* measure) defined as the sum of the ratios of misidentified crises and false alarms to the number of crisis and calm periods respectively. More formally, the *Total Error* is defined as $TME(c) = 2 - Se(c) - Sp(c)$ and corresponds to $1 - J(c)$.

3.3 Evaluation Criteria

Traditionally, the forecasting abilities of an EWS are assessed only on the basis of the crisis probabilities p_t , *i.e.* independently of the cut-off. In fact, two criteria are generally used, namely the *Quadratic Probability Score (QPS)* and the *Log Probability Score (LPS)*.⁴ The *QPS* statistic is simply a mean square error measure comparing the crisis probability (the prediction) with an indicator variable for the crisis. It is defined as:

$$QPS = \frac{2}{T} \sum_{t=1}^T (\hat{p}_t - y_t)^2, \quad (3.8)$$

where $\hat{p}_t$ represents the *ex-ante* forecast probability of crisis at time t and y_t is a dummy variable taking the value one when a crisis occurs at time t . *QPS* takes values from 0 to 1, with 0 indicating perfect accuracy. This metric originated in weather forecasting and has been introduced by Diebold and Rudebusch (1989). It relies on the sum of squared residuals as in a standard linear model.

However, these traditional criteria evaluate the EWS only on the basis of the gap between the crisis probability and the realization of the observed crisis variable. The cut-off, essential for an EWS, is not taken into account.

⁴The *Log Probability Score (LPS)*, corresponds to a loss function that penalizes large errors more heavily than *QPS*, with $LPS = -1/T \sum_{t=1}^T [(1 - y_t) \ln(1 - \hat{p}_t) + y_t \ln(\hat{p}_t)]$. This score ranges from 0 to ∞ , with $LPS = 0$ being perfect accuracy.

We thus propose to include the cut-off in the validation of EWS by assessing the forecasting abilities of a model *conditionally to all the values of the cut-off, i.e.* from 0 to 1, in a similar vein to robustness analysis. This constitutes the main advantage of this approach, as the predictive abilities of a “good” EWS should not break down for reasonable changes in the value of the cut-off.⁵

3.3.1 Cut-off based criteria

In this context, we propose an original evaluation criterion, the *Area under the ROC Curve (AUC)*, first developed by electrical engineers and radar engineers during World War II for detecting enemy objects in battle fields, and then used in medicine, machine learning and credit scoring literature.

Definition 4. *The ROC (Receiving Operating Characteristic) curve⁶ is a graphical tool which reveals the predictive abilities of an EWS. More exactly, it represents the trade-off between sensitivity and $1 - \text{specificity}$ for every possible cut-off. The ROC curve is thus obtained by representing all the couples $\{Se(c); 1 - Sp(c)\}$ corresponding to each value of the cut-off c ranging from 0 to 1 (see Figure 3.2).*

For a perfect EWS model, the ROC curve passes through the point (0,1), indicating that it correctly recognizes all crisis and non-crisis periods. On the contrary, a completely random guess about crisis would give a point along the diagonal line (the so-called line of no-discrimination) from the left bottom to the top right corners.

This criterion can be summarized by the *Area Under the ROC curve (AUC)* defined as:

$$AUC = \int_0^1 [Se(c) \times (1 - Sp(c))] d(1 - Sp(c)). \quad (3.9)$$

An area under the ROC curve approaching 1 indicates that the EWS is getting closer to the perfect classification. In contrast, the expected value of the AUC statistic for a random ranking is 0.5.

The AUC is straightforward to implement, since it can be estimated by using an average of a number of trapezoidal approximations. Another way to obtain the AUC consists in using a non-parametric kernel estimator⁷ follows:

$$AUC = \frac{1}{T_1 \times T_0} \sum_{j:y_j=0} \sum_{i:y_i=1} K(\hat{p}_j, \hat{p}_i), \quad (3.10)$$

where T_1 (T_0) is the number of crisis (calm) periods in the sample, and $K(\cdot)$ denotes a kernel function that depends on the estimated crisis probabilities in crisis periods ($p_i, \forall i$:

⁵Theoretically, an alternative approach that jointly validates the optimal cut-off and the crisis probabilities may exist. Nevertheless, this approach is not feasible in our context. Indeed, the accuracy and misclassification measures cannot be employed, as they have been used to identify the optimal cut-off and no other adequate measures have been proposed so far.

⁶Also called Correct Classification Frontier, as in Jorda et al. (2011).

⁷This non-parametric estimator of the AUC criterion has recently been considered by Jorda et al. (2011) in the EWS literature, so as to compare different specifications with the random model (AUC=0.5).

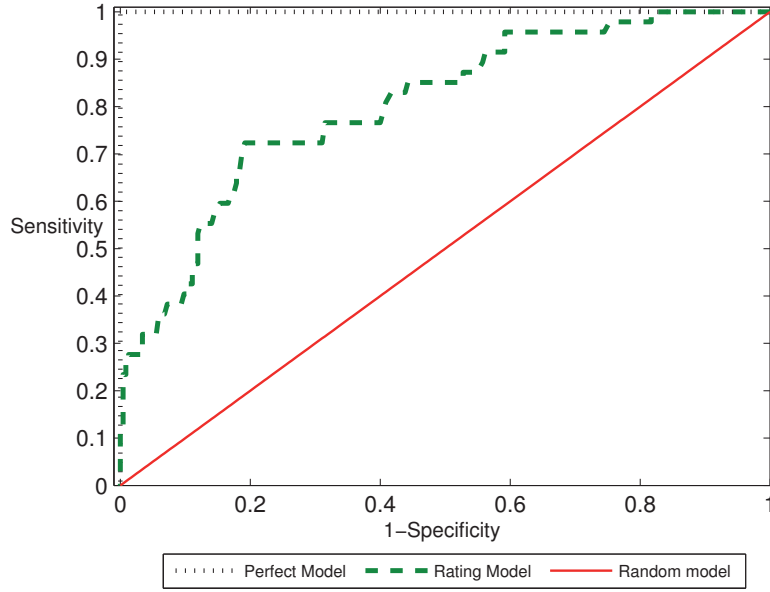


Figure 3.2: The ROC curve

$y_i = 1$) and calm periods ($p_j, \forall j : y_j = 0$) defined by:

$$K(\hat{p}_j, \hat{p}_i) = \begin{cases} 1, & \text{if } \hat{p}_i < \hat{p}_j \\ \frac{1}{2}, & \text{if } \hat{p}_i = \hat{p}_j \\ 0, & \text{if } \hat{p}_i > \hat{p}_j. \end{cases} \quad (3.11)$$

Our toolbox allows us to evaluate EWS models by taking into account the cut-offs (and therefore the crisis forecasts, the most important output of such a model) apart from relying solely on the crisis probabilities. The best model according to the *AUC* criterion is the outperforming EWS whatever the cut-off, and implicitly it is the best one conditional to the choice of the optimal cut-off. Indeed, taking into account the cut-off in the evaluation of an EWS can turn out to be crucial for an EWS with respect to a simple analysis based on *QPS*-type criteria.

3.3.2 Example

Let us consider a simple example of two EWS (denoted by A and B) with exactly the same forecasting abilities, but different estimated crisis probabilities. To be more exact, we suppose that the series of probabilities associated with model B , *i.e.* p_B , corresponds to an upward shift of the sequence of probabilities associated with model A , p_A , by a constant α : $p_A = p_B + \alpha$. It results that the two models have the same forecasting abilities, since the optimal cut-off for model B differs from that of model A only by α . The *sensitivity* and *specificity* series (and implicitly type I and type II

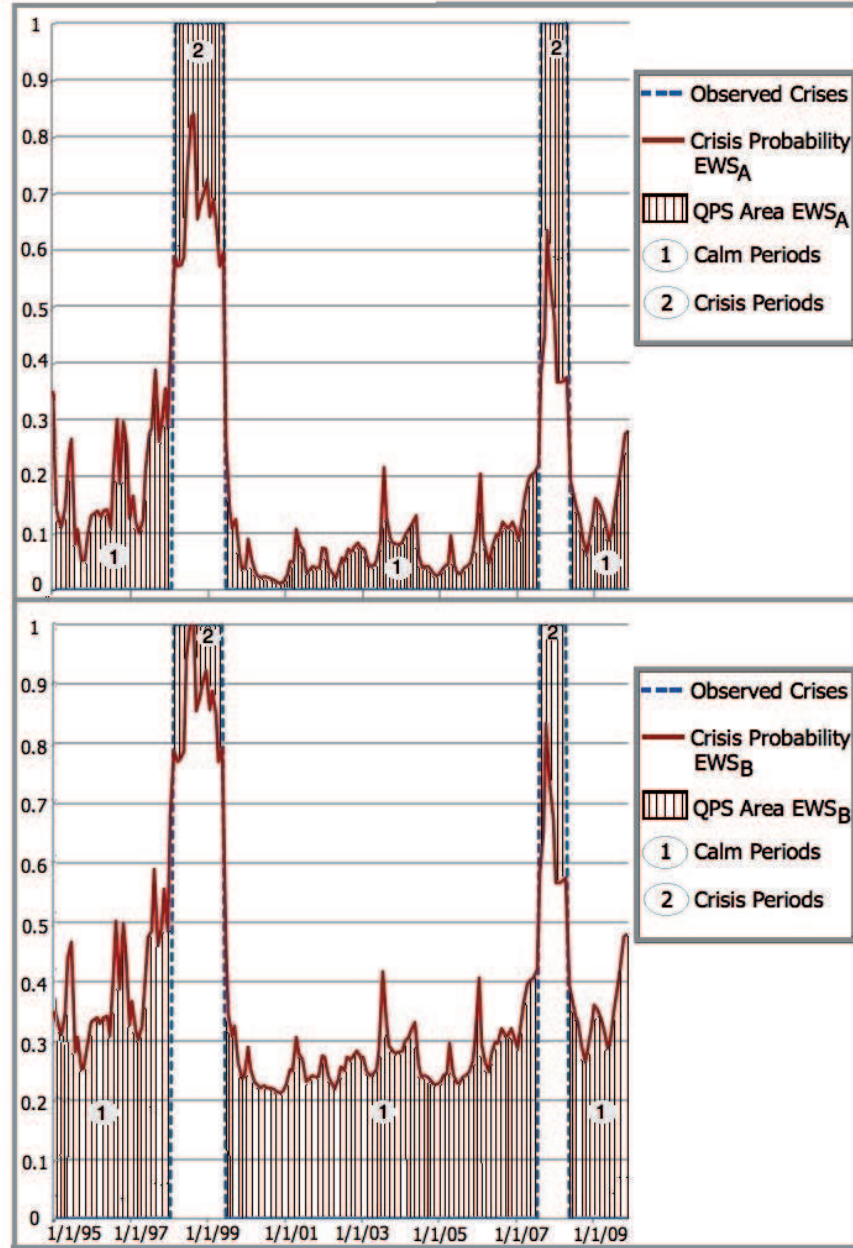


Figure 3.3: QPS - Graphical Approach

errors) associated with the two EWS are thus identical. And yet, in this context, QPS-type criteria wrongfully privilege one of the models. On the contrary, when taking into account the cut-off in the evaluation, *e.g.* AUC , we can conclude that the two models are utterly equivalent in terms of crisis forecasts.

Let us assume that $\alpha > 0$, where $\alpha \leq 1 - \max(p_B)$. In such a case, the crisis probabilities obtained from model B are always higher than those of model A , as illustrated in Figure 3.3.

Consider that the frequency of crisis occurrence is low and that ex-post we observe two crises of different but limited durations (which is the more common case in EWS

literature). QPS is based on the difference between the probabilities outputted by the EWS and unit (zero) during crisis (calm) periods. Accordingly, it corresponds to the sum of squared differences, depicted by hatched areas in Figure 3.3. It is thus clear that QPS for model B is higher than the one for model A if crises are not frequent, pointing out that model A has better predictive abilities. Still, the result depends on the frequency of observed crises (the length of hatch areas 1 and 2 in Figure 3.3). Formally, model A is preferred to model B if and only if:

$$QPS_B - QPS_A > 0 \Leftrightarrow 2 \sum_{t=1}^T [\alpha^2 - 2\alpha(y_t - \hat{p}_{A,t})] > 0. \quad (3.12)$$

This condition is fulfilled if:

$$\sum_{t=1}^T y_t < \frac{T\alpha + 2 \sum_{t=1}^T \hat{p}_{A,t}}{2}, \quad (3.13)$$

As a result, if the proportion of crisis periods, *i.e.* $1/T \sum_{t=1}^T y_t$, is small relatively to the ratio of the sum of probabilities to the number of periods, *i.e.* $1/T \sum_{t=1}^T p_{A,t}$ and the constant α , for $\alpha > 0$, the first model, A , is improperly considered to be more parsimonious than the latter, B . Besides, if the frequency of crises is below $\alpha/2$, eq. 3.13 is fulfilled independently of the sum of probabilities $\{p_{A,t}\}_{t=1}^T$. QPS privileges either model A or model B , even though they have identical forecasting abilities, except for the case where $\alpha = 0$, *i.e.* the two series of probabilities are identical. This finding is true provided that the cut-off has not been taken into account in the evaluation of the EWS. By contrast, the ROC evaluation criteria allow us to confirm this equivalence.

To illustrate this theoretical findings, let us consider the series of estimated probabilities, p_A , for two countries, Brazil (over the period 1994-2010) and Indonesia (over the period 1986-2009) obtained by estimating a logit model (see the empirical section for more details). Denote by p_B the sequence of probabilities obtained by shifting p_A upwards by $\alpha = 0.2$. The left part of Table 3.1 presents the QPS , and the AUC for each country and model.

Table 3.1: Example: Evaluation Criteria

Country	Model	QPS	AUC		CW	ROC
Brazil	EWS1	0.454030	0.639801	Statistic	5.537***	0
	EWS2	0.534030	0.639801	P-value	< 0.001	1
Indonesia	EWS1	0.231282	0.807832	Statistic	9.876***	0
	EWS2	0.311282	0.807832	P-value	< 0.001	1

Note: Two EWS with equal forecasting performance are compared by using evaluation criteria (QPS and AUC) as well as comparison tests (CW and ROC). The smaller the QPS , the better the model; the larger the AUC , the better the model. The null hypothesis of both test is the equality of forecasting abilities of the two models. The alternative indicates that the non-constraint model (EWS1) is better than the other one. The asterisks *, **, and *** denote significance at the 90%, 95% and 99% level, respectively.

Notice that *QPS* differs from one model to another, leading to the improper conclusion that the first model is slightly better than the second one and should be privileged for taking policy decisions. By comparison, the criteria based on the *ROC* curve are identical for the two models. To be more precise, not only the areas under the *ROC*, *i.e.* *AUC*, are equivalent, but also the *ROC curves* themselves.

Better still, apart from the aforementioned criteria (*QPS* and *AUC*), the comparison of two *EWS* models must rely on statistical test that we present in the next section.

3.4 Comparison Tests

Usually, the *EWS* literature aims to propose new econometric specifications (panel logit, Markov switching model, time varying probabilities Markov switching model etc.) or more frequently, new choices of explanatory variables in order to improve the crisis forecast ability (out-of-sample analysis) or the explanation of the crisis origins (in-sample analysis). These choices cannot be reduced to simple tests of significance, even if we are interested in the influence of a given variable on the crisis probability. It is well known that the significance of a parameter associated with a particular economic variable (in-sample) does not necessarily mean that this variable is able to improve the forecast ability of the *EWS*.

Hence, the *EWS* literature should be based on the comparison of forecasts issued from alternative models. However, this comparison is usually conducted only according to simple criteria such as *QPS* (with the drawbacks previously mentioned) without any statistical inference (see, for example, Kaminski, 2003; Arias and Erlandsson, 2005; Jacobs et al., 2008) even if they already exist in the statistical literature.

Accordingly, in the last step of our evaluation procedure, we propose to use a set of (model free) comparison tests in order to test the differences in crisis forecasts obtained from alternative *EWS* models. For that, let us consider two *EWS* models, denoted 1 and 2. Denote by $\{y_t\}_{t=1}^T$ the sequence of observed crises series and by $\{\hat{p}_{j,t}(c_1)\}_{t=1}^T$ the sequence of probabilities obtained from the j^{th} *EWS* model for $j = 1, 2$.

The most appropriate test in this context is the non parametric test of comparison of *ROC* curves (DeLong et al., 1988). It is based on the comparison of the areas under the *ROC* curves⁸ associated with the two *EWS* models, denoted AUC_1 and AUC_2 . The null of the test corresponds to the equality of areas under the *ROC* curves *i.e.*, $H_0 : AUC_1 = AUC_2$; In other words, neither of the models performs better than the other. DeLong et al. propose a test statistic based on the difference of *AUC* and use the theory on generalized U-statistics to construct an estimator of the variance of the difference (see Appendix 3.7.1 for technical details).

⁸Contrary to Jorda et al. (2011), who rely on a graphical comparison of the *AUC* for different models, we develop a statistical framework to evaluate *EWS*.

Definition 5. Under the null hypothesis $H_0 : AUC_1 = AUC_2$, the two EWS forecasts are equivalent and the AUC test statistic satisfies (DeLong et al., 1988):

$$W_{AUC} = \frac{(AUC_1 - AUC_2)^2}{\mathbb{V}(AUC_1 - AUC_2)} \xrightarrow[T \rightarrow \infty]{d} \chi^2(1), \quad (3.14)$$

At the same time, it is also possible to use the seminal test of comparison of forecast accuracy proposed by Diebold-Mariano (1995) and its specific version for nested models, proposed by Clark and West (2007). Both tests are based on the forecast errors of the two models, denoted $\{e_{1,t}\}_{t=1}^T$, and $\{e_{2,t}\}_{t=1}^T$, with $e_{j,t} = y_t - \hat{p}_{j,t}$ for $j = 1, 2$. The null corresponds to the hypothesis of equal forecasting accuracy, conditionally to a particular loss function $g(\cdot)$. We propose to use the MSFE, $g(e_{j,t}) = (y_t - \hat{p}_{j,t})^2$ that corresponds to half the QPS standard criterion, since $QPS = 2/T \sum_{t=1}^T g(e_{j,t})$.

Definition 6. Under the null hypothesis of equal predictive accuracy of the two EWS, $H_0: \mathbb{E}[(y_t - \hat{p}_{1,t})^2] = \mathbb{E}[(y_t - \hat{p}_{2,t})^2]$, the Diebold and Mariano (1995) test statistic DM verifies

$$DM = \frac{\sqrt{T}\bar{d}}{\sigma_{\bar{d},0}} \xrightarrow[T \rightarrow \infty]{d} N(0, 1), \quad (3.15)$$

where d_t denotes the loss differential, $d_t = (y_t - \hat{p}_{1,t})^2 - (y_t - \hat{p}_{2,t})^2$, $\bar{d}$ is the loss differential mean, $\bar{d} = (1/T) \sum_{t=1}^T d_t$ and $\sigma_{\bar{d},0}^2$ is the asymptotic long run variance of the loss differential.

Following standard practice, the long run variance $\sigma_{\bar{d},0}^2$ can be estimated with a Kernel estimator. Given the definition of the DM statistic, it is obvious that it fails to converge when the models are nested (since the denominator converges to zero). However, in EWS literature we often come across cases requiring to compare nested models. An appropriate test for nested models has been suggested by Clark and McCracken, (2001) and Clark and West, (2007).⁹

Note that the test of comparison of ROC curves relies on the AUC criterion whereas the DM test-statistic is based on the same loss function (MSE) as the QPS criterion. It follows that the ROC test takes into account not only the observed crises periods and the crises probabilities issued from two EWS specifications, as the DM (and its nested alternative CW) but also all the values of the cut-off.

Let us return to the previous example of Brazil and Indonesia in which we compare two EWS that have the same forecasting abilities. In the right part of Table 3.1 we

⁹Let us assume that model 1 is the parsimonious model and model 2 is the larger one, that reduces to model 1 if some of its parameters are set to 0. The corrected statistic proposed by Clark and West (2007), denoted CW , is defined as follows:

$$CW = \frac{\sqrt{T}\bar{f}}{\sigma_{\bar{f},0}} \xrightarrow[T \rightarrow \infty]{d} N(0, 1), \quad (3.16)$$

where $\hat{f}_t = (y_t - \hat{p}_{1,t})^2 - [(y_t - \hat{p}_{2,t})^2 - (\hat{p}_{2,t} - \hat{p}_{1,t})^2]$, $\bar{f}$ is the sample average of $\hat{f}_t$ and $\sigma_{\bar{f},0}^2$ is the sample variance of $\hat{f}_t - \bar{f}$.

present the test statistic and p-value for Clark-West (1997)'s test, based on a *QPS*-type loss function, and DeLong (1988) test, relying on *AUC* differences. It results that for these two countries, when the two series of probabilities are different enough ($\alpha = 0.2$), the W_{AUC} test does not reject the null hypothesis of equal forecasting abilities. On the contrary, the CW test leads to the rejection of the null hypothesis and consequently to an improper choice of model A over B .

We hence recommend the implementation of the *ROC* evaluation test as it allows us to gain more insights about the usefulness of different EWS specifications. Still, the *DM* (*CW*) test can also be used as a robustness check for their easy implementation, even though they do not always lead to a diagnostic as conclusive as the *ROC* test.

3.5 Empirical Application

We now propose an empirical application to illustrate the importance of the EWS evaluation procedure. This application focuses on the role of the yield spread (*i.e.*, long term -10 years- government bonds minus the short term -3 months- money rate) as a forward-looking indicator in the construction of EWS models. In a more general perspective, it can be viewed as an example of EWS specification where the main issue consists in assessing the importance of a leading indicator for correctly forecasting crises.

The yield spread can be considered as a forward interest rate that can be decomposed following the expectation hypothesis theory into an expected real interest rate and an expected inflation component (Estrella and Mishkin, 1996). It is hence linked to both changes in investors' expectations and expectations of future monetary policy. Since currency crises have been associated with credit growth sustained by excessive monetary expansion in many countries and investors anticipations can result in capital flight, aggravating a potential crisis, yield spreads can be assumed to reflect distress in the balance of payment. Moreover, since the yield spread seems to outperform other variables at long term forecasting horizons that are relevant from an investor's point of view, this variable is more forward-looking than other leading indicators (Estrella and Hardouvelis, 1991; Estrella and Trubin, 2006). Consequently, the use of the yield spread as a forecasting tool is even more compelling since it can signal the occurrence of a crisis in advance.

In this context, we propose to apply our evaluation methodology to assess the genuine usefulness of yield spread in forecasting currency crises.

3.5.1 EWS Specification

In order to assess the influence of yield spread, we consider a simple binary-choice model, *i.e.* a logit EWS which gives more weight to the tails of the distribution than a probit one. More formally, let y_{it} represent the binary crisis variable for country $i \in \{1, \dots, N\}$ at time $t \in \{1, \dots, T_i\}$. T_i denotes the number of time periods considered for the

i^{th} country (unbalanced panel). For each country, the crisis (logit) probability is defined as follows:

$$\Pr(y_{it} = 1) = \frac{\exp(\alpha_i + \beta_i' x_{it})}{1 + \exp(\alpha_i + \beta_i' x_{it})}, \quad i = 1, \dots, N, \quad (3.17)$$

where x_{it} denotes a vector of macroeconomic indicators, that includes the yield spread. α_i denotes a constant and β_i the vector of slope parameters. In this first specification, all parameters are country specific.

The approach generally used in the literature consists in estimating the binary EWS model in a panel framework by imposing some restrictions on the β_i parameters (see Berg and Patillo, 1999; Kumar et al., 2003 *inter alii*). It is well known that the panel approach is a way to reveal unobservable country heterogeneity and to increase the information set. This last point is particularly important in the specific context of currency crisis, given the relative scarcity of such events. However, this advantage has an obvious limit: the more heterogeneous countries are pooled in a “meta” model, the less the restrictions ($\beta_i = \beta$, for all i) on the slope parameters β_i are likely to make sense (even if we introduce individual effects α_i). This trade-off between more information and heterogeneity of slope parameters β_i can be summarized by a simple question, to pool or not to pool? To overcome this issue, Berg et al. (2008) recommend to construct country clusters, for which the slope parameters can be assumed to be homogeneous. There are two main approaches to construct such country clusters. The first one, proposed by Kapetanios (2003) is a pure statistical method based on an iterative procedure of homogeneity tests. The second approach focuses on macroeconomic similarities, crisis transmission mechanisms, contagion, etc.. We favor here the latter and consider two regional clusters, the first one including the Latin-American countries, and the second the South-Asian ones.¹⁰ For each country the crisis probability is then defined as follows:

$$\Pr(y_{it} = 1) = \frac{\exp(\alpha_i + \beta' x_{it})}{1 + \exp(\alpha_i + \beta' x_{it})} \quad \forall i \in \Omega_h, \quad (3.18)$$

where Ω_h is the h^{th} regional cluster, $h \in \{1, \dots, H\}$, and $\dim(\Omega_h) = N_h$ is the number of countries in the h^{th} cluster, so that $\sum_{h=1}^H N_h = N$. Besides, α_i represents the fixed effects (*i.e.*, the constant term specific to each country).

3.5.2 Data and Estimation

We consider a sample of twelve countries¹¹, for the period January 1980 to December 2010, extracted on a monthly frequency from the IMF-IFS database as well as the national banks of the countries under analysis via Datastream. The currency crisis indi-

¹⁰Note that, as a robustness check, we have also considered the pooled logit model, as well as the optimal clusters derived from the Kapetanios procedure.

¹¹Argentina, Brazil, Mexico, Peru, Uruguay, Venezuela, Indonesia, South Korea, Malaysia, Philippines, Taiwan and Thailand.

cator (the dependent variable), representing crises in the coming 24 months, is obtained by implementing the Kaminski, Lizondo and Reinhart (1998) modified dating method, thereafter *KLRm*, proposed by Lestano and Jacobs (2004) (see Appendix 3.7.2 for more details). Note that in the case of binary EWS models there is a debate related to the crisis dating quality, contrary to Markov-based models which do not require an *a-priori* identification of crises in the estimation step. However, the dating method impacts not only the estimation of an EWS, but also its evaluation, which means that the evaluation of Markov-based EWS depends on the dating method too. To check the sensitivity of our results to the dating method, we perform a robustness analysis based on the pressure index proposed Zhang (2001), instead of *KLRm*.

In all the estimated models (regional and pooled panel logit as well as country-by-country logit), the set of explicative variables includes growth of international reserves, growth of exports, growth of domestic credit over GDP, first difference of lending over deposit rate, first difference of industrial production index and yield spread. All variables are lagged one period. Among them, the yield spread plays a key role in our analysis since we aim to gauge its contribution to the improvement of the EWS. The other predictors are classic leading indicators for currency crises associated with devaluation pressure, loss of competitiveness, indebtedness, loan quality and recessions, respectively (see KLR; Berg and Patillo, 1999; *inter alii*). The procedure used to select these leading indicators is described in Appendix 3.7.2. A thorough attention has been given to the stationarity of the series, outliers and especially to the possible correlation among leading indicators (see Appendix 3.7.2 for more details). We also take into account the potential presence of serial correlation ¹² by using the sandwich estimator of the covariance matrix (Williams, 2000). This method is described in Appendix 3.7.3.

The estimation results for the model including the yield spread are presented in Table 3.2. It appears that the yield spread is one of the most important explanatory variables both in panel and time-series models. It is significant at a 5% level for the Latin-American cluster as well as for 11 out of 12 countries. The results for the pooled panel model confirm these findings. This first result implies that in all the countries considered a higher short term interest rate with respect to the long term one, *i.e.* a negative slope in yield spread, signals future balance of payment problems that lead to currency crises. The other explanatory variables generally have the correct sign too, but their significance plummets when regrouping the countries in a panel set without accounting for the heterogeneity of the estimated parameters.

¹²Berg and Coke (2004) show that considering a forecast horizon larger than 1 leads to autocorrelation in the crisis variable. This stylized fact is confirmed by Harding and Pagan (2006).

Table 3.2: EWS Estimation

	Panel Model			TS Model	
	Pooled	Region 1	Region 2	Significance	
				5%	1%
Growth of international reserves	−1.438 (−0.500)	−0.800 (−0.210)	−3.793 (−0.770)	5	4
Growth of exports	−2.872 (−0.550)	−3.088 (−0.510)	−4.677 (−0.650)	7	4
Growth of domestic credit over GDP	1.342** (2.700)	−0.268 (−0.210)	1.634*** (3.590)	6	3
Lending rate over deposit rate (first difference)	−0.005 (−0.130)	−2.451 (−1.610)	0.025 (0.710)	1	
Growth of industrial production (first difference)	−0.211 (−0.600)	−0.328 (−0.480)	−0.068 (−0.110)		
Yield spread	−1.522** (−2.630)	−1.049** (−3.170)	−2.404* (−1.740)	11	5
Constant				7	6

Note: The table presents the estimation results for the pooled panel model, regional panel model (South America and South Asia) and country-by-country (time-series) models. The figures between parentheses are t-statistics. The asterisks *, **, and *** denote significance at the 90%, 95% and 99% level, respectively. For the time-series models we present the number of countries for which a specific variable is significant at a given risk level.

3.5.3 EWS Evaluation

To analyze the importance of the yield spread in forecasting currency crises, we consider two specifications of our EWS model: *one that includes the yield spread indicator and another that does not include it* and compare their forecasting abilities. These models are estimated for the two optimal regional clusters (South America and South Asia). First, let us compare the two specifications (with and without spread) in a country-by-country setting by using a *QPS*-type criteria, as it is usually done in the literature. The left part of Table 3.3 displays the *QPS* corresponding to the two models for each of the twelve countries in our sample. The *QPS* criterion seems to confirm that the spread improves the forecasting abilities of the EWS for almost all the countries (10 out of 12).

While most of the papers in the literature would stop here and conclude to the importance of the yield spread for all countries, we propose to go further on and test if the spread is *significantly* important for a currency crisis EWS. As shown in section 3.4, since both specifications are nested, (the former can be reduced to the latter by imposing the nullity of the spread parameter), Clark-West's (2007) *CW* test is used to compare the forecasting ability of the two logit models for each country in the sample. The results are displayed in the right part of Table 3.3. The tests roughly confirm the importance of spread's contribution to currency-crises forecasting but only at 10% level, offering hence a more refined diagnostic than *QPS*-criteria. The use of formal tests shows that for some countries the introduction of the spread in the model does not improve the model's ability to correctly forecast currency crises. For instance, Figure 3.4 depicts the crisis probabili-

Table 3.3: EWS Evaluation: Regional Panel Model

	Evaluation Criteria			Comparison Tests	
	Model	QPS	AUC	CW test	W_{AUC} test
Argentina	with spread	0.276*	0.683	2.819** (0.005)	0.175 (0.676)
	without spread	0.288	0.705*		
Brazil	with spread	0.456*	0.659*	1.872* (0.061)	0.180 (0.672)
	without spread	0.461	0.646		
Indonesia	with spread	0.219*	0.803*	4.115 *** (<0.001)	17.02 *** (<0.0001)
	without spread	0.262	0.616		
Malaysia	with spread	0.256*	0.839*	1.792* (0.073)	5.645** (0.018)
	without spread	0.259	0.765		
Mexico	with spread	0.391*	0.611*	3.236 *** (<0.001)	4.970** (0.026)
	without spread	0.400	0.537		
Peru	with spread	0.467	0.559	0.510 (0.610)	0.670 (0.413)
	without spread	0.459*	0.601*		
Philippines	with spread	0.496	0.477	-2.056** (0.040)	2.379 (0.123)
	without spread	0.439*	0.537*		
South Korea	with spread	0.229*	0.829*	5.918 *** (<0.001)	15.51 *** (<0.001)
	without spread	0.260	0.691		
Taiwan	with spread	0.333*	0.687*	4.529 *** (<0.001)	34.09 *** (<0.001)
	without spread	0.351	0.443		
Thailand	with spread	0.230*	0.849*	4.879 *** (<0.001)	25.56 *** (<0.001)
	without spread	0.285	0.648		
Uruguay	with spread	0.223*	0.816*	3.367 *** (<0.001)	0.395 (0.530)
	without spread	0.270	0.786		
Venezuela	with spread	0.486*	0.544*	1.686* (0.092)	2.53 (0.112)
	without spread	0.488	0.471		

Note: QPS ranges from 0 to 2, the lower its level, the better the model. The AUC criterion takes values between 0.5 and 1, 1 being the perfect model. The best model according to each evaluation criteria is denoted by an asterisk (*). The null hypothesis of the comparison tests is the equality of predictive performance of the two models. The alternative of the Clark-West, CW , and Diebold-Mariano test, DM , is the statistical difference between the two criteria (it indicates that the model with the smaller QPS is better than the other one), while the alternative hypothesis of the DeLong, W_{AUC} , test is the statistical difference between the two areas (the model with a larger AUC is better). The asterisks *, **, and *** denote test significance at 90%, 95% and 99% level, respectively.

ties for Brazil issued from the model with spread (EWS1) and the model without spread (EWS2). Graphically, the series of probabilities seem almost identical, confirming the results of the CW test. These findings support the use of statistical tests to compare EWS instead of relying on simple QPS – type criteria.

In the next step, we propose to check this diagnostic by relying on criteria integrating the cut-off. We thus use the AUC evaluation criterion (see the left part of Table 3.3) and find evidence of a positive effect of the yield spread on the forecasting abilities of the EWS. For most of the countries, the AUC is always higher for the logit with yield spread relatively to the logit without yield spread. Still, relying on AUC criterion instead of

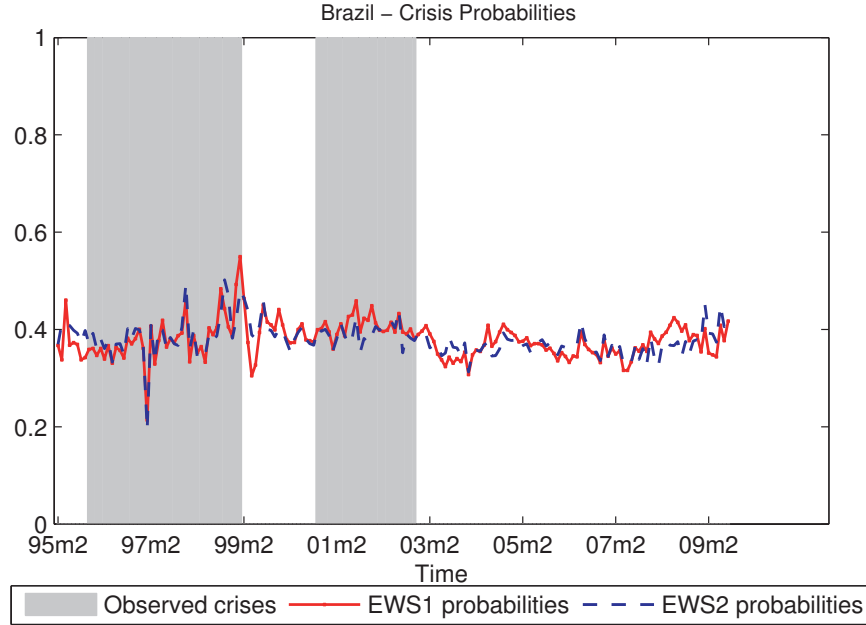


Figure 3.4: Brazil - Crisis probabilities

QPS, we identify one country, namely Argentina, for which the yield spread does not contribute to the improvement of the EWS.

However, the main impact of the cut-off is that the differences between criteria are generally proven to be statistically insignificant. The right part of Table 3.3 displays the DeLong's (1988) test statistics W_{AUC} and their p -values. Under the null, the areas under the *ROC* of both logit models are identical. These tests conclude to the rejection of the null only for 6 countries out of 12. In particular, for Argentina, Brazil, Peru, Philippines, Uruguay and Venezuela no gain in terms of sensitivity and specificity - no improvement in the type I and type II errors - results from the introduction of the spread in the EWS model. Indeed, this test leads to the conclusion that spread is important mostly for South-Asian countries. It appears that 5 out of the 6 countries for which this leading indicator improves the EWS belong to this cluster. For the South-American cluster, it seems that the yield spread does not significantly improve the crisis forecasts.

This application emphasizes the importance of inference for evaluating EWS. Using simple criteria, *QPS* in particular, can lead to misleading results. We thus propose two tests, *i.e.* *ROC* and $DM(CW)$. The *ROC* test, that takes into account the cut-off in the evaluation, relativizes the importance of spread in crisis forecasting and proves to be more performant than the *CW*. We hence argue that this test should be privileged in the evaluation of EWS models. Nevertheless, this test can be supplemented by the $DM(CW)$ test which, by contrast, is much easier to implement and already known in the forecasting literature.

3.5.4 Optimal cut-off

Previous results suggest that for some countries, *e.g.* Indonesia, Malaysia, Mexico, South Korea, Taiwan, and Thailand, the yield spread has a significant impact on the forecasting abilities of the EWS, whereas for others the influence of this variable is not significant. We now investigate the forecast accuracy of the outperforming model (with or without spread) for each country by identifying the optimal cut-off and calculating the associated percentage of correctly identified crisis (respectively calm) periods, *i.e.* *sensitivity* (respectively *specificity*). To emphasize the importance of the optimal cut-off in crisis forecasting, the time-series results are also presented in a similar vein to robustness check.

Table 3.4: EWS Optimal Cut-off: Descriptive Statistics

Optimal Cut-off	Time-Series			Regional Panel		
	c_{CSA}^*	c_{AM}^*	c_{NSR}	c_{CSA}^*	c_{AM}^*	c_{NSR}
Average	0.245	0.239	0.815	0.246	0.233	0.477
Std-deviation	0.079	0.117	0.130	0.084	0.092	0.213
Minimum	0.147	0.115	0.567	0.162	0.118	0.245
Maximum	0.371	0.437	0.993	0.378	0.388	0.955
Average Sensitivity	0.72	0.819	0.048	0.656	0.745	0.052
Average Specificity	0.72	0.711	1.000	0.654	0.635	1.000

Note: This table includes some descriptive statistics for the cut-offs. We select the optimal cut-off for the best model by using two methods (credit-scoring - *CSA*-, and accuracy measures - *AM*-). For comparison reasons we also present KLR's *NSR* cut-off. The corresponding average levels of correctly identified crisis and calm periods are also included.

The right part of Table 3.4 displays some descriptive statistics of the three cut-offs considered, *i.e.* credit-scoring, c_{CSA}^* , accuracy measures, c_{AM}^* and noise-to-signal ratio, c_{NSR} for the country-by country analysis. The major insight is that the *NSR* cut-off proposed by Kaminski et al. (1998) is always larger than the optimal cut-off we propose. The difference is significant, as the mean ratio is of 2 to 1 for the panel analysis and 3.3 to 1 for the country-by-country analysis. Besides, c_{NSR} is characterized by a larger dispersion relative to the other two cut-offs. It results that the average forecast performance of the EWS model is quite different given the cut-off choice. The use of an optimal cut-off leads on average to a correct identification of at least 2/3 of the crisis and calm periods, as sensitivity exceeds 72% on average and specificity outruns 71%.

On the contrary, the *NSR* cut-off leads to a perfect identification of calm periods (*specificity* = 100%). However, this accuracy with respect to calm periods is possible only to the detriment of the crisis ones, since the average sensitivity is equal to 5.2%. In other words, the average type I error (false negative) is larger when using c_{NSR} , while the average type II error (false alarms) is lower compared to c_{CSA}^* and c_{AM}^* . Thus, the optimal cut-offs are less sensitive to false alarms compared with missed crises. One

rationale behind this could be that policy makers and enterprises are possibly willing to take a 'crisis insurance' and to accept a possible false alarm rather than be taken by surprise by a crisis, especially since the costs of a false alarm are thought to be inferior to those engendered by an unexpected crisis.

Table 3.5: EWS Forecasting abilities

		Time-series			Regional Panel		
		cut-off	sensit	specif	cut-off	sensit	specif
Argentina (without spread)	c_{CSA}^*	0.151	0.703	0.703	0.185	0.703	0.648
	c_{AM}^*	0.120	0.892	0.624	0.185	0.703	0.648
	c_{NSR}	0.782	0.108	1.000	0.254	0.081	1.000
Malaysia (mixt)	c_{CSA}^*	0.293	0.935	0.939	0.171	0.839	0.810
	c_{AM}^*	0.293	0.935	0.939	0.169	0.935	0.798
	c_{NSR}	0.993	0.000	1.000	0.245	0.000	1.000

Note: The optimal model, as resulting from table 3.3, is chosen for each country. We select the optimal cut-off by using two methods (credit-scoring - CSA -, and accuracy measures - AM -). The corresponding proportion of correctly identified crisis (calm) periods, denoted *sensit*, (*specif*) are also presented. Note that type I and type II errors can be obtained as their complement. For comparison reasons, we also present the NSR cut-off.

The performance of the optimal model at the country level confirms our previous findings *i.e.* the use of an optimal cut-off improves significantly currency crises forecasts (see Table 3.5). Take the example of Argentina.¹³ The use of the NSR cut-offs leads to the correct identification of all the calm periods, but only 8.1% of the crisis periods. Assuming that a crisis occurs each ten years on average, it would take the model 123.46 years to correctly predict a currency crisis. By contrast, with our c_{CSA}^* cut-off the probability to correctly identify crisis periods rises to 70.3% and the time necessary to correctly identify a crisis reduces to 14.2 years. This increase in sensitivity (and drop in misidentified crises) is possible only at the cost of more false alarms. However, if we admit that the cost of a false alarm is lower than the one of a misidentified crisis, the gain becomes evident, especially since a large increase in the proportion of crises correctly identified (from 8.1% to 70.3%) is associated with a smaller reduction of *specificity* (from 100% to 64.8%). An extreme case is that of Malaysia. To correctly identify all calm periods, the value of the NSR cut-off rises until reaching the value of 0.245, missing all the crisis periods (*sensitivity* = 0). By contrast, the optimal cut-offs correctly identify not only the crisis periods (*sensitivity* = 0.839) but also the calm ones (*specificity* = 0.810). The time-series results confirm this forecast gain associated with the computation of the optimal cut-off.

Figure 3.5 depicts the cut-offs (z axis) against their associated *sensitivity* (x axis) and *specificity* (y axis) in a 3D scatterplot. The points on this figure correspond to the three types of cut-off estimated (CSA , NSR and AM) for the twelve countries in the

¹³The results for the other countries are available upon request.

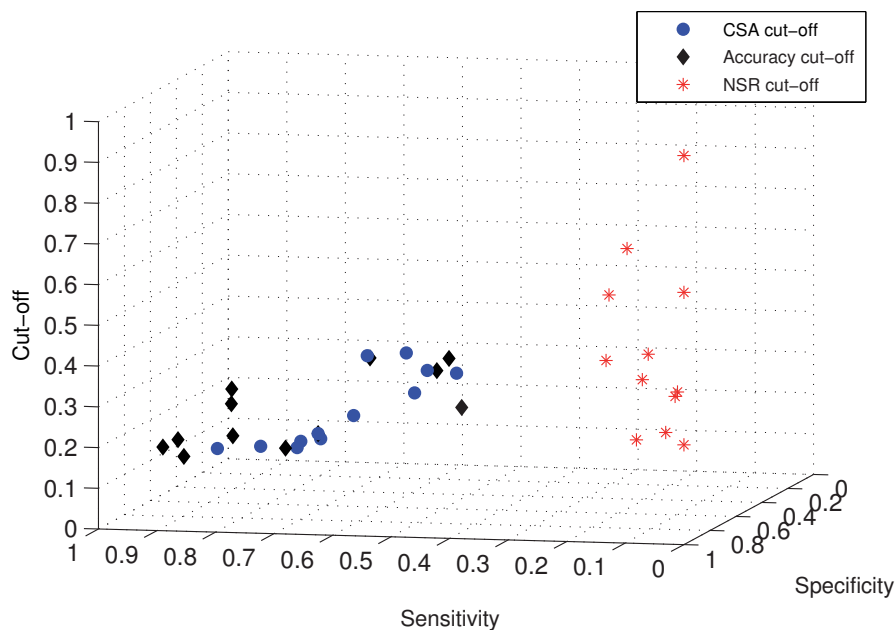


Figure 3.5: Cut-off - Regional Panel Models

sample. We observe that the values of optimal cut-off CSA and AM are relatively low (z axis) compared to the NSR ones. Besides, they are concentrated in a region with large *specificity* and *sensitivity*, indicating that these cut-offs correctly identify most of the crisis and calm periods. On the contrary, the NSR cut-off correctly identifies all calm periods at the expense of most of the crises. Graphically, these cut-offs are clustered around specificity equal to one (y axis) and small values of sensitivity (x axis).

As mentioned in section 3.2.1, the optimal cut-off can also be analyzed in terms of vulnerability to crisis. Figures 3.6 and 3.7 depict the currency crisis probability series issued from the optimal EWS model as well as the optimal credit-scoring cut-off c_{CSA}^* . Notice that the crisis probabilities during the calm periods in the second half of the sample are as elevated as before, suggesting that the forecasting abilities of macroeconomic indicators have not been improved recently.

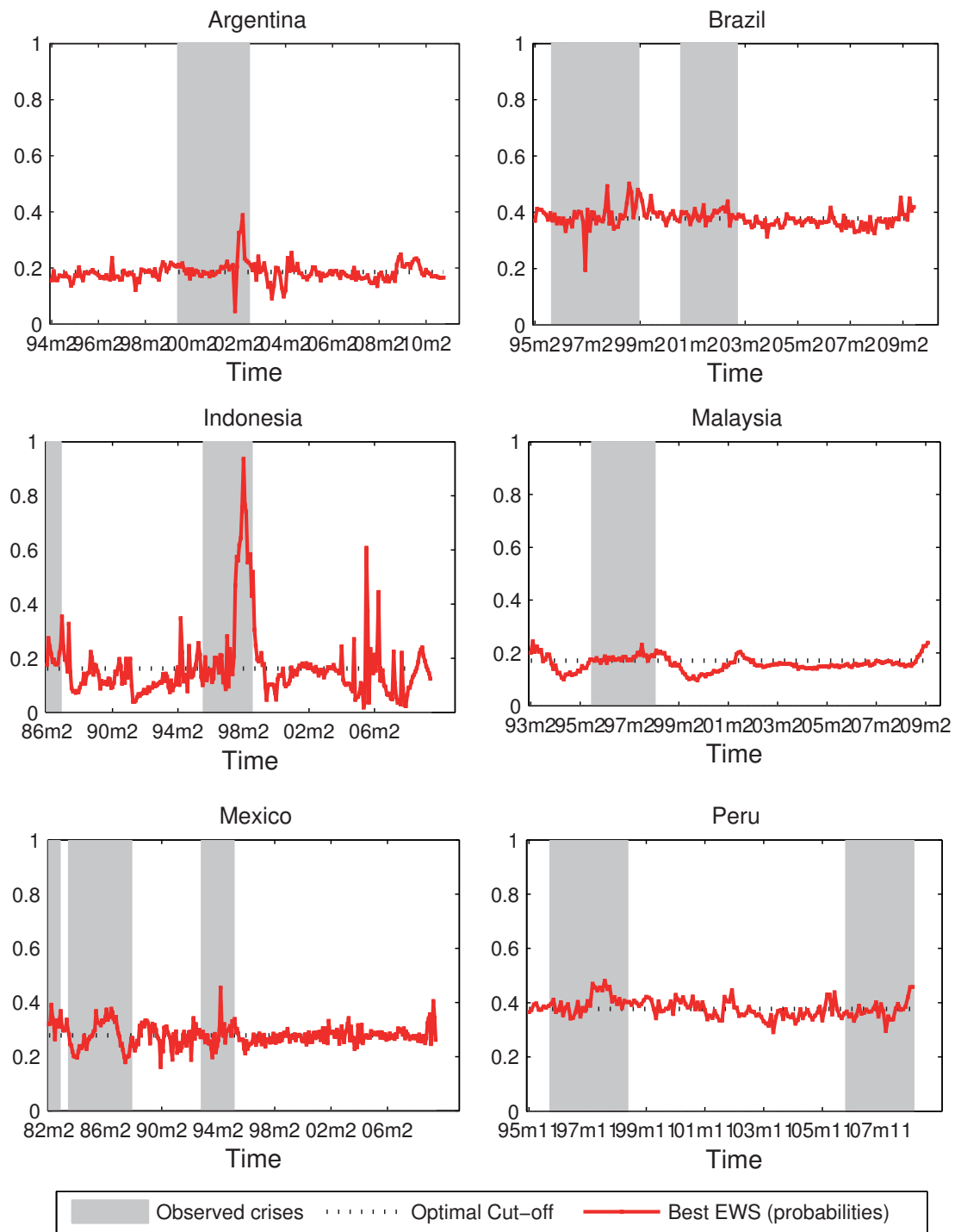


Figure 3.6: Crisis Probabilities - Time-Series Models

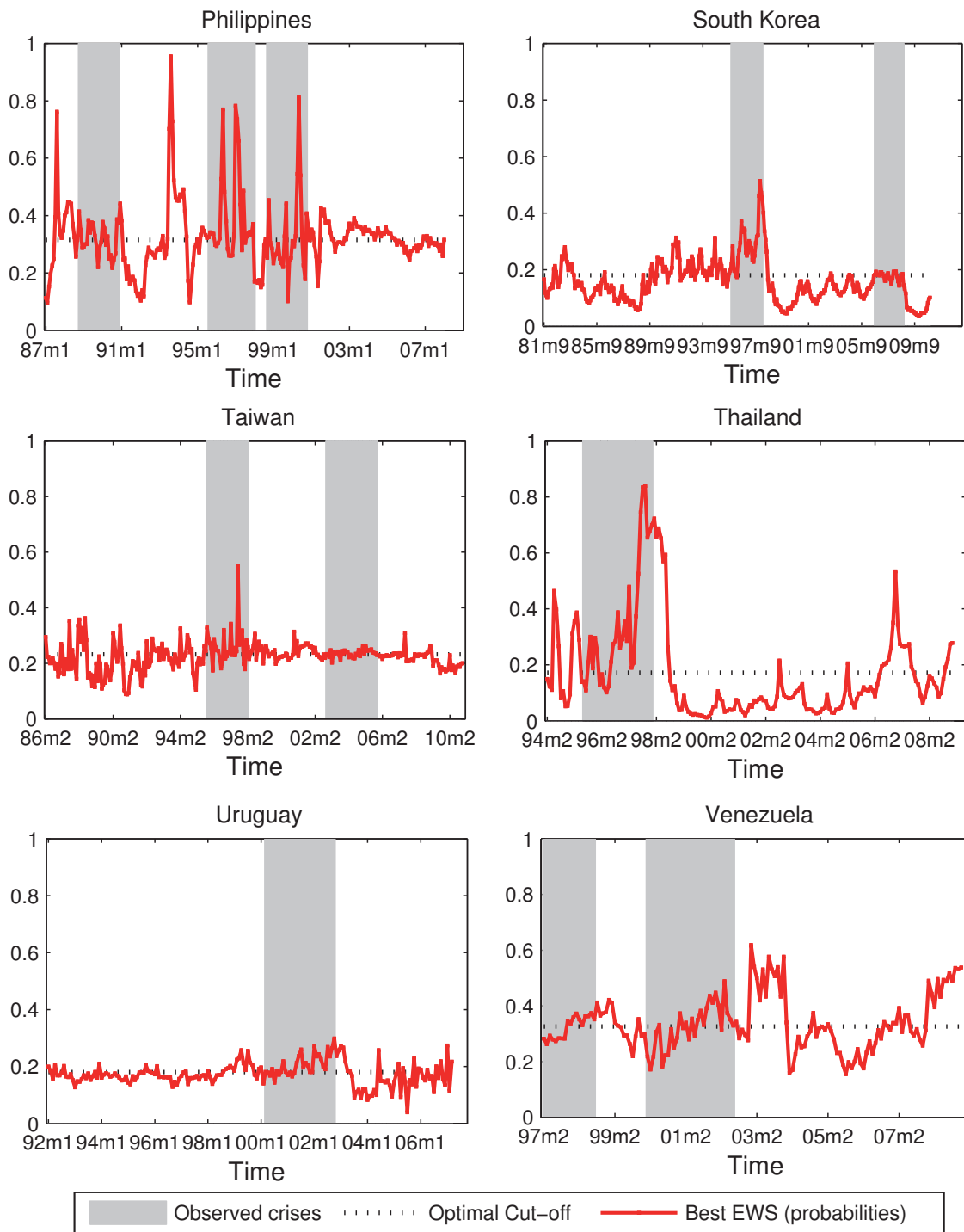


Figure 3.7: Crisis Probabilities - Regional Panel Models (continued)

We find that crisis probabilities during calm periods do not exhibit a downward trend, revealing a certain constant pressure in the exchange market. The highest pressure, as the optimal cut-off indicates, corresponds to Brazil, Peru, Philippines and Venezuela. Notice, however, that Brazil and Peru are characterized by a lower volatility than the other two countries during calm periods, while for Venezuela the model does not seem to perform too well. By contrast, the rest of the countries are characterized by a lower cut-off (around 0.2). Furthermore, the variance of the crisis probability series during observed crises is higher than the one characterizing calm periods in most countries. Among others, the recent financial crisis has left two of the twelve countries in our analysis, namely Peru, and South Korea, on the verge of a currency crisis (the *KLRm* and Zhang dating methods identify both events). At the same time, we also identify risky periods in the recent years that are not considered as *crises* by the dating method. It is particularly the case of Indonesia, Thailand, and Venezuela, for which crisis probabilities soar, indicating balance-of-payment vulnerability.

How come that for these countries our dating method does not identify a currency event after 2007? Looking at the three indicators on which the pressure index relies, *i.e.* relative changes in exchange rate, relative changes in international reserves and absolute changes in interest rate, we ascertain the idea that in these countries the drop in reserves, the exchange rate depreciation and the rise in interest rate are incomparably lesser than those registered during previous currency crises, and they are not simultaneous (except for Peru and South Korea).

The concept of resilience will be hence useful in complementing the early warning systems. A general definition of this concept relates to the ability (i) to recover quickly from a shock, (ii) to withstand the effect of a shock and (iii) to avoid the occurrence of simultaneous shocks. In the particular case of an EWS, the resilience can be defined as the deformation of the function linking the macroeconomic fundamentals and the crisis probabilities. Hence, the resilience of a country to crisis means that for constant macroeconomic fundamentals, the probability of observing a crisis diminishes either because the probabilities themselves drop or because the cut-off increases. Technically, the resilience relates to the parameter-instability issue. Since the resilience cannot be directly linked to the optimal cut-off, we propose here an indirect way to measure this concept, and illustrate it in a short example about crises in Philippines. We divide the sample in two subsamples of equal length and estimate the logit model without the yield spread on each subsample. The crisis probabilities obtained for each sub-period are then used to compute the corresponding optimal cut-off. In view of the definition of the resilience in an EWS context, we subsequently compute two sets of probabilities by relying on the two vectors of parameters (issued from the two logit models) and on the values of the

explanatory variables for the whole sample. The two crisis probabilities series obtained under this constant macroeconomic fundamentals hypothesis can thus be plotted against the observed crisis periods for the period 1987-2007 along with the two associated optimal cut-offs.

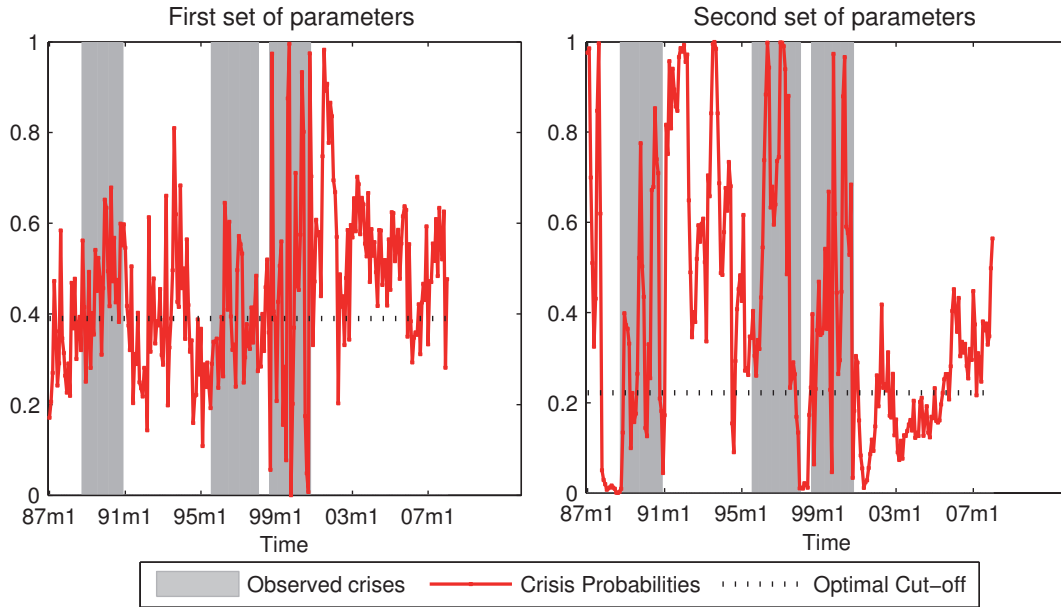


Figure 3.8: Philippines - Resilience

Let us first notice that the cut-off is smaller when using the parameters corresponding to the second subsample. The crisis periods are correctly identified in both cases as the crisis probabilities go beyond the cut-off in those periods. The calm periods are less precisely identified, even though the corresponding crisis probabilities diminish when considering the second set of parameters, especially in the second part of the sample. This reduction in probabilities provides some evidence of improvement in this country's resilience to crisis. Indeed, it is the instability of parameters of the EWS model and not an increase in the value the optimal cut-off that justifies this behavior of the resilience to crisis. A similar reasoning could be applied to the other countries in the sample. The two subsamples must, however, be defined so that they include both types of periods, which are necessary to compute the corresponding optimal cut-off.

3.5.5 Robustness Check

We now propose a sensitivity analysis of our results to the choice of crisis dating method. In a binary model (logit, probit, etc.), this choice impacts not only the evaluation results, but also the estimation of the parameters of the EWS. In the worst case scenario, we could argue that it is useless to assess the EWS forecast with respect to a dating scheme that could be invalid. The aim of our paper is neither to provide a new dating

methodology, nor to show that the *KLRm* is the best dating procedure. We consider only that it is possible to identify the crisis and the calm periods according to a binary scheme. This identification can be the result of economic experts' analyses or can be done through a pressure index approach. Whatever the methodology used, in this paper we simply assume that it is possible to well identify the crisis *ex-post*. If this assumption is not satisfied, no EWS evaluation is possible whatever the model used (logit, Markov, etc.).

Table 3.6: EWS Evaluation: Regional Panel Models (Robustness check)

	Evaluation Criteria			Comparison Tests	
	Model	QPS	AUC	CW test	W_{AUC} test
Argentina	with spread	0.487*	0.564*	2.221**	1.126
	without spread	0.500	0.508	(0.026)	(0.289)
Brazil	with spread	0.485*	0.677*	4.503***	13.83***
	without spread	0.496	0.547	(<0.001)	(<0.001)
Indonesia	with spread	0.361	0.565*	1.204	4.687**
	without spread	0.356*	0.455	(0.228)	(0.030)
Malaysia	with spread	0.419*	0.722*	5.132***	11.74***
	without spread	0.433	0.610	(<0.001)	(<0.001)
Mexico	with spread	0.339	0.438	-3.644***	15.03***
	without spread	0.334*	0.516*	(<0.001)	(<0.001)
Peru	with spread	0.396	0.382	-2.770**	19.21***
	without spread	0.375*	0.646*	(0.006)	(<0.001)
Philippines	with spread	0.334	0.540	-1.984**	16.69***
	without spread	0.308*	0.692*	(0.047)	(<0.001)
South Korea	with spread	0.271*	0.759*	5.642**	21.13***
	without spread	0.289	0.565	(<0.001)	(<0.001)
Taiwan	with spread	0.170*	0.669*	2.137**	5.676***
	without spread	0.174	0.541	(0.033)	(0.017)
Thailand	with spread	0.283*	0.895*	7.929***	45.47***
	without spread	0.396	0.559	(<0.001)	(<0.001)
Uruguay	with spread	0.280*	0.708*	3.137***	2.126
	without spread	0.317	0.604	(<0.001)	(0.145)
Venezuela	with spread	0.386	0.540*	0.734	0.036
	without spread	0.385*	0.528	(0.463)	(0.849)

Note: See note to table 3.3.

However, currency crises (and other crises) may not be precisely identified. That is why we recommend to conduct a robustness check of the evaluation procedure to the choice of the crisis dating method. Here, we propose to consider the *Zhang* dating method instead of *KLRm*. The robustness check findings for our EWS evaluation are summarized in Table 3.6. The *QPS* evaluation criterion indicates that the spread is important in

forecasting currency crises for 10 out of 12 countries, and AUC confirms these findings. As for the comparison tests, CW generally supports the alternative hypothesis that the model with spread outperforms the one without this variable (7 out of 12 countries at the 5% significance level). Instead, the W_{AUC} test rejects the null hypothesis of equal forecasting abilities for 6 countries at the 5% level in favor of the model with yield spread.

More precisely, for countries like Argentina, Indonesia, Malaysia, Peru, Philippines, South Korea, Taiwan, Thailand, Uruguay and Venezuela, the results obtained with the *Zhang* dating method go along the lines of our previous analysis, based on the *KLRm* dating method.

Nonetheless, changes in the dating method reflect in both the observed crisis series and the estimated crisis probabilities and thus in the relative comparison of the two models (with and without spread). More exactly, this time QPS and AUC favor the model without spread for three countries, *i.e.* Mexico, Peru and Philippines, while both comparison tests confirm this intuition.

Moreover, the three cut-offs considered (c_{CSA}^* , c_{AM}^* and c_{NSR}) have the same characteristics as in our previous analysis. Table 3.7 reports their descriptive statistics.

Table 3.7: EWS Optimal Cut-off: Descriptive Statistics (Robustness check)

Optimal Cut-off	Time-Series			Regional Panel		
	c_{CSA}^*	c_{AM}^*	c_{NSR}	c_{CSA}^*	c_{AM}^*	c_{NSR}
Average	0.269	0.262	0.805	0.271	0.275	0.514
Std-deviation	0.115	0.144	0.120	0.119	0.123	0.209
Minimum	0.105	0.087	0.638	0.101	0.098	0.157
Maximum	0.501	0.496	0.957	0.498	0.514	0.848
Average Sensitivity	0.733	0.834	0.112	0.603	0.611	0.028
Average Specificity	0.732	0.694	1.000	0.606	0.689	1.000

Note: See note to table 3.4.

For instance, the average NSR cut-off is at least twice the optimal ones, leading to the correct identification of all calm periods at the expense of most of the crisis ones, exactly as previously found. By contrast, the optimal cut-off lead to a better trade-off between type I and type II errors; by lowering the value of the cut-off, the number of crises correctly identified raises at a higher speed than the increase in false alarms. Our cut-offs thus lead to an average correct identification of approximatively 2/3 of the crisis and calm periods.

3.6 Conclusion

In this paper, we propose an original, model-free, evaluation toolbox for EWS. This general approach not only assesses the validity of EWS forecasts, but also allows us to compare the relative performance of alternative EWS. It is actually a two-step procedure

combining the evaluation of the competing EWS and the comparison of their forecasting abilities. We show both theoretically and empirically that the cut-off has to be taken into account in EWS evaluation since existing *QPS*-type evaluation criteria often lead to diagnostic error. More importantly, we argue that the significance of the difference in evaluation criteria for two alternative models has to be tested in a statistical framework. To this aim, we introduce several comparison tests. Then, the optimal cut-off, the one that best discriminates between crisis and calm periods (by simultaneously minimizing type I and type II errors) is identified for the outperforming model. Therefore the cut-off appears as a key element in economic actors' decisions as it labels a country as vulnerable or not at a given moment. Additionally, we assert that the optimal cut-off is different from the *NSR* one, previously used in the literature, and on top of that, it leads to a better trade-off between the two types of errors.

Our new methodology has four main advantages. First, it is model-free (it can be applied to any EWS, independent of the underlying econometric model). Second, it can be used to assess EWS for any type of crises (currency, banking, debt, etc.). Moreover, it can be used not only for an in-sample evaluation, but also to assess the out-of-sample forecasts of an EWS. Besides, it covers both the selection of the outperforming model and the crisis forecast, thus proving to be extremely useful for researchers and economic actors as well.

Applying our evaluation toolbox to a sample of twelve emerging countries from 1980 to 2010, we show that the criteria and tests including the cut-off should be favored as they allow us to better refine the forecasting abilities of EWS. Indeed, the yield spread appears to be an important indicator of currency crises in half of the countries when we rely on the *Area under the ROC comparison test*, whereas it first seems to be essential in all the countries considered (when using *QPS*-based tests like *Clark-West*). Furthermore, the optimal cut-off correctly identifies on average more than 2/3 of the crisis and calm periods, in contrast with the *NSR* one, that correctly forecasts all the calm periods at the expense of most of the crisis ones. These results contribute to the intensive discussion on the predictive power of EWS in the context of the latest crisis. Whereas Rose and Spiegel (2010, 2011) raise doubts about their utility, our findings tend to support studies showing that a set of variables (here the yield spread), which are consistent and statistically significant leading indicators for crises, exists (see Frankel and Saravelos, 2011, *inter alii*).

3.7 Appendix

3.7.1 Appendix: Comparison of ROC Curves Test

The non-parametric *test of comparison of ROC curves* has been proposed by DeLong et al. (1988). It is based on the comparison of the areas under the ROC curves associated with the two EWS models, denoted AUC_1 and AUC_2 . The null of the tests corresponds to the equality of areas under the ROC curves *i.e.*, $H_0 : AUC_1 = AUC_2$. The test statistic is defined as:

$$W_{AUC} = \frac{(AUC_1 - AUC_2)^2}{\mathbb{V}(AUC_1 - AUC_2)} \quad (3.19)$$

Under the null, it has an asymptotic $\chi^2(1)$ distribution. By definition, the asymptotic variance of the difference $\mathbb{V}(AUC_1 - AUC_2)$ is equal to:

$$\mathbb{V}(AUC_1 - AUC_2) = \mathbb{V}(AUC_1) + \mathbb{V}(AUC_2) - 2cov(AUC_1, AUC_2) \quad (3.20)$$

Each of these three elements can be estimated using a non parametric kernel estimator. Let us consider $\mathbb{V}$ the covariance matrix of the vector $(AUC_1 \ AUC_2)'$. A non-parametric kernel estimator of $\mathbb{V}$, denoted $\hat{\mathbb{V}}$, can be derived from the theory developed for generalized U-statistics by Hoeffding (1948) and Mann-Whitney statistics. Formally, we have:

$$\hat{\mathbb{V}}_{(2,2)} = (T_1)^{-1} \hat{S}_1_{(2,2)} + (T_0)^{-1} \hat{S}_0_{(2,2)} \quad (3.21)$$

where T_1 (respectively T_0) is the number of crisis (respectively calm) periods in the sample, and $\hat{S}_1$ (respectively $\hat{S}_0$) denotes the estimated variance for the crisis (respectively calm) periods.

$$\hat{S}_1 = \frac{1}{T_0^2 (T_1 - 1)} \times \sum_{i:y_i=1} \left(\begin{array}{c} \left[\sum_{j:y_j=0} K_1(\hat{p}_j, \hat{p}_i) - T_0 \times AUC_1 \right]^2 \\ \left[\sum_{j:y_j=0} K_1(\hat{p}_j, \hat{p}_i) - T_0 \times AUC_1 \right] \times \left[\sum_{j:y_j=0} K_2(\hat{p}_j, \hat{p}_i) - T_0 \times AUC_2 \right] \\ \left[\sum_{j:y_j=0} K_1(\hat{p}_j, \hat{p}_i) - T_0 \times AUC_1 \right] \times \left[\sum_{j:y_j=0} K_2(\hat{p}_j, \hat{p}_i) - T_0 \times AUC_2 \right] \\ \left[\sum_{j:y_j=0} K_2(\hat{p}_j, \hat{p}_i) - T_0 \times AUC_2 \right]^2 \end{array} \right) \quad (3.22)$$

Similarly, we have:

$$\hat{S}_0 = \frac{1}{T_1^2 (T_0 - 1)} \times \sum_{j:y_j=0} \left(\frac{\left[\sum_{i:y_i=1} K_1(\hat{p}_j, \hat{p}_i) - T_1 \times AUC_1 \right]^2}{\left[\sum_{i:y_i=1} K_1(\hat{p}_j, \hat{p}_i) - T_1 \times AUC_1 \right] \times \left[\sum_{i:y_i=1} K_2(\hat{p}_j, \hat{p}_i) - T_1 \times AUC_2 \right]} \right) \quad (3.23)$$

where $K(\cdot)$ denotes the kernel function of the estimated crisis probabilities both in crisis periods ($y_i = 1$) and calm periods ($y_j = 0$) defined by:

$$K(\hat{p}_j, \hat{p}_i) = \begin{cases} 1, & \text{if } \hat{p}_i < \hat{p}_j \\ \frac{1}{2}, & \text{if } \hat{p}_i = \hat{p}_j \\ 0, & \text{if } \hat{p}_i > \hat{p}_j \end{cases} \quad (3.24)$$

for each of the two models.

3.7.2 Appendix: Dataset

There is no official currency crisis dating method similar to the one NBER proposes for recessions. Therefore, a crisis episode is generally detected when an index of speculative pressure exceeds a certain threshold. Many alternative indexes have been developed and used for identifying currency crises. But they are all non-parametric termination rules that take into consideration the size of the movements in a combination of a number of series. Lestano and Jacobs, (2004) compare several currency crisis dating methods, aiming to identify the one that recognizes most of the crises categorized by the IMF for the 1997 Asian flu. They conclude that the KLR modified index, the Zhang original index (Zhang, 2001), and extreme values applied to the KLR modified index perform best.

Following their results, we identify crisis periods using the KLR modified pressure index (KLRm) which, unlike the KLR index, also includes interest rates:

$$KLRm_{it} = \frac{\Delta e_{it}}{e_{it}} - \frac{\sigma_e}{\sigma_r} \frac{\Delta r_{it}}{r_{it}} + \frac{\sigma_e}{\sigma_{ir}} \Delta ir_{it}, \quad (3.25)$$

where e_{it} denotes the exchange rate (*i.e.*, units of country i 's currency per US dollar in period t), $r_{n,t}$ represents the foreign reserves, while ir_{it} is the interest rate. Meanwhile, the

standard deviations σ_X are actually the standard deviations of the relative changes in the variables, $\sigma(\Delta X_{it}/X_{it})$, where X denotes each variable separately, including the exchange rate and the foreign reserves, with $\Delta X_{it} = X_{it} - X_{i,t-6}$. For the interest rate, σ_{ir} is the standard deviation of the absolute changes in interest rate. For both subsamples, the threshold equals two standard deviations above the mean:¹⁴

$$Crisis_{it} = \begin{cases} 1, & \text{if } KLRm_{it} > 2\sigma_{KLRm_{it}} + \mu_{KLRm_{it}} \\ 0, & \text{otherwise.} \end{cases} \quad (3.26)$$

To check the robustness of our results to the dating method, we also consider the *Zhang* pressure index instead of the *KLRm*. It is defined as follows:

$$Crisis_{it} = \begin{cases} 1, & \text{if } \begin{cases} \frac{\Delta e_{it}}{e_{it}} > \beta_1 \sigma'_{e_{it}} + \mu_{e_{it}} & \text{or} \\ \frac{\Delta r_{n,t}}{r_{it}} < \beta_2 \sigma'_{r_{it}} + \mu_{r_{it}} \end{cases} \\ 0, & \text{otherwise.} \end{cases} \quad (3.27)$$

where $\sigma'_{e_{it}}$ is the standard deviation of $(\Delta e_{it}/e_{it})$ in the sample of (t-36, t-1), and $\sigma'_{r_{it}}$ is the standard deviation of $(\Delta r_{it}/r_{it})$ in the sample of (t-36, t-1). The thresholds are set to $\beta_1 = 3$ and $\beta_2 = -3$. Contrary to the *KLRm* index, the interest rates are excluded from the *ZCC* and the thresholds used are time-varying for each component.

From a macroeconomic point of view, it is more important to know if there will be a crisis in a certain horizon than in a certain month, because this time period allows the state to take steps to prevent the crisis. Consequently, we define for each country $C24_t$, which corresponds to y_t from our general framework, and thus serves as the crisis dummy variable taking the value 1 if there will be a crisis in the following 24 months and 0 otherwise:

$$C24_{it} = \begin{cases} 1, & \text{if } \sum_{j=1}^{24} Crisis_{it+j} > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (3.28)$$

Several explanatory variables from three economic sectors are considered (Lestano et al., 2003) on a monthly frequency and denoted in US dollars:

1. External sector: the one-year growth rate of international reserves, the one-year growth rate of imports, the one-year growth rate of exports, the ratio of M2 to foreign reserves, and the one-year growth rate of M2 to foreign reserves.
2. Financial sector: the one-year growth rate of M2 multiplier, the one-year growth rate of domestic credit over GDP, the one-year growth rate of real bank deposits, the real

¹⁴In the case of *KLR* the threshold equals three standard deviations; however, in this case, Taiwan would never register any currency crisis, which is historically not accurate. For example, Taiwan was not exempted from the Asian crisis in 1997.

interest rate, the lending rate over deposit rate, and the real interest rate differential.

3. Domestic real and public sector: the industrial production index.

As in Kumar, (2003), we reduce the impact of extreme values by using the formula: $f(x_t) = \text{sign}(x_t) \times \ln(1 + |x_t|)$. Traditional first generation (Im, Pesaran, Shin, 2003 and Maddala and Wu, 1999) and second generation (Bai and Ng, 2000 and Pesaran, 2007) panel unit root tests are performed, leading to the rejection of the null hypothesis of stochastic trend except for the lending rate over deposit rate and industrial production index indicators. Hence, these series are substituted by their first differences.

Finally, we identify the most correlated leading indicators for each country. Two indicators are considered as being correlated for a certain country if Pearson's correlation coefficient is higher than a 30% threshold. It seems that growth of real exchange rate and real interest rate are highly correlated with most indicators for all countries, whereas the first difference of lending rate over deposit rate, the first difference of the industrial production index and yield spread are the least correlated ones with all the other indicators for all the 12 countries. The competing models are defined such that no couple of indicators is correlated in more than 4 countries. We hence identify the leading indicators by minimizing the AIC and BIC information criteria of the pooled panel data models, *i.e.*, growth of international reserves, growth of exports, growth of domestic credit over GDP, first difference of lending over deposit rate, first difference of industrial production index and yield spread. The missing values through the series are replaced using cubic splines interpolation, but when the series revealed missing values at the beginning of the sample, such as "the one-year growth of terms of trade" or "yield spread", the corresponding observations are dropped from the analysis, leading to an unbalanced panel framework. Table 3.8 shows the period covered by the leading indicators for each of the 12 countries.

3.7.3 Appendix: A Robust Estimator of the Variance of the Parameters

To compute robust estimators of the variance for logit models we use a sandwich estimator. Technically, the covariance matrix of the estimators is asymptotically equal to the inverse of the hessian matrix: $\mathbb{V}(\hat{\beta}) = -H(\hat{\beta})^{-1}$. However, this is appropriate only if we employ the real Data Generating Process (DGP). For a more permissive method from this point of view, we define the covariance matrix as follows:

$$\mathbb{V}(\hat{\beta}) = (-H(\hat{\beta})^{-1})\mathbb{V}(g(\hat{\beta}))(-H(\hat{\beta})^{-1}), \quad (3.29)$$

where $H(\hat{\beta})^{-1}$ is the inverse of the hessian matrix, and $\mathbb{V}(g(\hat{\beta}))$ is the variance of the gradient. Using the empirical variance estimator of the gradient we find that:

$$\mathbb{V}(\hat{\beta}) = -T/(T-1)H(\hat{\beta})^{-1}\left\{\sum_{t=1}^T g_t(\hat{\beta})g_t(\hat{\beta})'\right\}(-H(\hat{\beta})^{-1}), \quad (3.30)$$

Table 3.8: Database

Country	Period
Argentina	February 1994 - December 2010
Brazil	February 1995 - August 2009
Indonesia	February 1986 - August 2009
Malaysia	February 1993 - April 2009
Mexico	February 1982 - August 2009
Peru	November 1995 - January 2009
Philippines	January 1987 - February 2008
South Korea	September 1981 - December 2010
Taiwan	February 1986 - December 2010
Thailand	February 1994 - January 2009
Uruguay	January 1992 - April 2007
Venezuela	February 1997 - December 2008

Note: Data availability.

which is a robust variance estimator for the time-series model.

The main advantage of this sandwich method is that it can also be applied in the case of grouped data, as in our case. It is important to note that in the current situation each country from a cluster is a group of time-series observations that are correlated. Thus, the observations corresponding to a country are not treated as independent, but rather the countries themselves which form the clusters, are considered independent. Therefore, instead of using $g_t(\hat{\beta})$, we use the sum of $g_t(\hat{\beta})$ for each country, while T is replaced by the number of countries in a cluster. These changes ensure the independence of so-called “superobservations” entering the formula (Gould et al., 2005).

Chapter 4

Currency Crises Early Warning Systems: why they should be Dynamic¹

This paper introduces a new generation of Early Warning Systems (EWS) which takes into account the dynamics, *i.e.* the persistence in the binary crisis indicator. We elaborate on Kauppi and Saikonen (2008), which allows to consider several dynamic specifications by relying on an exact maximum likelihood estimation method. Applied so as to predict currency crises for fifteen countries, this new EWS turns out to exhibit significantly better predictive abilities than the existing models both within and out of the sample, thus vindicating dynamic models in the quest for optimal EWS.

4.1 Introduction

The recent subprime crisis has renewed the interest for Early Warning Systems (EWS). In principle, they should be able to ring before the occurrence of a financial crisis (banking, currency, debt, etc.), letting enough time for authorities to implement adequate rescuing policies to prevent or at least to smooth the perverse effects of the turmoil. Unfortunately, the existing EWS have remained silent at the edge of the recent financial crisis, leading researchers to renew their models (see Rose and Spiegel, 2010). This paper follows the same objective by emphasizing the importance of crisis dynamics for the new generation of EWS. However, in this paper we restrict our attention to currency crises.

At first sight, understanding why detecting a crisis appears so difficult is fastidious as forecasting techniques have substantially improved over the last decades. This difficulty actually lies in the specificity of EWS, that aim at accurately detecting the occurrence of a crisis, which is by essence a binary variable taking the value of one when the event occurs,

¹This chapter is based on Candelon, Dumitrescu and Hurlin (2010).

and the value of zero otherwise. In this context, it is not possible to directly implement the methods proposed in times series econometrics such as vector autoregression. Thus, following Kaminski, Lizondo and Reinhart (1998), the first EWS was elaborated for currency crises, upon a signalling approach. Using a large set of potentially informative variables, they identified a threshold beyond which a crisis is signaled. The properties of such an EWS clearly depend on this cut-off point. Kaminski et al. (1998) estimated it as the threshold value that minimizes the ratio between the number of crises incorrectly and correctly detected, also called the noise-to-signal ratio. Alternative estimation methods are also available (see Candelon et al., 2012). Once the variable specific threshold is determined, it is possible to build an aggregate indicator as a weighted combination of the variables, where each weight corresponds to the inverse of the associated noise to signal ratio. Hence, the so built EWS should exhibit a positive trend as the occurrence of a crisis increases.

Berg and Patillo (1999) proposed to use a static panel probit model as an alternative to the signalling approach. Hence, the binary crisis variable is treated as endogenous and explained by a set of macroeconomic variables. Evaluation criteria, such as the quadratic probability score (QPS) and the log probability score (LPS), indicate that their EWS exhibit better forecasting abilities (within and out of the sample) than the Kaminski et al. (1998) one. Several extensions have been proposed: Kumar et al., (2003) advocate the use of panel logit instead of panel probit. Fuertes and Kalotychou (2007) and Berg et al. (2008) analyze the presence of country clusters and their consequences for the EWS. Bussiere and Fratzscher (2006) suppose that a post-crisis specific period may be present, and consider the crisis as a ternary variable instead of a binary one, thus developing a multinomial logit EWS (Bussiere and Fratzscher 2006). Moreover, as the estimation methods for panel discrete-choice variables are quite standard and available in almost all econometric softwares, this type of EWS has been extensively implemented in applied studies.

Nevertheless, both previous EWS are static and assume that the crisis probability depends only on a set of macroeconomic variables, representing the implemented economic policies. This assumption is not supported by most empirical studies which show that the longer a country is in a crisis period, the higher the probability to exit the crisis will be, whatever the political reaction (see Tudela, 2004). Besides, Berg and Coke (2004) showed that EWS are *per nature* autoregressive, as they should ring not only one period before the occurrence of a crisis but during j periods, where j is the forecast horizon. Hence, it appears difficult for a static model to reproduce such a property.

To overcome the absence of dynamics, another mainstream of the literature proposes EWS elaborated on Markov-switching models (Abiad, 2003; Martinez-Peria, 2002;

Fratzcher, 2003). This type of EWS can take into consideration dynamic processes which are specific to the crisis or non-crisis regime. Nevertheless, these models consider a market pressure index, which is a continuous indicator of the stress faced by a country's currency. Although this approach is *per se* interesting, it has been shown by Candelon et al. (2009) that its predicting abilities are lower compared to the Berg and Patillo EWS (1999). Moreover, to the best of our knowledge, a panel version of the Markov model is not available.

Therefore, our paper proposes a new generation of EWS which reconciles the binary-choice property of the crisis variable and the dynamic dimension of this phenomenon. Particular attention is given to the specification, estimation and evaluation of such models. To be more exact, we consider not only the exogenous source of crisis persistence, *i.e.* macroeconomic variables, but also several sources of endogenous persistence of crises. Actually, the endogenous dynamics of crises can be apprehended in several ways. First, it can be included as a lagged binary crisis variable. Notice the existence of threshold effects in this case, as a crisis is identified only if the index goes beyond a certain threshold. Second, dynamics can be introduced via the past index associated with the probability of being in a crisis regime. Thus, the EWS to be estimated relies on an autoregressive (AR) model, where the lagged index summarizes all the past information of the system. Finally, both types of dynamics can be simultaneously considered. Given all these different specifications, the estimation methodology proposed should be flexible enough to allow for specification tests. We thus rely on the recent paper of Kauppi and Saikonen (2008) that proposes an exact Maximum Likelihood estimation fitted to all these time-series models.²

Our dynamic EWS framework has several advantages. First, it allows to easily estimate and compare several binary EWS specifications, the static one included. Indeed, beyond being easy to program in most common econometric softwares and not time intensive (results are obtained in a few second), this framework allows to detect the EWS specification having the best forecasting abilities both in-sample and out-of-sample. Second, it can be applied to estimate EWS for any type of crises, not only currency crises as in our application, provided that the crisis indicator is dichotomic. Besides, our specifications allow for the presence of past macro-economic variables representing the economic policies experienced by a certain country, source of exogenous crisis persistence. Most importantly, since panel analysis has been privileged in currency-crisis literature, we extend this time-series framework to a dynamic fixed effects panel EWS by elaborating on Carro (2007).³

²A previous attempt to estimate one specific dynamic specification has been proposed by Falcetti and Tudela (2006) using a smoothly simulated likelihood estimation.

³All Matlab codes are available from the authors upon request.

In an empirical analysis, we aim to build a currency crisis EWS for a sample of fifteen emerging countries. The predictive abilities (within and out of the sample) of this new EWS compared to a wide range of alternative EWS (in particular the Markov-switching and the static logit models) are then investigated using the unified evaluation framework proposed by Candelon et al. (2012). Anticipating on our results, it turns out that the dynamic model including the lagged binary dependent variable outperforms (in-sample and out-of sample) the other specifications. Moreover, it appears that our new EWS has good out-of-sample forecasting abilities even when the forecast horizon increases. More precisely, considering a 24 months crisis variable, our dynamic EWS identifies correctly most of the out-of-sample crisis and calm periods for most of the countries. All in all, it turns out that the dynamic logit EWS including the lagged binary crisis indicator outperforms its main competitors, namely the static logit and the Markov-switching, thus vindicating dynamic specifications in the quest for the optimal currency-crisis EWS model.

This paper is structured as follows: the new estimation methods used for dynamic binary-choice EWS both in time-series and panel are presented in section 4.2. The database, the currency crisis dating methods as well as the estimation results are scrutinized in section 4.3. Section 4.4 proposes both a within-sample and an out-of-sample comparison of the forecasting abilities of the models, while section 4.5 concludes.

4.2 A Dynamic Specification of EWS

To date, almost all EWS models are static and do not exploit the persistence property of the crisis captured by lagged endogenous indicators. In a more general context, this issue goes along the lines of a well-known econometric debate on the exogenous/endogenous persistence of a phenomenon. This paper is hence the first one to consider a dynamic version of EWS based on an exact maximum likelihood estimation by relying on Kauppi and Saikkonen (2008).

4.2.1 Specification and Estimation

Let us consider first the time-series version of the dynamic binary-choice EWS. We denote by $\{y_{n,t}\}_{t=1}^T$ the currency crisis binary variable for country n , taking the value 1 during crisis periods and 0 otherwise, and by $\{x_{n,t}\}_{t=1}^T$ the matrix of explanatory variables, whose first column is a unit vector. For ease of computation, hereafter n will be omitted.

The one-step-ahead dynamic specification accounting both for the influence of the

lagged binary variable and that of the lagged index takes the form of:⁴

$$\Pr_{t-1}(y_t = 1) = F(\pi_t) = F(\delta\pi_{t-1} + \alpha y_{t-1} + x_{t-1}\beta), \quad (4.1)$$

where $\Pr_{t-1}(y_t = 1)$ is the conditional probability given the information set we have at our disposal at time $t - 1$, π_t is the index at time t , and F is a distribution function, *i.e.* gaussian in the case of the probit model and logistic for the logit model. In this paper we use the heavy-tailed logistic distribution, considered to be more appropriate for the study of extreme events such as crises. It is worth nothing that this method can be applied so as to construct EWS for any type of crisis (banking, currency, sovereign debt, etc.) as long as the dependent variable is dichotomic.

The main advantage of this general framework is that it allows to estimate and then compare different alternative specifications in an exact maximum likelihood framework. More precisely, we first consider the pure static model, as in the rest of the literature in which the occurrence of currency crises is explained only by exogenous macroeconomic variables (x_{t-1}). This constitutes the benchmark model, in which crises are persistent only if the changes in economic indicators are themselves persistent (exogenous persistence). At the same time, we introduce three dynamic specifications, in which we also allow for endogenous persistence of crisis.

Indeed, a dynamic model including the lagged value of the binary dependent variable y_{t-1} is proposed. In this case, we can assess the impact of the regime prevailing in the previous period on the crisis probability. Note the existence of threshold effects, as the index must go beyond a certain threshold to set off a crisis in the previous period. Next, a dynamic model including the lagged index π_{t-1} is implemented. This time an increase in the index is linearly transmitted to the next period, hence always increasing the probability of crisis. Finally, the most complex dynamic model, including both the lagged dependent variable y_{t-1} and the lagged index π_{t-1} is estimated. Nevertheless, one should not lose sight of the fact that since in the last two models δ is an autoregressive parameter, it has to satisfy the usual stationarity condition, *i.e.* the roots of the corresponding polynomial lie outside the unit circle. Otherwise, the crisis becomes perpetual, which is counterintuitive. To tackle this problem, a constrained maximum likelihood estimation is implemented and described in Appendix 3.7.1. In fact, the log-likelihood function takes the general form of:

$$\text{LogL}(\theta) = \sum_{t=1}^T l_t(\theta) = \sum_{t=1}^T [y_t \log F(\pi_t(\theta)) + (1 - y_t) \log(1 - F(\pi_t(\theta)))], \quad (4.2)$$

⁴It can be noted that the model for h -step-ahead forecasts can be obtained by repetitive substitutions. For more details, see Kauppi and Saikkonen (2008). Besides, different lags of the exogenous variables can be considered, but in this paper we restrict our attention to only one lag ($t - 1$) as it is generally done in the literature.

where θ is the vector of parameters. The ML estimators have the desired large-sample properties. Besides, we tackle the autocorrelation problem induced by the construction of a j months ahead crisis variable by considering a Gallant correction for the covariance matrix. Given the maximum-likelihood framework, these dynamic time-series models are easy to implement using any existing econometric software.

4.2.2 Panel Framework

Instead of considering EWS for individual countries and hence applying a time-series approach, several papers (Berg and Patillo, 1998; Kumar et al., 2003) favor a panel data approach (with fixed effects) by pooling the information available in several countries. In this section we thus propose a dynamic version of the fixed effects panel model by elaborating on Carro (2007). Note that this framework can embody all the dynamic specifications proposed in time-series. This constitutes an important contribution of our paper to the EWS literature.

Let us consider the panel version of the general dynamic binary-choice EWS:

$$\Pr(y_{it} = 1) = F(\delta\pi_{t-1} + \alpha y_{it-1} + \beta x_{it-1} + \eta_i), \text{ for } t = 0, 1, 2, \dots, T, \text{ and } i = 1, 2, \dots, N, \quad (4.3)$$

where N is the number of individuals in the panel, T represents the number of time series observations for each individual, and η_i accounts for the permanent unobserved heterogeneity between individuals. Since we do not impose any distributional assumption to $\{\eta_i\}_{i=1}^N$, they are treated as parameters to be estimated, and our approach is one with fixed effects. Besides, we assume no cross-sectional dependence. The dependent variable y_{it} equals 1 if there is a crisis at time t and 0 in the opposite case. Moreover, x_{it-1} represents the matrix of explicative variables. The log-likelihood of the model conditioned on the first observation, often called concentrated likelihood takes the following form:

$$\text{LogL}(\theta, \eta_i) = \sum_{i=1}^N \text{LogL}_i(\theta, \eta_i) = \sum_{i=1}^N \sum_{t=1}^T [y_{it} \log(F_{it}) + (1 - y_{it}) \log(1 - F_{it})], \quad (4.4)$$

where θ represents the vector of parameters. As usual, the estimated parameters maximize the log-likelihood function, which means that they solve the first order conditions with respect to θ and with respect to η_i . Most importantly, the estimation of θ depends on $\hat{\eta}_i$, which means that $\hat{\theta}$ is a convergent estimator of θ_0 only when $\hat{\eta}_i$ is a convergent estimator of η_{i0} , that is when $T \rightarrow \infty$. Thus, the central issue here, as in any non-linear panel model with fixed effects, is how to deal with this incidental parameters problem.

The solution proposed by Carro (2007) actually consists in a numerical substitution of the fixed effects (η_i) in the estimation of θ . Thus, at each step N non-linear equations are solved so as to estimate $\{\hat{\eta}_i\}_{i=1}^N$ by using $\hat{\theta}$ obtained at the previous step and then

the estimated values of $\hat{\eta}_i$ are introduced into the first order condition corresponding to the concentrated likelihood so as to estimate $\hat{\theta}$. To be more exact, the estimation of η is nested in the algorithm that maximizes the concentrated log-likelihood, so that at each iteration $N + 1$ non-linear optimizations are realized using the Gauss-Newton algorithm (the first N ones correspond to the fixed effects, while the last one corresponds to the θ parameters).

Moreover, in order to reduce the estimation bias from $O(T^{-1})$ to $O(T^{-2})$ without increasing the asymptotic variance Carro proposed a modification of the first order condition expressed in terms of the original parameters of the model (hereafter MMLE). Consequently, the modified score for a certain country takes the following form:

$$d_{\theta Mi}(\theta) = d_{\theta Ci}(\theta, \hat{\eta}_i(\theta)) - \frac{1}{2} \frac{1}{d_{\eta \eta i}(\theta, \hat{\eta}_i(\theta))} \left(d_{\theta \eta \eta i}(\theta, \hat{\eta}_i(\theta)) + d_{\eta \eta \eta i}(\theta, \hat{\eta}_i(\theta)) \frac{\partial \hat{\eta}_i(\theta)}{\partial \theta} \right) + \frac{\partial / \partial \eta_i (\mathbb{E}[d_{\theta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i])}{\mathbb{E}[d_{\eta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]} \Big|_{\eta_i = \hat{\eta}_i(\theta)} - \frac{\mathbb{E}[d_{\theta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]}{\mathbb{E}[d_{\eta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]} \Big|_{\eta_i = \hat{\eta}_i(\theta)} * \frac{\partial / \partial \eta_i (\mathbb{E}[d_{\eta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i])}{\mathbb{E}[d_{\eta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]} \Big|_{\eta_i = \hat{\eta}_i(\theta)}, \quad (4.5)$$

where $d_{\theta Ci}(\theta, \hat{\eta}_i(\theta))$ is an individual's score from the concentrated likelihood (hereafter MLE):

$$d_{\theta Ci}(\theta, \hat{\eta}_i(\theta)) = \frac{y_{it} - F_{it}(\theta, \eta_i)}{F_{it}(\theta, \eta_i)(1 - F_{it}(\theta, \eta_i))} \left(y_{it} + \frac{\partial \hat{\eta}_i(\theta)}{\partial \theta} \right). \quad (4.6)$$

The MMLE first order condition corresponding to the entire panel can be obtained by adding the individual MMLE scores. At the same time, the corresponding standard errors can be easily obtained from the principal diagonal of the covariance matrix, which is given by the inverse of the Hessian accounting for the fixed effects. The computation of the modified score and Hessian matrix for a logit model in the case of the second dynamic specification, *i.e.* the one including the lagged binary variable, which corresponds to the fixed-effects panel model estimated in the empirical application, is detailed in Appendix 4.6.2.

The main advantage of Carro's (2007) estimation method consists in its simplicity of implementation since it is based on the first derivatives of the log-likelihood function. On top of that, it allows for the estimation of all the dynamic specifications introduced in the previous subsection. More exactly, the same computational trick as in Appendix 4.6.1 can be used to assure the stationarity of the autoregressive parameter δ). Finally, given the reduction of the bias, the estimators have good asymptotic properties.

4.3 Dynamic EWS Estimation

To prove the importance of crisis dynamics in the specification of an EWS, in this section we tackle the estimation results of our dynamic models relatively to that of existing EWS models. Recall that the dynamic EWS models we propose can be estimated for all the types of crises, *i.e.* banking, currency, debt, etc., as long as the dependent variable is dichotomic. In this application, however, we restrict our attention to currency crises. Nevertheless, before starting the analysis *per se*, let us present some data-related issues.

4.3.1 Dataset

Monthly data expressed in US dollars covering the period 1985-2008 for 15 emerging countries⁵ have been extracted from the IMF-IFS database and the national banks of the countries under analysis via Datastream. Several explanatory variables from two economic sectors were selected (see Candelon et al. 2012, Berg et al., 2008, Lestano et al., 2003).

1. External sector: the one-year growth rate of international reserves, the one-year growth rate of imports, the one-year growth rate of exports, the ratio of M2 to foreign reserves, and the one-year growth rate of M2 to foreign reserves.

2. Financial sector: the one-year growth rate of M2 multiplier, the one-year growth rate of domestic credit over GDP, real interest rate, and real exchange rate overvaluation.

As in Kumar (2003), we treat outliers by dampening every variable using the formula: $f(x_t) = \text{sign}(x_t) * \ln(1 + |x_t|)$, so as to reduce the impact of extreme values. Traditional first generation (Im, Pesaran, Shin, 2003) as well as (Maddala and Wu 1999) and second generation (Bai and Ng, 2001 and Pesaran, 2007) panel unit root tests are performed, leading to the rejection of the null hypothesis of stochastic trend for all explanatory variables. Besides, the gaps through the series are replaced with the mean value of each series.

Since we aim to evaluate the forecasting abilities of dynamic logit models, we proceed to a general selection from the aforementioned exogenous variables, leading to the choice of only two macroeconomic variables. To be more exact, this selection is based on previous results found in the literature, on the correlation between the indicators, as well as on the explanatory power of each variable. The first lag of these variables is then introduced into the models as a control variable, namely the one-year growth rate of international reserves and the one-year growth rate of M2 to foreign reserves.

⁵Argentina, Brazil, Chile, Indonesia, Israel, Malaysia, Mexico, Morocco, Peru, Philippines, South Korea, Turkey, Thailand, Uruguay and Venezuela.

4.3.2 Dating Currency Crises

The most common method leading to the identification of currency crisis periods implies the computation of an index of speculative pressure. If this index exceeds a certain threshold, a crisis episode is identified. As in Candelon et al. (2012), we base our choice on the results of Lestano and Jacobs (2004). We hence identify crisis periods using the KLR modified pressure index (KLRm) which, unlike the KLR index, also includes interest rates:

$$\text{KLRm}_{n,t} = \frac{\Delta e_{n,t}}{e_{n,t}} - \frac{\sigma_e}{\sigma_r} \frac{\Delta r_{n,t}}{r_{n,t}} + \frac{\sigma_e}{\sigma_i} \Delta i_{n,t}, \quad (4.7)$$

where $e_{n,t}$ denotes the exchange rate (*i.e.*, units of country n 's currency per US dollar in period t), $r_{n,t}$ represents the foreign reserves of country n in period t , while $i_{n,t}$ is the interest rate in country n at time t . Meanwhile, the standard deviations σ_X are actually the standard deviations of the relative changes in the variables $\sigma(\Delta X_{n,t}/X_{n,t})$, where X denotes each variable separately, including the exchange rate and the foreign reserves, with $\Delta X_{n,t} = X_{n,t} - X_{n,t-6}$. Besides, we use the standard deviations of the absolute changes in interest rates $\sigma_{\Delta i_{n,t}}$.⁶ The threshold equals two standard deviations above the mean, so that the crisis at time t indicator is given by:

$$C1_{n,t} = \begin{cases} 1, & \text{if } \text{KLRm}_{n,t} > 2\sigma_{\text{KLRm}_{n,t}} + \mu_{\text{KLRm}_{n,t}} \\ 0, & \text{otherwise.} \end{cases} \quad (4.8)$$

From macroeconomic point of view, it is more important to know if there will be a crisis in a certain horizon than in a certain month, because this time period allows the authorities to take steps to prevent the crisis. Consequently, we also check the out of sample forecasting abilities of a $C24_t$ variable (hereafter $C24$ variable), which serves as the crisis dummy variable for each country, taking the value of 1 if there will be at least one crisis in the following 24 months and 0 otherwise:

$$C24_{n,t} = \begin{cases} 1, & \text{if } \sum_{j=1}^{24} C1_{n,t+j} > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (4.9)$$

Note that the variables $C1$ and $C24$ correspond to y_t from our general framework.

4.3.3 Optimal Country Clusters

As Berg et al. (2008) have pointed out, pooling all available countries into one panel model might not be the best alternative especially in terms of forecasting abilities of the model. Indeed, a major issue in panel analysis is related to the heterogeneity of

⁶Additionally, we take into account the existence of higher volatility in periods of high inflation, and consequently the sample is split into high and low inflation periods. The cut-off corresponds to a six month inflation rate higher than 50%.

slope parameters. Nevertheless, a viable alternative to time-series estimation might be represented by a panel including only poolable countries. To be more precise, by poolable countries we mean a group of countries for which the slope parameters corresponding to the time-series models are statistically equal to the ones of a panel model including the same group of countries, *i.e.*, $\beta_i = \beta_p$, where β_i is the vector of parameters for country i and β_p is the vector of parameters corresponding to the panel model. It is Kapetanios (2003) who proposed a sequential procedure based on an Hausman type statistic that tests the homogeneity of parameters between different countries grouped together in the same panel. Thus, it allows to isolate country clusters for which the null hypothesis of homogeneity of parameters cannot be rejected.

Following their recommendations, we apply the dynamic panel model on two optimal clusters (11 and respectively 2 countries out of 15) which are identified by using Kapetanios's methodology (Kapetanios, 2003). For the two non-poolable countries (Israel and South Korea) only time-series models are estimated.

4.3.4 Estimation Results

As aforementioned, the main advantage of this estimation framework is that it allows to estimate and compare different EWS specification in an exact maximum likelihood framework. We thus estimate the three types of dynamic EWS we propose, as well as the benchmark, *i.e.* the static EWS model for each of the 15 countries under analysis in the time-series framework. More precisely, we consider the static model (labeled Model 1), a dynamic one including the *lagged binary* dependent variable (labeled Model 2), a dynamic one including the *lagged index* (called Model 3) and, last but not least, a dynamic model which includes both the *lagged binary* dependent variable and the *lagged index* (called Model 4). We also remind that the dependent variable is $C1$ (crisis at time t).

In a first step, we select the most parsimonious dynamic specification by relying on the Schwarz Information Criterion (hereafter SBC). This goodness of fit indicator reveals that the right-hand-side variables have important explanatory power especially when the lagged dependent variable and/or the lagged index are present in the model, *i.e.* the dynamic models (see Table 4.1).

More specifically, the lowest values of the SBC criterion correspond to the dynamic models and in particular to Model 2, *i.e.* including the lagged crisis indicator, which seems to be the most adequate dynamic specification for most of the countries (8 out of 15).⁷ To put it another way, the goodness of fit indicator is a clear indication of the fact that dynamic specifications generally outperform the static one. Nevertheless, the best model for estimation is not necessarily the outperforming one in forecasting. A proper

⁷Nevertheless, the third model seems better for countries like Brazil and Thailand, whereas the static model seems more parsimonious than the dynamic specifications for the countries registering a very small number of crisis periods (only one or two periods), countries for which no model works well since we face a rare event data problem.

Table 4.1: SBC information criterion (time-series logit models)

Country	Model 1	Model 2	Model 3	Model 4
	SBC	SBC	SBC	SBC
Argentina	57.62	57.47	63.07	62.72
Brazil	88.39	87.72	87.20	91.94
Chile	49.03	54.53	54.81	58.68
Indonesia	54.90	49.80	58.70	55.05
Israel	26.42	30.49	31.88	35.14
South Korea	16.75	22.33	22.33	31.44
Malaysia	50.16	40.35	54.51	45.66
Mexico	101.1	67.70	105.4	72.29
Marocco	27.86	33.43	33.60	39.15
Peru	62.50	53.73	66.37	59.22
Philippines	75.14	70.75	76.82	75.00
Thailand	33.21	32.47	30.43	35.30
Turkey	44.78	50.04	45.13	48.86
Uruguay	62.06	58.89	66.65	64.35
Venezuela	74.60	68.01	82.78	73.48

Note: Model 1 is the static model (the benchmark), considering that crises occurrence can be explained only by macroeconomic factors. Models 2 to 4 are dynamic, including the lagged binary crisis variable (Model 2), the lagged index (Model 3), and both of them respectively (Model 4), apart from macroeconomic leading indicators. Bold values correspond to the best model according to SBC for each country.

statistical assessment framework for the forecasting performance of static and dynamic models needs hence to be implemented, which is done in the next section.

In a second step, we analyze the signs of the estimated parameters for the best dynamic model, *i.e.*, Model 2, which includes the lagged binary crisis indicator (see Tables 4.2, 4.3, 4.4 and 4.5 which show the results of the ML estimates for the four EWS specifications considered).

These results roughly correspond to our a priori expectations. Indeed, if an increase in a country's *growth of international reserves* indicator is observed at a certain moment in time, a decline in the probability of occurrence of currency crises is presumed, since it is perceived as an indicator of currency non-vulnerability, *i.e.*, a negative coefficient of the *growth of international reserves* is awaited. Besides, the probability of currency-crisis emergence is supposed to escalate if an expansion of the *growth of M2 to reserves* is noticed in the previous period. To be more exact, if the growth of the amount of money in circulation overruns the *growth of international reserves*, the currency is perceived as unstable and a speculative attack is foreseeable. Thus, a positive coefficient of the *growth of M2 to reserves* is expected. Nonetheless, several countries register strange signs, *i.e.* negative and significant coefficients, generally due to the fact that the two macroeconomic variables capture mainly the information not filtered by the lagged binary variable. Most importantly, the coefficient of the lagged binary dependent variable is most of the time significant and has a positive sign (except for the countries for which the crisis lasts only

one period). It follows that the probability of being in a crisis increases if a crisis period prevailed in the previous period. This clearly indicates that crisis' persistence should be accounted for in order to improve currency crisis EWS.

Since the second model appears to outperform the other dynamic specifications, we opt for the dynamic panel methodology in the form including the *lagged binary* dependent variable along with the selected macroeconomic indicators. We actually estimate two models; the first one uses the whole sample of countries (pooled model) while the other one relies only on the poolable ones (optimal country clusters). The estimation results for these dynamic panel logit models with fixed effects are reported in Table 4.6.

It can be noticed that the estimated coefficients for the growth of international reserves and the lagged binary variable are significant for all the significance levels, *i.e.*, 1%, 5% and 10%. Moreover, their signs are similar from one model to another, and go along the lines of the time-series analysis, confirming the economic intuition that a higher growth of international reserves lowers the crisis probability. Similarly, problems in the balance-of-payment in the previous period, synthesized by a lagged crisis indicator equal to 1, tends to increase the probability of continuing in a crisis. On the contrary, the M2 to reserves indicator is generally not significant.

Table 4.2: Estimation results (time-series logit models)

Country	Indicator	Model 1	Model 2	Model 3	Model 4
Argentina	Intercept	-4.905*** (1.117)	-4.828*** (0.809)	-4.432*** (1.158)	-4.184*** (1.047)
	Lagged binary variable		2.527** (1.062)		2.384** (0.915)
	Growth of international reserves	-7.703** (3.783)	-5.398* (2.676)	-7.135*** (3.266)	-4.832** (2.255)
	Growth of M2 to reserves	-1.602 (1.051)	-1.156 (0.806)	-1.394 (0.955)	-0.880 (0.693)
	Lagged index			0.106 (0.107)	0.143 (0.135)
Brazil	Intercept	-3.293*** (0.481)	-3.579*** (0.423)	-0.621*** (0.167)	-0.808*** (0.246)
	Lagged binary variable		2.363*** (0.874)		0.655 (0.596)
	Growth of international reserves	-3.289 (1.997)	-2.433 (1.753)	-1.087*** (0.285)	-1.034*** (0.330)
	Growth of M2 to reserves	-0.047 (0.069)	-0.042 (0.057)	0.017*** (0.004)	0.018*** (0.004)
	Lagged index			0.835*** (0.039)	0.794*** (0.060)
Chile	Intercept	-4.413*** (0.747)	-4.405*** (0.758)	-2.104*** (0.812)	-0.984** (0.478)
	Lagged binary variable		-4.140*** (1.397)		1.613* (0.900)
	Growth of international reserves	-1.447 (3.754)	-1.395 (3.839)	-0.086 (0.312)	0.788*** (0.282)
	Growth of M2 to reserves	-3.029 (2.303)	-3.051 (2.359)	-0.993 (0.968)	0.090 (0.091)
	Lagged index			0.531*** (0.158)	0.809*** (0.098)
Indonesia	Intercept	-5.583*** (0.927)	-5.640*** (0.913)	-8.501*** (2.009)	-6.971*** (1.334)
	Lagged binary variable		3.961*** (1.454)		4.246** (1.886)
	Growth of international reserves	-2.738 (3.912)	-0.086 (3.967)	-5.183 (8.408)	-0.537 (5.706)
	Growth of M2 to reserves	-10.12*** (2.608)	-6.355*** (1.938)	-15.45*** (5.029)	-8.574*** (3.224)
	Lagged index			-0.481*** (0.083)	7-0.207 (0.169)

Note: Continued on the next page.

Table 4.3: Estimation results (time-series logit models) - continued

Country	Indicator	Model 1	Model 2	Model 3	Model 4
Israel	Intercept	-8.228*** (1.418)	-23.07** (11.063)	-9.323*** (2.382)	-37.22 (65.094)
	Lagged binary variable		-10.10 (6.360)		-20.65 (37.37)
	Growth of international reserves	-54.41*** (10.26)	-209.8* (112.3)	-59.41*** (10.89)	-374.2 (669.5)
	Growth of M2 to reserves	-22.64*** (4.615)	-89.33* (47.22)	-23.81*** (5.249)	-168.5 (304.3)
	Lagged index			-0.124 (0.155)	0.157** (0.079)
South Korea	Intercept	-33.65*** (5.959)	-41.55*** (7.351)	-24.83 (68.57)	-303.6 (25930)
	Lagged binary variable		76.76*** (17.19)		466.7 (39249)
	Growth of international reserves	-265.9*** (48.85)	-266.6*** (40.94)	-314.2 (1158)	-1033 (82816)
	Growth of M2 to reserves	-176.9*** (31.83)	-31.05*** (9.330)	-131.0 (385.3)	-178.4 (20333)
	Lagged index			(0.511) (0.397)	(-0.623) (11.96)
Malaysia	Intercept	-4.253*** (0.760)	-5.246*** (1.039)	-5.952*** (1.144)	-6.177*** (1.197)
	Lagged binary variable		6.092*** (1.752)		6.596*** (2.035)
	Growth of international reserves	-12.78*** (2.759)	-5.090 (4.138)	-17.21*** (4.967)	-6.822 (5.363)
	Growth of M2 to reserves	-5.640*** (1.723)	-0.448 (1.895)	-8.381** (3.447)	-1.656 (1.985)
	Lagged index			-0.374** (0.150)	-0.188*** (0.071)
Mexico	Intercept	-3.188*** (0.681)	-4.343*** (0.684)	-4.290*** (0.801)	-5.200*** (0.883)
	Lagged binary variable		5.927*** (1.893)		6.932*** (2.639)
	Growth of international reserves	-5.135 (3.298)	-0.746 (2.459)	-6.857 (5.033)	-1.221 (3.665)
	Growth of M2 to reserves	-2.543 (1.983)	1.173 (3.662)	-3.452 (3.510)	1.383 (4.995)
	Lagged index			-0.329 (0.242)	-0.196*** (0.066)

Note: Continued on the next page.

Table 4.4: Estimation results (time-series logit models) - continued

Country	Indicator	Model 1	Model 2	Model 3	Model 4
Morocco	Intercept	-6.812*** (1.020)	-6.809*** (1.033)	-1.899*** (0.298)	-1.530*** (0.189)
	Lagged binary variable		-0.514 (1.866)		-2.588*** (0.769)
	Growth of international reserves	-6.700*** (1.422)	-6.699*** (1.421)	-1.473** (0.687)	-1.116** (0.506)
	Growth of M2 to reserves	-17.80*** (2.960)	-17.79*** (2.960)	-4.971*** (1.687)	-3.897*** (1.240)
	Lagged index			0.729*** (0.042)	0.777*** (0.025)
Peru	Intercept	-5.151*** (0.961)	-5.939*** (0.840)	-2.787*** (0.709)	-6.793*** (2.993)
	Lagged binary variable		3.264*** (0.999)		3.600*** (1.341)
	Growth of international reserves	-13.23*** (2.519)	-11.77*** (2.808)	-7.706*** (1.626)	-13.44* (7.037)
	Growth of M2 to reserves	-0.476 (0.696)	-0.574 (0.712)	-0.439 (0.253)	-0.525 (0.941)
	Lagged index			0.477*** (0.094)	-0.144 (0.493)
Philippines	Intercept	-3.300*** (0.532)	-3.754*** (0.534)	-5.178*** (1.319)	-5.647*** (1.590)
	Lagged binary variable		2.967*** (1.131)		2.832** (1.299)
	Growth of international reserves	-10.595*** (3.192)	-7.135** (3.261)	-16.586*** (5.407)	-15.833** (6.510)
	Growth of M2 to reserves	-5.645*** (1.427)	-3.280** (1.296)	-8.879*** (2.702)	-6.493*** (2.470)
	Lagged index			-0.473*** (0.121)	-0.436 (0.233)
Thailand	Intercept	-6.387*** (1.545)	-8.068*** (2.776)	-13.81 (32.12)	-12.30 (24.86)
	Lagged binary variable		5.497*** (1.496)		2.514*** (1.205)
	Growth of international reserves	-40.56*** (11.65)	-35.99 (20.73)	-113.1 (266.6)	-87.37 (208.7)
	Growth of M2 to reserves	5.873** (2.790)	10.84** (4.866)	37.13 (90.47)	29.79 (70.38)
	Lagged index			0.579*** (0.043)	0.525*** (0.180)

Note: Continued on the next page.

Table 4.5: Estimation results (time-series logit models) - continued

Country	Indicator	Model 1	Model 2	Model 3	Model 4
Turkey	Intercept	-5.721*** (1.627)	-5.556*** (1.651)	-10.21 (6.460)	-10.90 (6.817)
	Lagged binary variable		0.767 (1.609)		2.253 (1.507)
	Growth of international reserves	-15.11*** (3.997)	-13.43** (5.899)	-29.89 (22.26)	-27.99 (24.67)
	Growth of M2 to reserves	-4.562** (2.045)	-4.241* (2.545)	-7.292 (4.589)	-7.443 (6.055)
	Lagged index			-0.502*** (0.095)	-0.617*** (0.166)
Uruguay	Intercept	-4.717*** (0.693)	-5.053*** (0.634)	-7.485*** (2.309)	-4.464*** (0.931)
	Lagged binary variable		2.971*** (1.046)		2.923*** (0.926)
	Growth of international reserves	-10.17*** (2.638)	-7.704*** (1.776)	-15.731*** (4.845)	-6.640*** (2.196)
	Growth of M2 to reserves	-2.336 (2.404)	-2.803 (1.735)	7-3.053 (2.947)	-3.017** (1.521)
	Lagged index			-0.537** (0.246)	0.136 (0.159)
Venezuela	Intercept	-6.088*** (1.111)	-5.914*** (1.196)	-1.427*** (0.389)	-5.304*** (1.309)
	Lagged binary variable		3.470** (1.432)		3.370*** (1.210)
	Growth of international reserves	-15.48*** (4.092)	-11.72*** (4.059)	-4.779*** (1.233)	-10.33*** (3.884)
	Growth of M2 to reserves	-2.164** (0.940)	-0.235 (1.314)	0.359*** (0.115)	0.017 (1.338)
	Lagged index			0.804*** (0.037)	0.103 (0.207)

Note: Model 1 is the static model (the benchmark), considering that crises occurrence can be explained only by macroeconomic factors. Models 2 to 4 are dynamic, including the lagged binary crisis variable (Model 2), the lagged index (Model 3), and both of them respectively (Model 4), apart from macroeconomic leading indicators. Robust standard errors are reported in the parentheses. The asterisks ***, **, and * denote significance at the 1%, 5% and 10% significance level.

Table 4.6: Estimation results (panel logit models)

Indicator	Coefficients		
	All countries	Poolable countries (cluster1)	Poolable countries (cluster2)
Lagged binary variable	4.383*** (0.304)	4.294*** (0.332)	3.608*** (0.955)
Growth of international reserves	-4.092*** (0.665)	-3.614*** (0.681)	-7.496*** (2.695)
Growth of M2 to reserves	-0.542* 0.298	-0.550* (0.325)	-0.459 (0.776)

Note: The first model is the pooled panel one, that includes all countries. The second one corresponds to the first optimal cluster (11 countries), while the last one is for the second optimal cluster (2 countries). Israel and South Korea are not included in either of the optimal clusters. The standard errors are reported in the parentheses. The asterisks ***, **, and * denote significance at the 1%, 5% and 10% significance level.

4.4 Forecasts Evaluation

So far we have seen that accounting for the currency crisis dynamics matters, more exactly, we have proved that the introduction of the lagged binary dependent variable into the model improves the estimation of currency crises probabilities. In this section we go one step further and statistically test the in-sample one-step-ahead forecasting abilities of the static and dynamic currency crisis EWS models. Most importantly, the out-of-sample one-step ahead predictive abilities of the best model are checked. To do this, we apply the validation methodology developed in Candelon et al. (2012) (see Appendix 4.6.3). Finally, a robustness check is performed by considering one-step-ahead out-of-sample forecasts of the 24 months crisis variable C_{24} .

4.4.1 In-Sample Analysis

To check the within sample forecasting abilities of the static and dynamic time-series models, the whole dataset is considered (January 1986 - February 2008). Moreover, for comparability reasons, we gauge the forecasting abilities of dynamic Markov switching models. To this aim, a Markov-switching model is estimated for each country (see Abiad, 2003; Arias and Erlandson, 2005; Candelon et al. 2009). Nevertheless, contrary to the static model that has been previously used in the literature, our approach is based on a dynamic perspective, materialized in a switch of the lagged binary dependent variable from one regime to another, *i.e.*, from crisis to calm periods and vice-versa. Once the filtered probabilities are computed, the model for each country is evaluated. Besides, we first consider the time-series framework and then the panel one.

We assess the forecasting abilities of these models by considering the QPS and AUC evaluation criteria as well as the CW (Clark West, 2007) and DM (Diebold-Mariano, 1995) comparison tests. The left part of Table 4.7 reports the performance assessment

criteria for the three time-series models, *i.e.* static logit, dynamic logit, and dynamic Markov-switching.

Table 4.7: Evaluation criteria

	Time-series Models						Dynamic Panel Models			
	Static logit		Dynamic logit		Dynamic Markov		Pooled		Optimal Clusters	
	QPS	AUC	QPS	AUC	QPS	AUC	QPS	AUC	QPS	AUC
Argentina	0.043	0.938	0.033*	0.946	0.754	0.871	0.036	0.950*	0.036	0.950*
Brazil	0.069	0.710	0.068*	0.799	0.502	0.783	0.082	0.806	0.081	0.810*
Chile	0.022*	0.606*	0.022*	0.601	1.198	0.010	0.026	0.475	0.025	0.443
Indonesia	0.044	0.979	0.024*	0.981*	1.199	0.817	0.032	0.913	0.032	0.913
Israel	0.011*	0.994*	0.011*	0.994*	0.480	0.536	0.011*	0.983	--	--
South Korea	0.000*	1.000*	0.003	1.000	0.171	0.969	0.013	0.999	--	--
Malaysia	0.039	0.978*	0.015*	0.978*	0.239	0.955	0.017	0.978*	0.017	0.977
Mexico	0.073	0.784	0.038*	0.880	0.598	0.971*	0.044	0.924	0.044	0.923
Marocco	0.008*	0.888*	0.008*	0.888*	0.436	0.028	0.008*	0.271	0.048	0.910
Peru	0.059	0.974	0.038*	0.989	1.121	0.268	0.039	0.988	0.016	0.994*
Philippines	0.062	0.915	0.049	0.935	0.526	0.948	0.048	0.905	0.035*	0.977*
Thailand	0.021	0.995	0.011*	0.998*	1.203	0.828	0.016	0.994	0.035	0.965
Turkey	0.026*	0.976	0.027	0.978	0.386	0.992*	0.035	0.976	0.049	0.965
Uruguay	0.051	0.959	0.034	0.966	0.546	0.995*	0.035	0.964	0.008*	0.967
Venezuela	0.066	0.955	0.046	0.961	0.997	0.863	0.049	0.964	0.037*	0.989*

Note: The AUC criteria takes values between 0.5 and 1, 1 being the perfect model, while QPS ranges from 0 to 2, 0 being perfect accuracy. The asterisk * indicates the best model according to each evaluation criteria for the 15 countries.

Recall that the higher (lower) the *AUC* (the *QPS*), the better the model. It is thus evident that the outperforming model for most of the countries is the dynamic logit. By contrast, the Markov EWS registers very high *QPS* values, and sometimes *AUC* values that are worse than those of a random model, hence pointing out that this type of model, in spite of its advantages, does not perform well in forecasting. These findings support our intuition that the dynamic Markov model is not even as good as the static model. However, we must rely on comparison tests to check if the differences between the models are significant or not and for which countries. To this aim, we consider the *CW* test for nested models (static vs. dynamic logit) and the *DM* test for the non-nested ones (different estimation method). The results obtained are presented in Table 4.8.

Table 4.8: Comparison tests

	SL vs. DL ^a (Clark-West)		DL vs. DMK ^b (Diebold-Mariano)		DPL optimal vs. DMK ^c (Diebold-Mariano)		DL vs. DPL pooled ^d (Diebold-Mariano)		DL vs. DPL optimal ^e (Diebold-Mariano)	
	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
Argentina	2.144**	0.032	25.54***	< 0.001	25.61***	< 0.001	0.715	0.474	0.662	0.508
Brazil	1.136	0.256	8.004***	< 0.001	7.830***	< 0.001	1.496	0.135	1.491	0.136
Chile	1.036	0.300	21.89***	< 0.001	21.71***	< 0.001	1.343	0.179	1.391	0.164
Indonesia	2.526**	0.012	21.89***	< 0.001	48.04***	< 0.001	1.402	0.161	1.431	0.152
Israel	0.434	0.664	10.04***	< 0.001	--	--	0.139	0.889	--	--
South Korea	1.037	0.300	5.157***	< 0.001	--	--	1.304	0.192	--	--
Malaysia	2.656**	0.008	5.985***	< 0.001	5.885***	< 0.001	0.782	0.434	0.803	0.422
Mexico	3.291***	< 0.001	29.02***	< 0.001	25.81***	< 0.001	0.834	0.404	0.801	0.423
Marocco	0.668	0.504	9.726***	< 0.001	8.701***	< 0.001	1.033	0.302	2.320**	0.020
Peru	2.776**	0.006	40.70***	< 0.001	22.39***	< 0.001	0.138	0.890	-1.446	0.148
Philippines	2.358**	0.018	9.094***	< 0.001	6.976***	< 0.001	-0.059	0.953	-0.651	0.515
Thailand	2.067**	0.039	21.77***	< 0.001	18.12***	< 0.001	0.863	0.388	1.541	0.123
Turkey	2.067**	0.039	11.10***	< 0.001	6.048***	< 0.001	-0.832	0.406	1.075	0.282
Uruguay	2.827**	0.005	14.33***	< 0.001	10.89***	< 0.001	0.160	0.873	-1.746	0.081
Venezuela	2.977***	< 0.001	16.40***	< 0.001	11.98***	< 0.001	0.755	0.450	-2.308**	0.021

a - Static vs. Dynamic time series logit model.

b - Dynamic time-series logit model vs. Markov switching.

c - Dynamic panel logit model (poolable countries) vs. Markov switching.

d - Dynamic time-series logit model vs. Dynamic panel logit (all countries)

e - Dynamic time-series logit model vs. Dynamic panel logit (optimal clusters)

Note: The null hypothesis of both Clark-West and Diebold-Mariano tests is the equality of predictive performance of the two models. The alternative of both tests is the statistical difference of the loss associated with the two models. In the case of the Clark-West it means that the dynamic model is better than the static one (the test-statistic is positive if the loss corresponding to the dynamic model is inferior to the one associated with the static one). Similarly, the alternative of the Diebold Mariano test indicates that the first model is better than the other one if the test-statistic is positive. Under the null hypothesis, both test statistics follow a normal distribution. The asterisks ***, **, and * denote significance at the 1%, 5% and 10% significance level, respectively.

To be more precise, we first compare the static and the dynamic logit models (SL vs. DL) and show that the dynamic time-series specification outperforms the static one for most of the countries. As we have already seen in the SBC information criterion results, there are several countries for which the static model seems better than the dynamic. Here we have the statistical proof that for Brazil, Chile, Israel, South Korea and Morocco the static and dynamic models have similar forecasting abilities. Indeed, Clark-West's test rejects the null hypothesis of equal forecasting abilities for 10 out of the 15 countries. It thus corroborates our main finding that accounting for the endogenous persistence of the crisis by introducing the lagged binary dependent variable into the model matters for the forecasting abilities of currency crisis EWS models. Moreover, we compare the dynamic logit and the dynamic Markov models (DL vs. DMK), our findings stressing the forecasting superiority of the logit models versus the dynamic Markov ones.

It is thus obvious that in the quest for the optimal EWS policymakers and researchers should account for this persistence feature of the crisis indicator. To emphasize the importance of this modelling, we scrutinize the abilities of this country-by-country dynamic logit to discriminate between crisis and calm periods. The left part of Table 4.9 indicates the optimal cut-off for each country and the associated percentage of correctly forecasted crisis (calm) periods, *i.e sensitivity (specificity)* in the time-series framework.

Table 4.9: Optimal cut-off identification

Country	Time-series Models						Dynamic Panel Model			
	Static logit			Dynamic logit			Dynamic Markov			
	Cut-off	Se	Sp	Cut-off	Se	Sp	Cut-off	Se	Sp	Sp
Argentina	0.011	1.000	0.767	0.010	1.000	0.767	0.672	0.857	0.857	0.930
Brazil	0.064	0.500	0.859	0.042	0.700	0.847	0.026	0.900	0.694	0.898
Chile	0.014	0.667	0.714	0.014	0.667	0.721	< 0.001	1.000	< 0.001	0.966
Indonesia	0.061	1.000	0.917	0.036	0.909	0.969	0.908	0.818	0.992	0.984
Israel	0.066	1.000	0.992	0.203	1.000	0.992	0.998	0.500	0.973	--
South Korea	0.005	1.000	1.000	0.606	1.000	1.000	0.999	1.000	0.938	--
Malaysia	0.033	1.000	0.938	0.008	1.000	0.860	0.823	1.000	0.895	0.996
Mexico	0.244	0.643	0.972	0.034	0.857	0.988	0.567	1.000	0.873	0.968
Marocco	0.010	1.000	0.883	0.010	1.000	0.883	0.000	1.000	0.000	0.769
Peru	0.112	1.000	0.944	0.069	1.000	0.956	0.998	0.077	0.960	0.977
Philippines	0.038	1.000	0.784	0.022	1.000	0.733	0.987	1.000	0.882	0.973
Thailand	0.183	1.000	0.984	0.174	1.000	0.996	0.999	0.875	0.887	0.851
Turkey	0.071	0.889	0.957	0.107	0.889	0.969	0.897	1.000	0.957	0.813
Uruguay	0.031	1.000	0.851	0.021	1.000	0.843	0.937	1.000	0.984	0.000
Venezuela	0.065	0.929	0.892	0.092	0.929	0.948	0.999	1.000	0.725	0.956

Note: For each country we identify the optimal cut-off by using the accuracy measures method, so as to give more weight to the correct identification of crisis periods (*sensitivity*). *Se* stands for *sensitivity* (percentage of crises correctly forecasted by the EWS), while *Sp* represents *specificity* (percentage of calm periods correctly identified by the model). The higher *Se* and *Sp*, the better the model.

The optimal cut-off for each country has been identified by relying on accuracy measures thus giving more weight to the correct identification of crisis periods (*sensitivity*). It seems that both the static and dynamic country per country models are characterized by small values of the cut-offs (they range between 0.008 and 0.606) in contrast with the cut-offs associated with the dynamic Markov EWS. Moreover, both crisis and calm periods are very well forecasted by the dynamic logit model, *i.e.* sensitivity and specificity lay between 66.7% and 100% for each country while in the case of the static model they range between 50% and 100%. At the same time, the *sensitivity* and *specificity* of the dynamic Markov model vary a lot from one country to another, *i.e.* they sometimes reach their maximum, 1, but other times they drop to their minimum, 0, as well, indicating that dynamic Markov models are not as good as the logit ones are. This means that the lagged dependent variable has improved explanatory power and discriminates very well between calm and crisis periods, which further motivates the use of dynamic EWS. Indeed, for most of the countries the crisis probabilities issued from the dynamic logit model are quite low during real calm periods and they are very high in the real crises periods (see Figures 4.1 and 4.2), reinforcing the idea that the dynamic model outperforms the static one.

To be more precise, this model correctly forecasts most of the currency crisis episodes that have been recorded and analyzed by other studies (Abiad, 1993; Dabrowski, 2003; Glick and Hutchison, 1999), while the static model seems to be less efficient.

As for the panel framework, the right part of Table 4.7 indicates that the evaluation criteria have similar values from one model to another (pooled panel and optimal clusters panel respectively) which are quite close (but lesser) to the time-series logit ones. At the same time, the right part of Table 4.8 sets out the fact that the dynamic panel models do not have better forecasting abilities than the dynamic time series EWS, thus supporting the results obtained by Berg et al (2008). To be more precise, the null hypothesis of equal forecasting abilities cannot be rejected for either of the panel models. The only exception is Venezuela, for which the use of a dynamic panel model with optimal clusters significantly improves the EWS forecasting abilities. Besides, it appears that not only time-series dynamic logit models but also dynamic panel logit models are better than dynamic Markov-switching specifications (see the middle part of Table 4.8). Finally, the values of the optimal cut-off for the dynamic optimal clusters panel model are similar to the ones registered for the dynamic time-series logit model, even though fairly smaller *i.e.*, they vary between 0.001 and 0.74 (see Table 4.9). Similarly, the *sensitivity* and the *specificity* corresponding to the two types of dynamic models resemble.

All in all, our findings prompt the fact that there are gains from using a dynamic EWS specification which includes the lagged binary crisis indicator.

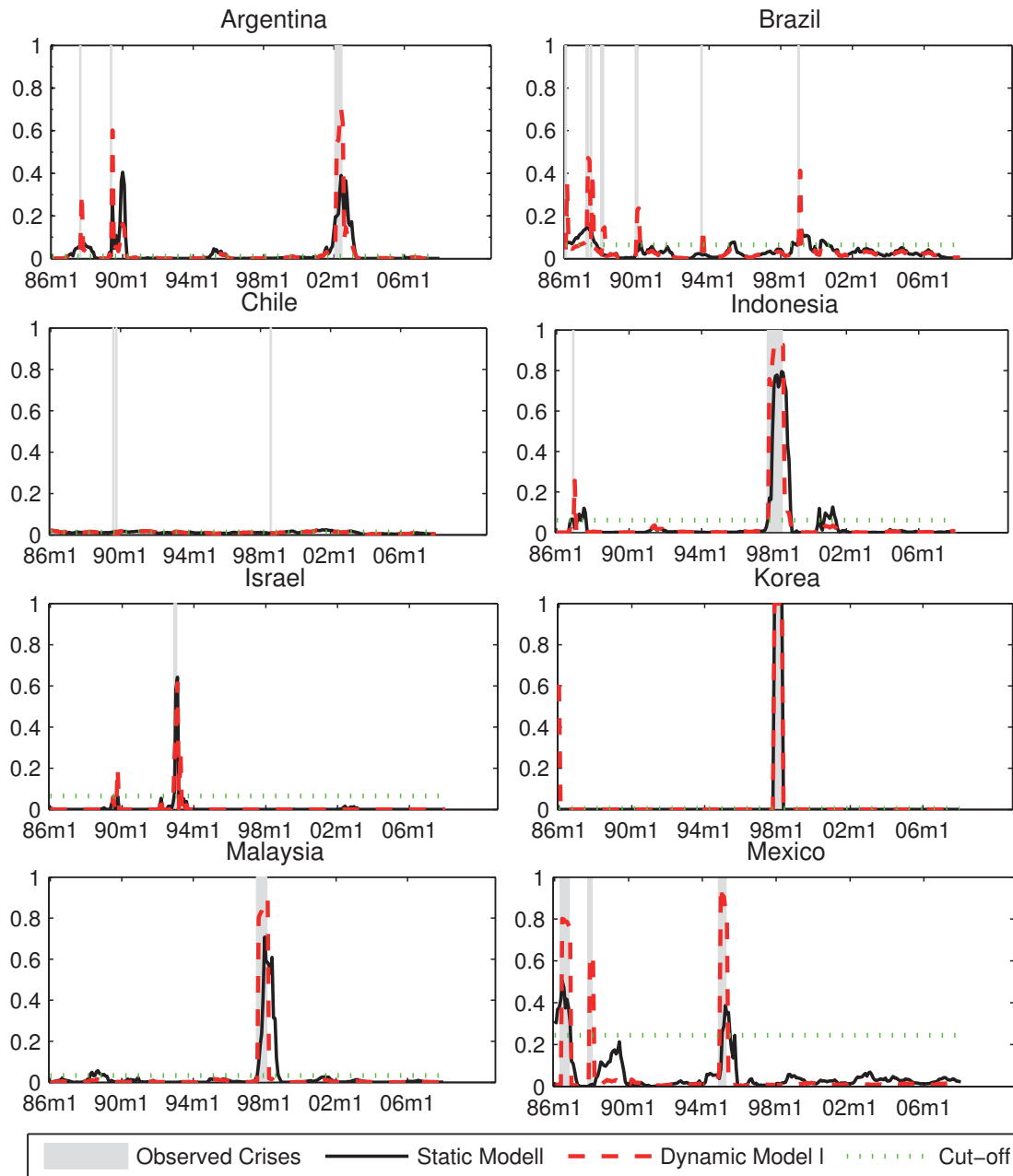


Figure 4.1: Predicted probability of crisis - in sample

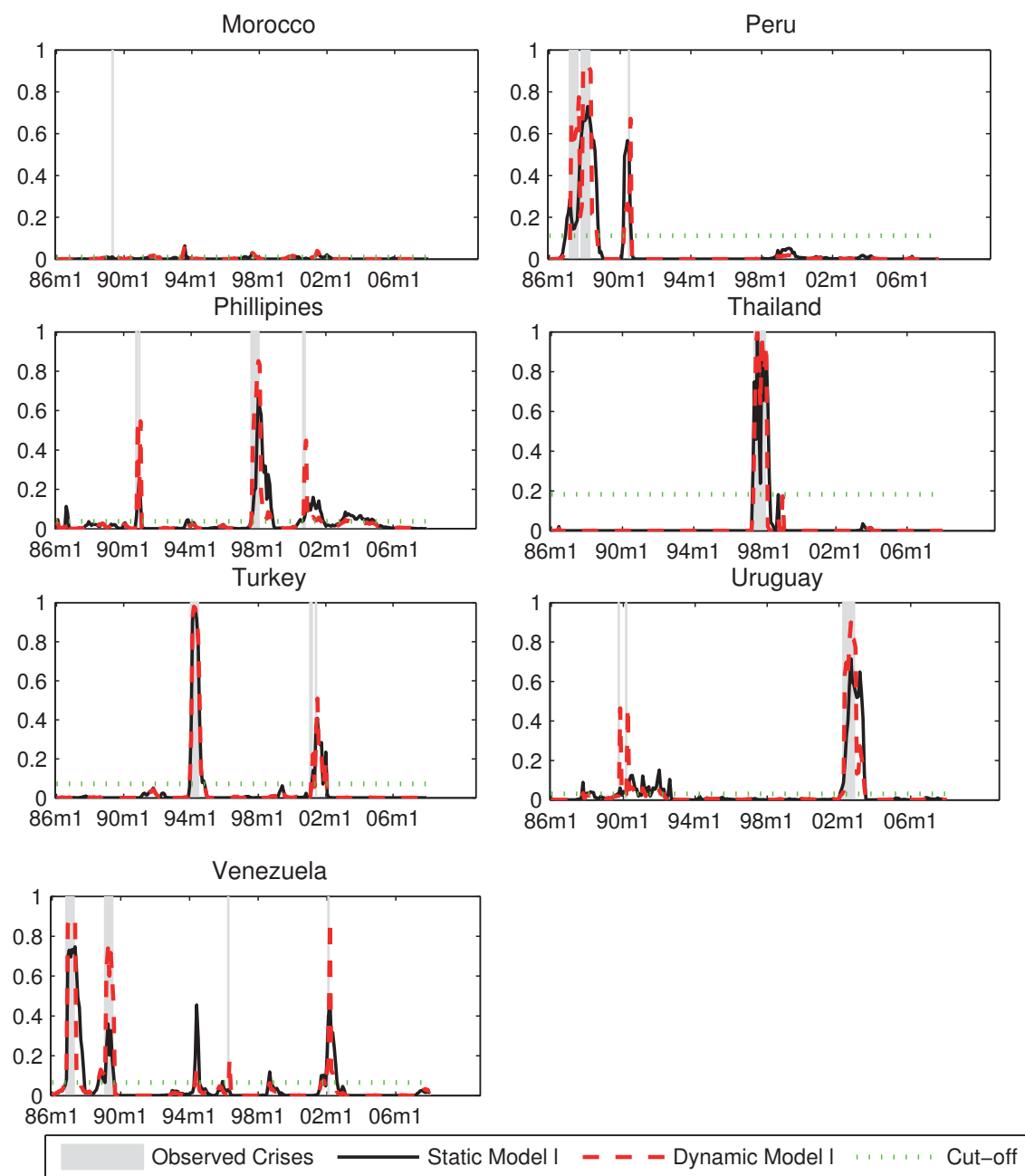


Figure 4.2: Predicted probability of crisis - in sample (continued)

4.4.2 Out-of-sample analysis

We now check the out-of-sample performance of our dynamic time-series EWS model. More precisely, a dynamic time-series logit model is estimated over the January 1986 - December 1996 window, and the estimated parameters are used to compute the probability of having a crisis in January 1997. This estimation and forecasting is then reconducted for the February 1986-January 1997 period to obtain the out-of-sample crisis probability for February 1997 and so on. Note that the evaluation methodology previously used can be easily implemented for the series of out-of-sample crisis probabilities obtained for each country.

The left part of Table 4.10 reports the Clark-West comparison test statistic and p-value for the out-of-sample forecasts issued by the static and the dynamic time-series logit models for a one month horizon ($C1$).

Table 4.10: Comparison tests (Out of sample exercise)

Country	Static $C1$ vs. Dynamic $C1$		Static $C24$ vs. Dynamic $C24$	
	statistic	p-value	statistic	p-value
Argentina	1.383	0.167	4.011***	< 0.001
Brazil	-0.387	0.699	3.835***	< 0.001
Chile	-0.329	0.742	3.969***	< 0.001
Indonesia	1.752*	0.080	-1.803*	0.071
Israel	-1.360	0.174	-4.312***	< 0.001
South Korea	1.004	0.315	1.287	0.198
Malaysia	2.016**	0.044	-1.665*	0.096
Mexico	3.735***	< 0.001	-3.397***	< 0.001
Marocco	-3.791***	< 0.001	3.177***	< 0.001
Peru	2.268**	0.023	4.612***	< 0.001
Philippines	2.037**	0.042	-0.933	0.351
Thailand	1.593	0.111	1.376	0.169
Turkey	-1.711*	0.087	4.020***	< 0.001
Uruguay	2.244**	0.025	1.051	0.293
Venezuela	0.813	0.416	3.669***	< 0.001

Note: The null hypothesis of the Clark-West test is the equality of predictive performance of the two models. Under the null hypothesis, the test-statistic follows a normal distribution. Its alternative indicates that the dynamic EWS is better than the static one if the test-statistic is positive (the loss associated with the dynamic model is inferior to the one corresponding to the static model). Otherwise, the alternative indicates that the static model outperforms the dynamic one. The asterisks ***, **, and * denote significance at the 1%, 5% and 10% significance level, respectively.

It appears that the dynamic logit model outperforms the static one for half of the countries, thus demonstrating the robustness of our in-sample findings. The out-of-sample results of the validation criteria support our results (they are available upon request).

Furthermore, Figures 4.3 and 4.4 present the out-of-sample crisis probabilities from January 1997 to February 2008 for each of the 12 countries registering at least one crisis in the within-sample period (January 1986 - December 1996).⁸

It results that for the countries which faced more than one month of currency crisis, the EWS forecasting probability is very low in calm periods while it is very high during crisis periods. It hence goes along the lines of our in-sample results. On the contrary, when countries faced only one period of crisis, the model's forecasting abilities are disappointing. These results are confirmed by Table 4.11.

Table 4.11: Optimal cut-off identification (Out-of-sample exercise)

	C1 forecasts			C24 forecasts		
	cut-off	Sensitivity	Specificity	cut-off	Sensitivity	Specificity
Argentina	0.011	0.800	0.938	0.124	0.679	0.868
Brazil	0.110	< 0.001	0.992	0.588	0.500	0.891
Chile	0.021	< 0.001	0.812	0.206	1.000	0.746
Indonesia	0.036	0.800	0.952	0.136	0.944	0.724
South Korea	--	--	--	0.087	1.000	0.576
Malaysia	--	--	--	0.138	0.714	0.508
Philippines	0.065	0.625	0.857	0.336	0.579	0.458
Thailand	--	--	--	0.109	1.000	0.405
Turkey	0.002	0.660	0.670	0.155	1.000	0.491
Uruguay	0.028	0.875	0.944	0.480	0.226	0.932
Venezuela	0.162	< 0.001	0.977	0.353	0.625	0.645

Note: For each country we identify the optimal cut-off by using the accuracy measures method, so as to give more weight to the correct identification of crisis periods (sensitivity). The values of the cut-off are calculated on the basis of the in-sample dataset (January 1986 - December 1996). The other countries, *i.e.* Israel, South Korea, Malaysia, Mexico, Morocco, Peru and Thailand do not register any in-sample or out-of-sample crises. *sensitivity* is the percentage of crises correctly forecasted by the EWS), while *specificity* represents the percentage of calm periods correctly identified by the model. The higher the *sensitivity* and *specificity*, the better the model.

Indeed, there are countries like Brazil, Chile and Venezuela that know only one out-of-sample period of crisis, which is not forecasted in due time. Such a result is driven by the low amount of crisis observations: the number of 0 is huge, hence causing bias in the estimation of the model. This finding supports the use of longer forecast horizons as it is usually done in the literature. Considering a forecast horizon of j months increases “artificially” the number of ‘ones’ observed in the sample, which should improve the quality of the estimation and hence that of the forecasting ability. Besides, results should be available for a larger number of countries since, contrary to the previous case, there are more chances of registering values of ‘one’ in the within-sample period.

In this context, a robustness check is performed by employing the $C24$ crisis variable. Nevertheless, the use of a 24-month-ahead crisis variable also introduces autocorrelation (see Berg and Coke, 2004) which may be problematic, especially in the case of a dynamic

⁸Without at least one in-sample crisis, the optimal cut-off needed in order to evaluate the out-of-sample forecasting abilities of the model cannot be computed.

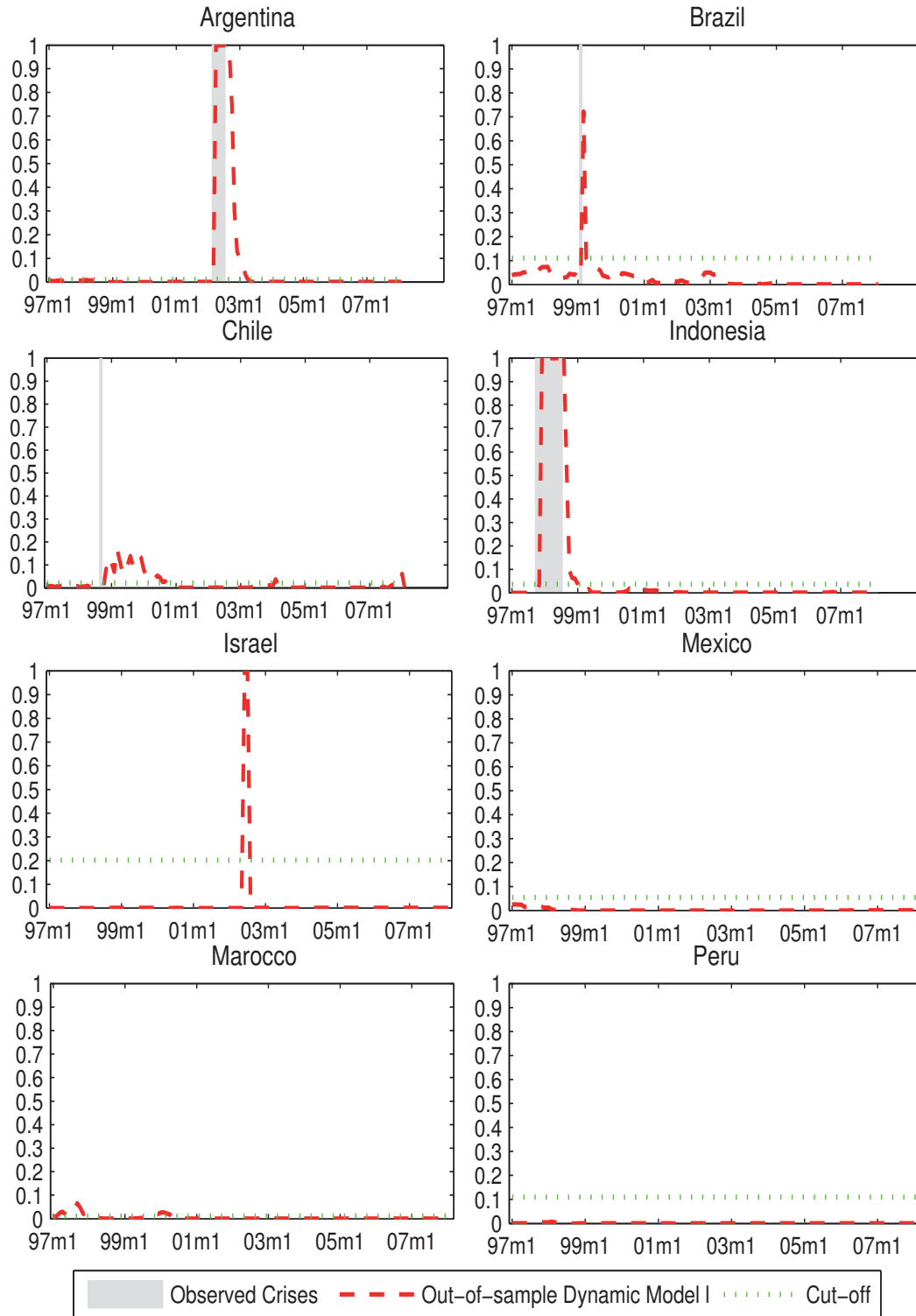
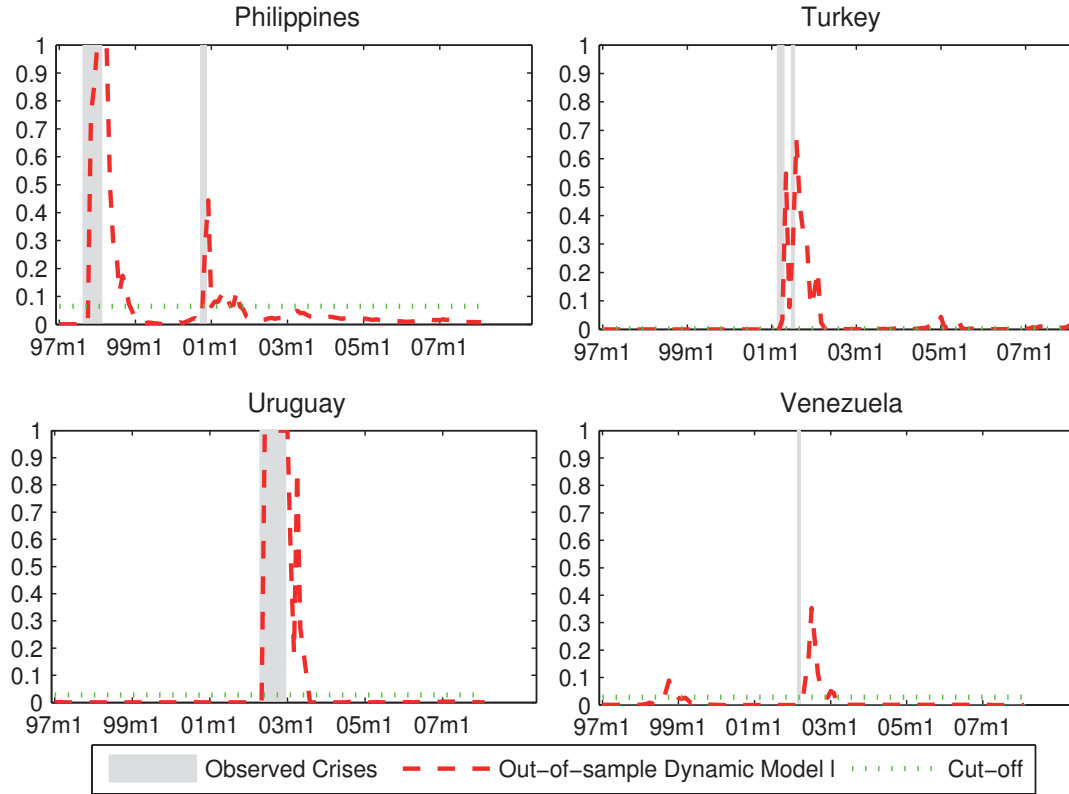


Figure 4.3: Predicted probability of crisis (C1) - out-of-sample


 Figure 4.4: Predicted probability of crisis ($C1$) - out-of-sample (continued)

model, since the lagged binary variable may be correlated with the error term.⁹ In order to avoid this situation, we consider that the probability of having at least one crisis in the next 24 months $\Pr(C24_t = 1)$ depends on the exogenous macroeconomic variables (x_{t-1}) and on the state of the economy during the previous 24 months, which is given by the 25th lag of the $C24$ variable ($C24_{t-25}$).¹⁰

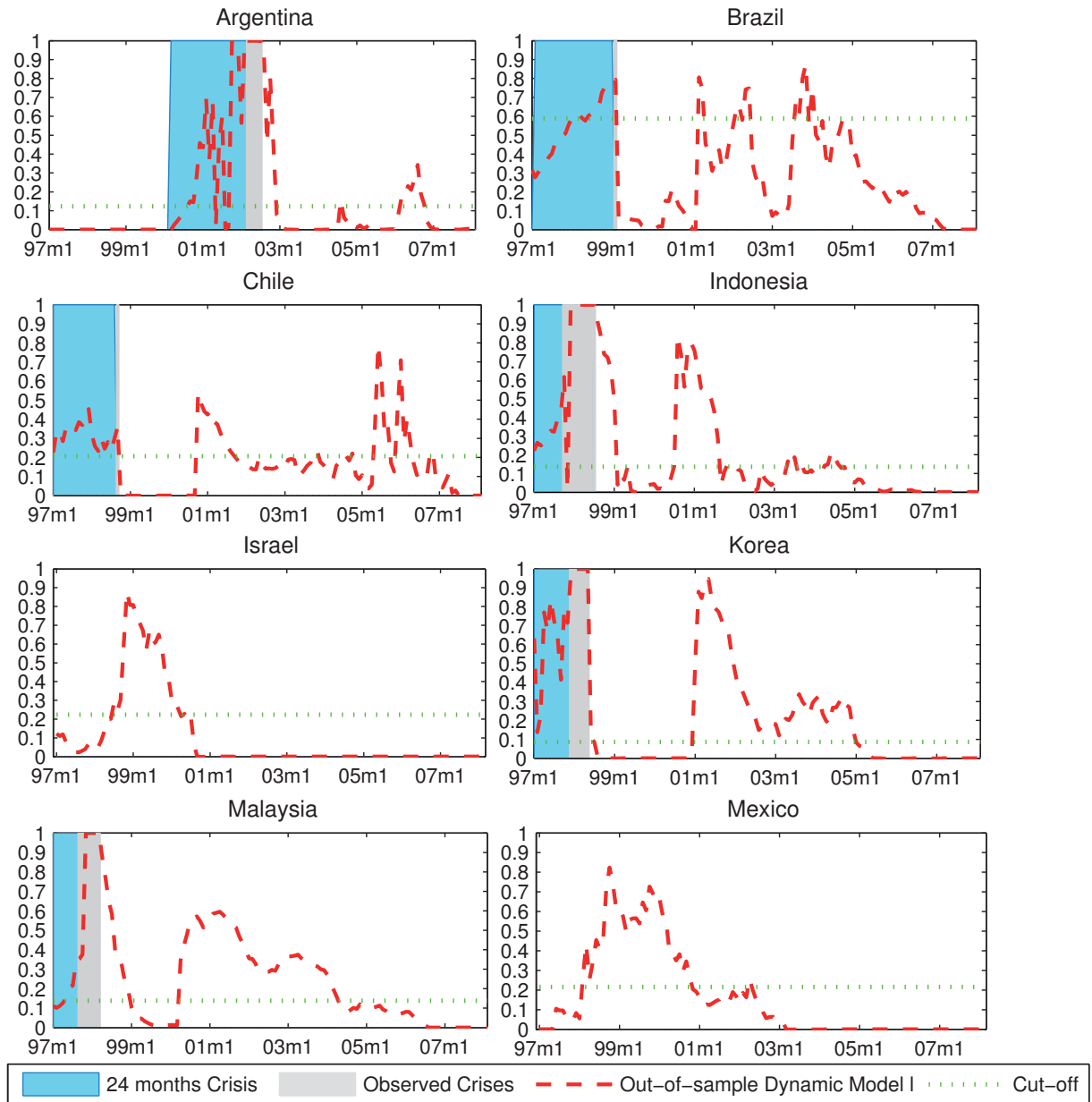
Before detailing the out-of-sample results for this new dynamic EWS model, we compare its forecasting abilities to the ones of a static $C24$ EWS (see the right part of Table 4.10) by using Clark-West test for nested models. It appears that the dynamic model has better forecasting abilities than the static one for 7 countries (at a 5% significance level) and for 9 countries (at a 10% significance level) respectively.

Moreover, Figures 4.5 and 4.6 present the out-of-sample probabilities corresponding to this new binary variable.

As expected, the results seem quite noisy, since we have increased the forecasting window from one to 24 months. Indeed, the crisis are always correctly forecasted with at least a few months before their occurrence but there are also a lot of false alarms.

⁹Harding and Pagan, (2009) note that the binary variables dating crisis, recessions as well as bulls and bears are 'constructed binary variables' which, contrary to those encountered in microeconomics, are not i.i.d. When constructed binary variables are used as a dependent variable, test statistics robust to the serial correlation and heteroskedasticity characterizing the binary variable must be employed.

¹⁰To be more exact, the model becomes: $\Pr_{t-1}(C24_t = 1) = \Lambda(\delta C24_{t-25} + x_{t-1}\beta)$.

Figure 4.5: Predicted probability of crisis (C_{24}) - out-of-sample

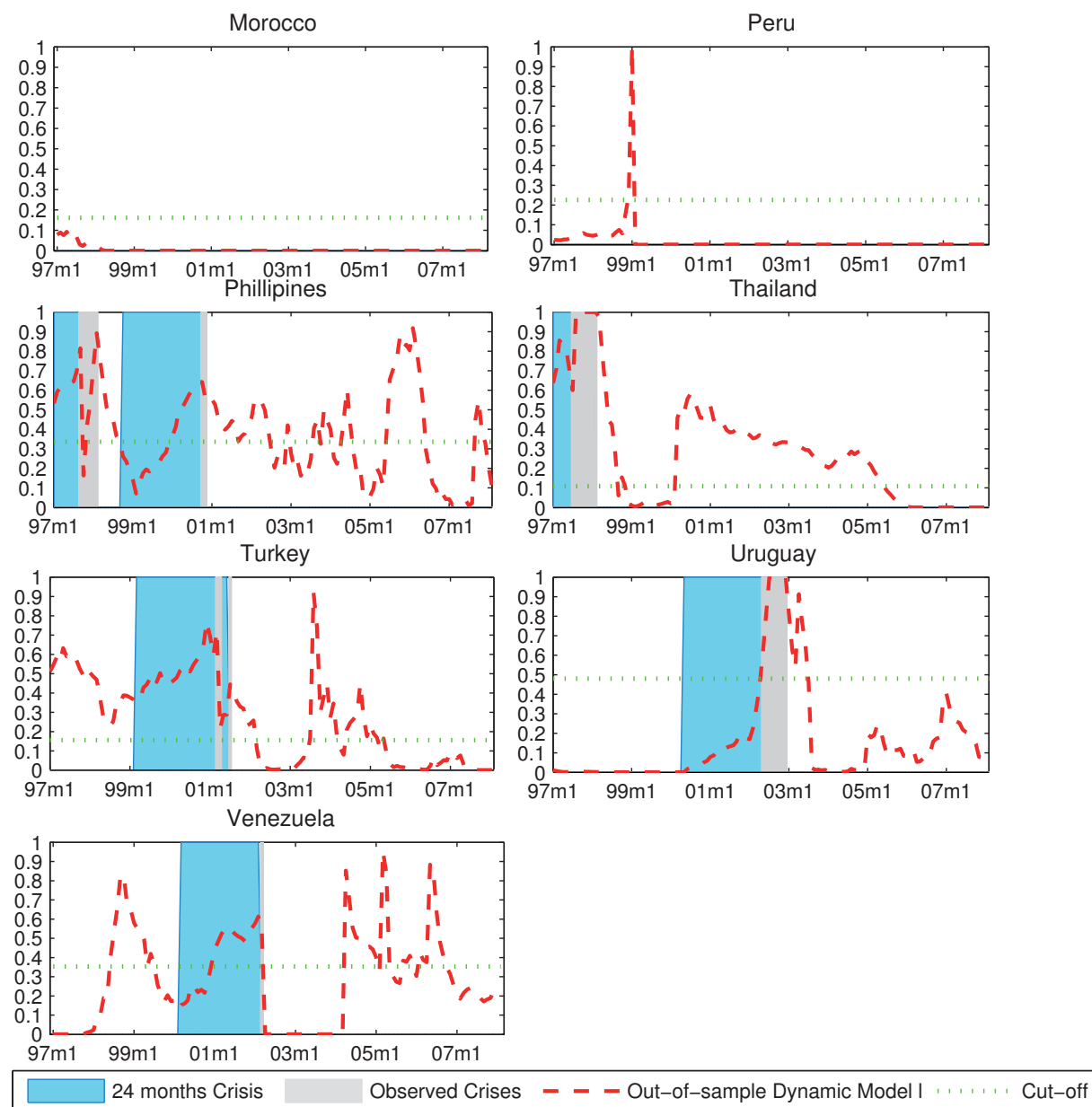


Figure 4.6: Predicted probability of crisis (C_{24}) - out-of-sample (continued)

Nevertheless, at a closer look it can be noticed that most of the false alarms appear around 24 months after the real crisis, as a result of the introduction of the 25th lagged binary dependent variable. These events can be thus considered as 'predictable' and should not be taken into account when analyzing the predictive abilities of this model. In such a case, only Brazil, Chile and Philippines face major false alarms in the out-of-sample period we consider. As showed in the case of $C1$, but for Morocco and Peru, the countries which never faced a crisis in the out-of-sample period are characterized by false alarms. No model can accurately forecast the non-occurrence of crises in such a case.

The cut-offs used for the out-of-sample series of crisis probabilities in the case of $C24$ as well as the percentage of crisis and calm periods correctly identified are reported in Table 4.11. First, there are countries for which the crisis in 24 months and the calm periods are quite well forecasted by the two models, but there are also countries like Brazil, Thailand and Uruguay for which the occurrence of the crisis is not forecasted in due time. Similarly, there are countries like Philippines, Thailand and Turkey, for which the calm periods are not well identified by the $C24$ model. However, as already discussed, the actual predictive abilities of this model would be better if we kept in mind that the false alarms escalating 24 months after the crisis are actually expected, as a side-effect of the modelling.

All in all, the dynamic EWS model including the lagged binary crisis indicator has good forecasting abilities not only in-sample but also out-of-sample. These findings vindicate dynamic EWS models.

4.5 Conclusion

This paper provides evidence of the importance of crisis dynamics to adequately forecast crises and shows that future EWS models should integrate this dynamic. It is actually the first one to consider an exact ML methodology, elaborated on Kauppi and Saikonen (2008), to estimate a dynamic discrete-choice EWS. In a second part, it extends this methodology to panel by drawing on the works of Carro (2007).

Several conclusions can be drawn from the empirical application of this methodology to construct currency crisis EWS models. First, we show that dynamic logit models consistently outperform static ones as well as Markov-switching. This conclusion is drawn from the within sample forecast exercise. Such a result is corroborated by the out-of-sample forecast exercise performed by considering both a crisis at time t and a crisis in 24 months variable. One reasoning behind this is that the dynamics captures the auto-correlation, the persistence observed in such EWS. Second, looking at their forecasting ability, it turns out that dynamic EWS deliver good forecasting probabilities, beyond the optimal cut-off in the crisis periods and underneath this threshold the rest of the time.

There is no doubt that in the quest for a new generation for financial crisis EWS, dynamics should constitute a key characteristic that would deliver more adequate signals to prevent financial turmoils. Let us hope that policy makers could exploit these signals to tame such painful events.

4.6 Appendix

4.6.1 Appendix: Constrained Maximum Likelihood Estimation (Kauppi and Saikkonen, 2008)

Let us recall the general form of the model in the case of a logistic distribution function Λ : $\Pr_{t-1}(y_t = 1) = \Lambda(\delta\pi_{t-1} + \alpha y_{t-1} + x_{t-1}\beta)$. Following Kauppi and Saikkonen, we set the initial value π_0 to $(\bar{x}\beta)/(1 - \delta)$, $\bar{x}$ being the sample mean of the exogenous variables. The initial condition for the β vector of parameters is given by an OLS estimation, while the initial δ is set to 0. Moreover, since δ is an autoregressive parameter, a constrained maximum likelihood estimation must be implemented. Nevertheless, the same results can be reached in a faster and easier way, by using a transformation of the δ parameter in the classical maximum likelihood process. Thus, to solve this problem, we denote by ψ the new maximization parameter, identified so that δ is equal to $\psi/(1 + |\psi|)$, *i.e.*, δ takes values in the interval $[0,1]$.

Hence, the log-likelihood function takes the form of:

$$\text{LogL}(\theta) = \sum_{t=1}^T l_t(\theta) = \sum_{t=1}^T [y_t \log \Lambda(\pi_t(\theta)) + (1 - y_t) \log(1 - \Lambda(\pi_t(\theta)))], \quad (4.10)$$

where θ is the vector of parameters $\theta = [\psi, \alpha, \beta]$.

It is noticed that in view of the parameter transformation from δ to ψ , the maximization covariance matrix corresponds to the parameters $[\psi, \alpha, \beta]$, and not to the initial parameters $[\delta, \alpha, \beta]$. Thus, we must proceed to a change of the covariance matrix from the first space to the second one. To this end, we use Taylor's theorem to calculate the approximation of the transformation function around the point ψ_0 . To be more exact, since the estimated parameter $\hat{\delta} = f(\hat{\psi})$, where $f(\hat{\psi}) = \hat{\psi}/(1 + |\hat{\psi}|)$, the approximation becomes:

$$\hat{\delta} = f(\hat{\psi}) \simeq f(\psi_0) + \frac{\partial f(\hat{\psi})'}{\partial \psi} \Big|_{\psi_0} (\hat{\psi} - \psi_0). \quad (4.11)$$

Nevertheless, we aim at finding the variance of δ , and thus, using the formula $\mathbb{V}(a'X) = a'\mathbb{V}(X)a$, we obtain:

$$\mathbb{V}(\hat{\delta}) \simeq \frac{\partial f(\hat{\psi})'}{\partial \psi} \Big|_{\psi_0} \mathbb{V}(\hat{\psi}) \frac{\partial f(\hat{\psi})}{\partial \psi} \Big|_{\psi_0} \quad (4.12)$$

Since $\hat{\psi} \xrightarrow{P} \psi_0$, we can replace ψ_0 with the estimator $\hat{\psi}$ in eq. 4.12:

$$\mathbb{V}(\hat{\delta}) \simeq \frac{\partial f(\hat{\psi})'}{\partial \psi} \Big|_{\hat{\psi}} \mathbb{V}(\hat{\psi}) \frac{\partial f(\hat{\psi})}{\partial \psi} \Big|_{\hat{\psi}} \quad (4.13)$$

Last but not least, the first derivative of the transformation function $f(\hat{\psi})$ with respect to $(\hat{\psi})$ can be computed through finite differences. Consequently, the standard errors obtained as the square root of the elements laying on the first diagonal of the covariance matrix are consistent with the $[\delta, \alpha, \beta]$ vector of parameters. More exactly, a Gallant correction based on a Parzen kernel (Gallant, 1987) is used for the covariance matrix. Kauppi and Saikonnen (2008) argue that robust standard errors can be obtained as the diagonal elements of the matrix $\hat{J}(\hat{\theta})^{-1} \hat{I}(\hat{\theta}) \hat{J}(\hat{\theta})^{-1}$, where $\hat{I}(\hat{\theta}) = T^{-1}(\sum_{t=1}^T \hat{d}_t \hat{d}_t' + \sum_{t=1}^T w_{Tj} \sum_{t=j+1}^T (\hat{d}_t \hat{d}_{t-j}' + \hat{d}_{t-j} \hat{d}_t'))$, $\hat{d}_t = \partial l_t(\hat{\theta}) \partial \theta$, and where $J(\theta) = \text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T (\partial^2 l_t(\theta) \partial \theta \partial \theta')$. On top of that, we consider that the robust covariance matrix should be used not only for h -periods-ahead forecasts, $h > 1$ (as in Kauppi and Saikonnen, 2008) but also for one period-ahead forecasts, since the logistic distributional hypothesis imposed to the error term might not always hold and, most importantly, since this covariance matrix specification is robust to autocorrelation, automatically introduced when considering an EWS (see Berg and Coke, 2004).

4.6.2 Appendix: Modified Maximum Likelihood Estimation (Carro, 2007)

As previously mentioned, the dynamic panel logit models with fixed effects is estimated by solving $N + 1$ non-linear equations based on the modified score of each individual, which takes the following form:

$$\begin{aligned} d_{\theta M i}(\theta) = d_{\theta C i}(\theta, \hat{\eta}_i(\theta)) - \frac{1}{2} \frac{1}{d_{\eta \eta i}(\theta, \hat{\eta}_i(\theta))} & \left(d_{\theta \eta \eta i}(\theta, \hat{\eta}_i(\theta)) + d_{\eta \eta \eta i}(\theta, \hat{\eta}_i(\theta)) \frac{\partial \hat{\eta}_i(\theta)}{\partial \theta} \right) \\ + \frac{\partial / \partial \eta_i (\mathbb{E}[d_{\theta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i])}{\mathbb{E}[d_{\eta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]} \Big|_{\eta_i = \hat{\eta}_i(\theta)} & - \frac{\mathbb{E}[d_{\theta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]}{\mathbb{E}[d_{\eta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]} \Big|_{\eta_i = \hat{\eta}_i(\theta)} \\ * \frac{\partial / \partial \eta_i (\mathbb{E}[d_{\eta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i])}{\mathbb{E}[d_{\eta \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]} \Big|_{\eta_i = \hat{\eta}_i(\theta)}, \end{aligned} \quad (4.14)$$

where $d_{\theta C i}(\theta, \hat{\eta}_i(\theta))$ is an individual's score from the concentrated likelihood (MLE):

$$d_{\theta C i}(\theta, \hat{\eta}_i(\theta)) = \frac{y_{it} - F_{it}(\theta, \eta_i)}{F_{it}(\theta, \eta_i)(1 - F_{it}(\theta, \eta_i))} \left(y_{it} + \frac{\partial \hat{\eta}_i(\theta)}{\partial \theta} \right) \quad (4.15)$$

From the first order condition of η_i , $d_{\eta_i}(\theta, \eta_i) = \sum_{t=1}^T (y_{it} - F_{it}(\theta, \eta_i)) / F_{it}(\theta, \eta_i)(1 - F_{it}(\theta, \eta_i)) = 0$, it can be derived that the estimators $\hat{\eta}_i$, $i = 1, 2, \dots, N$ solve the following equation:

$$\sum_{t=1}^T y_{it} \frac{f_{it}(\theta, \eta_i)}{F_{it}(\theta, \eta_i)(1 - F_{it}(\theta, \eta_i))} = \sum_{t=1}^T \frac{F_{it}(\theta, \eta_i) f_{it}(\theta, \eta_i)}{F_{it}(\theta, \eta_i)(1 - F_{it}(\theta, \eta_i))} \quad (4.16)$$

Deriving eq. 4.15 with respect to θ for the second dynamic binary specification, *i.e.* the one including the lagged binary indicator, we can obtain $\partial\hat{\eta}_i(\theta)/\partial\theta$:

$$\partial\hat{\eta}_i(\theta)/\partial\theta = -\frac{\sum_{t=1}^T \frac{X_i Z}{F_{it}(\theta, \eta_i)^2 (1 - F_{it}(\theta, \eta_i)^2)}}{\sum_{t=1}^T \frac{Z}{F_{it}(\theta, \eta_i)^2 (1 - F_{it}(\theta, \eta_i)^2)}}, \quad (4.17)$$

where $X_i = y_{t-1}, x_{i1}, x_{i2}, \dots, x_{iK}$, is the explanatory variable corresponding to the $\theta = \alpha, \beta_1, \beta_2, \dots, \beta_K$ parameter we analyze, K is the number of explanatory variables, and $Z = y_t[f'_{it}(\theta, \eta_i)F_{it}(\theta, \eta_i)(1 - F_{it}(\theta, \eta_i)) - f_{it}^2(\theta, \eta_i)(1 - 2F_{it}(\theta, \eta_i))] - f_{it}^2(\theta, \eta_i)F_{it}^2(\theta, \eta_i) - F_{it}^2(\theta, \eta_i)(1 - F_{it}(\theta, \eta_i))f'_{it}(\theta, \eta_i)$. Let us remind that $F_{it}(\theta, \eta_i)$ is the cumulative distribution function, $f_{it}(\theta, \eta_i)$ is the density function and $f'_{it}(\theta, \eta_i)$ is the first derivative of the density function. Thus, in the case of a logit model $F_{it}(\theta, \eta_i) = \exp(\alpha y_{it-1} + x_{it-1}\beta + \eta_i)/(1 + \exp(\alpha y_{it-1} + x_{it-1}\beta + \eta_i))$, $f_{it}(\theta, \eta_i) = \exp(\alpha y_{it-1} + x_{it-1}\beta + \eta_i)/(1 + \exp(\alpha y_{it-1} + x_{it-1}\beta + \eta_i))^2$, and $f'_{it}(\theta, \eta_i) = \exp(\alpha y_{it-1} + x_{it-1}\beta + \eta_i)(1 - \exp(\alpha y_{it-1} + x_{it-1}\beta + \eta_i))/(1 + \exp(\alpha y_{it-1} + x_{it-1}\beta + \eta_i))^3$.

To put it another way, the partial derivative of the η gradient with respect to θ is given by the implicit functions theorem:

$$\partial\hat{\eta}_i(\theta)/\partial\theta_k = -\frac{\partial d_{\eta_i}(\theta, \hat{\eta}_i)/\partial\theta_k}{\partial d_{\eta_i}(\theta, \hat{\eta}_i)/\partial\eta_i}, \quad (4.18)$$

where $k=1,2,\dots,K$, K being the number of explanatory variables considered in the model, $\partial d_{\eta_i}(\theta, \hat{\eta}_i)/\partial\theta_k = \partial^2 \text{LogL}(\theta, \eta_i)/\partial\eta_i \partial\theta_k|(\theta, \hat{\eta}_i)$, and $\partial d_{\eta_i}(\theta, \hat{\eta}_i)/\partial\eta_i = \partial^2 \text{LogL}(\theta, \eta_i)/\partial^2 \eta_i|(\theta, \hat{\eta}_i)$. The estimation of the parameters by classical MLE is straightforward since $d_{\theta C_i}(\theta, \hat{\eta}_i(\theta)) = 0$ and $d_{\eta_i}(\theta, \eta_i) = 0$ can be easily computed and solved. However, aiming to reduce the estimation bias, the implementation of MMLE becomes compulsory, for which further information regarding the expectance of the first order condition and the derivatives of this expectance is required.

In view of the MMLE estimation, we derive the following elements for the α parameter, corresponding to the lagged binary variable:

$$d_{\alpha\eta_i}(\theta, \eta_i) = \frac{\partial^2 \text{LogL}_i}{\partial\alpha \partial\eta_i} = -\sum_{t=1}^T y_{i,t-1} f_{it}(\alpha y_{it-1} + \beta x_i + \eta_i), \quad (4.19)$$

$$d_{\eta_i\eta_i}(\theta, \eta_i) = \frac{\partial^2 \text{LogL}_i}{\partial^2 \eta_i} = -\sum_{t=1}^T f_{it}(\alpha y_{it-1} + \beta x_i + \eta_i), \quad (4.20)$$

$$d_{\alpha\eta_i}(\theta, \eta_i) = - \sum_{t=1}^T y_{i,t-1} f'_{it}(\alpha y_{it-1} + \beta x_i + \eta_i), \quad (4.21)$$

$$d_{\eta_i\eta_i}(\theta, \eta_i) = - \sum_{t=1}^T f'_{it}(\alpha y_{it-1} + \beta x_i + \eta_i), \quad (4.22)$$

where f is the logistic pdf and f' is the first derivative of the logistic pdf.

Next, we aim at calculating the expectation of the derivatives $d_{\alpha\eta_i}$ and $d_{\eta_i\eta_i}$. Thus, in a first step we calculate the probability at time t that a crisis will occur in country i in the next 24 months given the initial value of the binary dependent variable y_{i0} , the fixed effects η_i and the explanatory variables x_i ($\Pr(y_{it} = 1|y_{i0}, \eta_i, x_i)$):

$\Pr(y_{i1} = 1|y_{i0}, \eta_i, x_i) = F_{it}(\alpha y_{i0} + \beta x_i + \eta_i)$ starting point. For $t > 1$:

$$\begin{aligned} \Pr(y_{it} = 1|y_{i0}, \eta_i, x_i) &= \Pr(y_{it-1} = 1|y_{i0}, \eta_i, x_i)(F_{it}(\alpha + \beta x_i + \eta_i) - F_{it}(\beta x_i + \eta_i)) \\ &\quad + F_{it}(\beta x_i + \eta_i). \end{aligned} \quad (4.23)$$

Moreover, $\Pr(y_{i0} = 1|y_{i0}, \eta_i, x_i) = y_{i0}$.

In the second step the expectation of the two derivatives can be calculated:

$$\mathbb{E}[d_{\alpha\eta_i}(\theta, \eta_i)|y_{i0}, \eta_i, x_i] = - \sum_{t=1}^T \mathbb{E}[y_{it-1} f'_{it}(\alpha y_{it-1} + \beta x_i + \eta_i)|y_{i0}, \eta_i, x_i], \text{ where} \quad (4.24)$$

$$\mathbb{E}[y_{it-1} f'_{it}(\alpha y_{it-1} + \beta x_i + \eta_i)|y_{i0}, \eta_i, x_i] = f'_{it}(\alpha + \beta x_i + \eta_i) \Pr(y_{it-1} = 1|y_{i0}, \eta_i, x_i). \quad (4.25)$$

and

$$\mathbb{E}[d_{\eta_i\eta_i}(\theta, \eta_i)|y_{i0}, \eta_i, x_i] = - \sum_{t=1}^T \mathbb{E}[f'_{it}(\alpha y_{it-1} + \beta x_i + \eta_i)|y_{i0}, \eta_i, x_i], \text{ where} \quad (4.26)$$

$$\begin{aligned} \mathbb{E}[f'_{it}(\alpha y_{it-1} + \beta x_i + \eta_i)|y_{i0}, \eta_i, x_i] &= f'_{it}(\alpha + \beta x_i + \eta_i) \Pr(y_{it-1} = 1|y_{i0}, \eta_i, x_i) \\ &\quad + f'_{it}(\beta x_i + \eta_i)(1 - \Pr(y_{it-1} = 1|y_{i0}, \eta_i, x_i)) \\ &= \Pr(y_{it-1} = 1|y_{i0}, \eta_i, x_i)(f'_{it}(\alpha + \beta x_i + \eta_i) \\ &\quad - f'_{it}(\beta x_i + \eta_i)) + f'_{it}(\beta x_i + \eta_i), \end{aligned} \quad (4.27)$$

The last elements needed in the gradient function are the derivatives of the two expectations of $d_{\alpha\eta_i}$ and $d_{\eta_i\eta_i}$ with respect to the fixed effect η_i . Nevertheless, to compute these elements, the derivative of the probability of occurrence of a crisis within 24 months at

time t in country i with respect to η_i must be first calculated:

$$\frac{\partial}{\partial \eta_i} \Pr(y_{i1} = 1 | y_{i0}, \eta_i, x_i) = f_{i1}(\alpha y_{i0} + \beta x_i + \eta_i) \quad (4.28)$$

$$\begin{aligned} \frac{\partial}{\partial \eta_i} \Pr(y_{it} = 1 | y_{i0}, \eta_i, x_i) &= \frac{\partial}{\partial \eta_i} \Pr(y_{it-1} = 1 | y_{i0}, \eta_i, x_i) (F_{it}(\alpha + \beta x_i + \eta_i) \\ &\quad - F_{it}(\beta x_i + \eta_i)) \\ &\quad + \Pr(y_{it} = 1 | y_{i0}, \eta_i, x_i) (f_{it}(\alpha + \beta x_i + \eta_i) - f_{it}(\beta x_i + \eta_i)) \\ &\quad + f_{it}(\beta x_i + \eta_i). \end{aligned} \quad (4.29)$$

Finally, the last elements needed in the formula of the gradient are obtained:

$$\begin{aligned} \frac{\partial}{\partial \eta_i} (\mathbb{E}[d_{\alpha \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]) &= - \sum_{t=1}^T \frac{\partial}{\partial \eta_i} \mathbb{E}[y_{it-1} f(\alpha y_{it-1} + \beta x_{it-1} + \eta_i) | y_{i0}, \eta_i, x_i] \\ &= f'(\alpha + \beta x_{it-1} + \eta_i) \Pr(y_{it-1} = 1 | y_{i0}, \eta_i, x_i) \\ &\quad + f(\alpha + \beta x_{it-1} + \eta_i) \frac{\partial}{\partial \eta_i} \Pr(y_{it-1} = 1 | y_{i0}, \eta_i, x_i), \end{aligned} \quad (4.30)$$

and respectively

$$\begin{aligned} \frac{\partial}{\partial \eta_i} (\mathbb{E}[d_{\eta_i \eta_i}(\theta, \eta_i) | y_{i0}, \eta_i, x_i]) &= - \sum_{t=1}^T \frac{\partial}{\partial \eta_i} \mathbb{E}[f(\alpha y_{it-1} + \beta x_{it-1} + \eta_i) | y_{i0}, \eta_i, x_i] \\ &= \frac{\partial}{\partial \eta_i} \Pr(y_{it-1} = 1 | y_{i0}, \eta_i, x_i) (f(\alpha + \beta x_{it-1} + \eta_i) \\ &\quad - f(\beta x_{it-1} + \eta_i)) \\ &\quad + \Pr(y_{it-1} = 1 | y_{i0}, \eta_i, x_i) (f'(\alpha + \beta x_{it-1} + \eta_i) \\ &\quad - f'(\beta x_{it-1} + \eta_i)) + f'(\beta x_{it-1} + \eta_i). \end{aligned} \quad (4.31)$$

The individual score corresponding to the slope parameters, *i.e.* β , can be computed in a similar way. Nevertheless, we must take into account the fact that $\partial F / \partial \beta_k = x_{it,k} f$, where $x_{it,k}$ is the k^{th} explanatory variable, which is known, contrary to y_{t-1} from the score corresponding to the lagged binary variable parameter α .

As for the covariance matrix, it is the inverse of the the MMLE Hessian matrix, which is calculated accounting for the fixed effects by using the following formula:¹¹

$$\begin{aligned} \sum_{i=1}^N & \left(\frac{\partial^2 \text{LogL}_i(\theta, \hat{\eta}_i(\theta))}{\partial \theta \partial \theta} + \frac{\partial^2 \text{LogL}(\theta, \eta_i)}{\partial \theta \partial \eta_i} \Big|_{\eta_i = \hat{\eta}_i(\theta)} \frac{\partial \hat{\eta}_i(\theta)}{\partial \theta} \right. \\ & \left. + \left[\frac{\partial^2 \text{LogL}_i(\theta, \eta_i(\theta))}{\partial \eta_i \partial \theta} \Big|_{\eta_i = \hat{\eta}_i(\theta)} + \frac{\partial^2 \text{LogL}_i(\theta, \eta_i(\theta))}{\partial \eta_i \partial \eta_i} \Big|_{\eta_i = \hat{\eta}_i(\theta)} \frac{\partial \hat{\eta}_i(\theta)}{\partial \theta} \right] \frac{\partial \hat{\eta}_i(\theta)}{\partial \theta} \right). \end{aligned} \quad (4.32)$$

¹¹It is also possible to correct for autocorrelation as in the case of time-series models, by using a “sandwich estimator” for the covariance matrix.

4.6.3 Appendix: Evaluation Methodology

To show the usefulness of the dynamic EWS, it is necessary to scrutinize their forecasting abilities (especially in out-of-sample exercises). To this aim, we implement the EWS evaluation toolbox developed by Candelon et al. (2012). The main advantage of this framework is that it can be applied to any EWS outputting crisis probabilities, both in-sample and out-of-sample. To be more precise, in a first step we rely on different evaluation criteria and comparison tests so as to identify the outperforming model, while in a second step we gauge the optimal model's ability to discriminate between crisis and calm periods by identifying the optimal cut-off for each country.

Accordingly, we consider both classic EWS evaluation measures, such as the QPS criterion and its corresponding comparison test Diebold-Mariano (1995) (as well as the nested version, Clark-West, 2007) and newer ones for the EWS literature, which take the cut-off into account in the evaluation and thus lead to a more refined diagnostic, *i.e.* the *Area Under the ROC criterion* (hereafter *AUC*). The QPS criterion is a mean-squared-error measure that compares the crisis probabilities (the forecasts issued by the EWS, $\text{Pr}_{t-1}(y_t = 1)$) with the crisis occurrence indicator y_t :

$$QPS = \frac{2}{T} \sum_{t=1}^T (\text{Pr}_{t-1}(y_t = 1) - y_t)^2. \quad (4.33)$$

At the same time, AUC is a credit-scoring criteria, that reveals the predictive abilities of an EWS by relying on all the values of the cut-off, *i.e.* the threshold used to compute crisis forecasts $\hat{y}_t(c)$, ($c \in [0, 1]$):

$$AUC = \int_0^1 [Se(c) \times (1 - Sp(c))] d(1 - Sp(c)), \quad (4.34)$$

where $Se(c)$ represents the *sensitivity*, *i.e.* the proportion of crises correctly identified by the EWS for a given cut-off c and $Sp(c)$ is the *specificity*, *i.e.* the proportion of calm periods correctly identified by the model for a cut-off equal to c . Furthermore, the null hypothesis of Diebold-Mariano (1995) and Clark-West (2007) tests is the equality of forecasting abilities of the two models. Their alternative is the statistical difference of forecasting abilities (the model with the lower loss according to a certain loss function is the better). For both tests, we use the MSFE loss function. The test statistic is

$$Stat = \frac{\sqrt{T}\bar{d}}{\sigma_{\bar{d},0}} \xrightarrow[T \rightarrow \infty]{d} N(0, 1), \quad (4.35)$$

where $\bar{d}$ is loss differential mean for the Diebold-Mariano test and where it is the mean of a modified loss differential in the case of the Clark-West test. Next, the optimal

cut-off is identified for each country by maximizing the Youden index (J) which is an accuracy measure arbitrating between type I and type II errors (misidentified crises and false alarms):

$$c^* = \arg \max_{c \in [0,1]} J(c), \quad (4.36)$$

where $J(c) = Se(c) + Sp(c) - 1$. A model's ability to correctly discriminate between crisis and calm periods is then given by *sensitivity* and *specificity*. This optimal cut-off can also be interpreted in terms of country vulnerability to crisis. The higher the cut-off and the larger the variance of the probability series during calm periods, the more the country seems prone to crisis. However, these results depend upon the underlying EWS model and should thus be analyzed with caution. For more details on this evaluation method, see Candelon et al. (2012).

Chapter 5

Modeling Financial Crises Mutation¹

The recent financial turmoils in Latin America and Europe have led to a concatenation of several events from currency, banking and sovereign debt crises. This paper proposes a multivariate dynamic probit model that encompasses the three types of 'currency, banking and sovereign debt' and allows us to investigate the potential causality between all three crises. To achieve this objective, we propose a methodological novelty consisting in an exact maximum likelihood method to estimate this multivariate dynamic probit model, thus extending Huguenin, Pelgrin and Holly's method (2009) to dynamic models.

Using a sample of emerging countries which experienced financial crises, we find that mutations from banking to currency (and vice-versa) are quite common. More importantly, the trivariate model turns out to be more parsimonious in the case of the two countries (Ecuador and South Africa) which suffered from the 3 types of crises. These findings are strongly confirmed by a conditional probability and an impulse-response function analysis, highlighting the interaction between the different types of crises and advocating hence the implementation of trivariate models whenever it is feasible.

5.1 Introduction

Since the tulipmania,² economic literature has recorded numerous turmoils affecting the foreign exchange market (currency crisis), the banking market (banking crisis) and the government foreign debt (sovereign debt market). Nevertheless, recent episodes have proved that most of the time crises do not remain restricted to a single market, but tend to spill-over into another one. Analyzing the crisis events over a period of a hundred years in a sample of 56 countries, Bordo et al. (2001b) have shown that the *ex-post* probability of twin, banking and currency crises has strongly increased since WWII. Similarly, using data back to the XIX century, Kaminsky and Reinhart (2008) present evidence of a strong

¹This chapter is based on Candelon, Dumitrescu, Hurlin and Palm (2011).

²Kindelberger (2000) calls this event the first financial crisis listed in history. It has affected the Dutch tulip market in 1636.

connection between debt cycles and economic crises in an analysis of both cross-country aggregates and individual country histories.

Nevertheless, some historical events showed that bidirectional feed-back between crises was not always sufficient to get an exhaustive picture of a turmoil. For example, the Ecuadorian crisis in 1999 affected first the banking sector, subsequently impacted simultaneously the Sucre³ and the country's public finance. More recently, the European crisis emerged as a banking distress succeeding the collapse of the U.S. real estate bubbles. It took a sovereign debt dimension when some European countries (as Greece, Ireland, Portugal), penalized by the recessive consequences of the banking credit crunch or by the public safety plans set up to stabilize the financial system, came close to default. A third dimension is now reached with the increase in volatility of the Dollar/Euro exchange rate as well as the rumors over a split of the Euro area.

The balance sheet approach provides the theoretical framework to analyze the potential spill-over from one crisis to another. Using such an accounting framework, Rosenberg et al. (2005) and, more recently, Candelon and Palm (2010) show how balance sheets are linked across sectors. Consecutively, the transmission of a shock from one country's economy to that of another country will become visible in their balance sheet. The financial crisis takes then another shape.

It appears thus evident that an accurate financial crisis model has to take the mutability of a crisis into account. In a seminal paper, Glick and Hutchinson (1999) model twin crises and assess the extent to which each type of crisis provides information about the likelihood of the other one. Their approach relies, in a first step, on individual models for currency and banking crises. In a second step, the global model is estimated by using the instrumental variables method so as to tackle the potential endogeneity bias. Implemented on a pooled sample of 90 industrial and developing countries over the 1975 – 1997 period, they find that the twin crisis phenomenon was most common in financially liberalized emerging markets during the Asian crisis. Nevertheless, from a methodological point of view, the use of a two-step approach is not free of criticism with respect to the endogeneity problem. Moreover, the use of a panel framework driven by the shortness of the time dimension will require some degree of homogeneity among countries. Finally, Glick and Hutchinson (1999) do not consider sovereign debt crises, focusing exclusively on twin crises.

This paper extends their study in several ways: First, it considers a multivariate dynamic probit model that encompasses the three types of crises (currency, banking and sovereign debt), thus allowing to investigate not only the potential mutation from currency to banking crises or vice versa, but also the mutability of them into a sovereign

³The Ecuadorian currency has been replaced by the U.S. dollar on March, 13, 2000.

debt crisis and vice-versa. Second, this paper introduces a methodological novelty by proposing an exact maximum likelihood approach to estimate this multivariate dynamic probit model. In a related study, Dueker (2005) estimates a dynamic qualitative VAR model of business cycle phase using simulation based methods.⁴ However, as shown by Huguenin, Pelgrin and Holly (2009) for a static model, a multivariate probit model cannot be accurately estimated using simulation methods. Its estimation requires hence to derive an exact maximum-likelihood function. We thus generalize the univariate dynamic probit model developed by Kauppi and Saikkonen (2008) to a multivariate level and derive its exact likelihood, allowing to obtain converging and efficient parameter estimates. Third, applied to a large sample of emerging countries, we show that in the bivariate case mutations of a banking crisis into a currency crisis (and vice-versa) have been quite common, confirming hence Glick and Hutchinson's results (1999). More importantly, for the two countries (Ecuador and South Africa) which suffered from the 3 types of crises, the trivariate model turns out to be more parsimonious, thus supporting its implementation anytime when it is feasible.

The rest of the paper is organized as follows. Section 5.2 presents a multivariate dynamic probit model. In section 5.3 we describe the Exact Maximum Likelihood method as well as some numerical procedures to estimate the multivariate dynamic probit model. In section 5.4, the multivariate dynamic probit model is estimated for 17 emerging countries in its bivariate (twin crises) or trivariate form.

5.2 A Multivariate Dynamic probit Model

In this section we propose and describe the multivariate dynamic probit model, that allows us to identify and characterize by different means crisis mutation phenomena among several markets.

Consider M latent continuous variables $y_{m,i,t}^*$ representing the pressure on the market m in country i , $i \in 1, \dots, I$ at time period $t \in 1, \dots, T$. The observed variable $y_{m,i,t}$ takes the value 1 if a crisis occurs on market m , in country i at period t and the value 0 otherwise. For simplicity, the country index is removed in the sequel of the paper as the model is estimated separately for each country. The term 'market' refers to the banking sector, the market for public debt and the foreign currency market. It is used as a synonym for the type of crisis. Denote by y_t^* and y_t the $M \times 1$ vectors with elements $y_{m,t}^*$ and $y_{m,t}$ respectively.

In line with Kauppi and Saikkonen (2008), consider the stochastic process y_t (M-variate) and x_t (K-variate), where y_t is a vector of binary variables taking on the values zero and one and x_t is a vector of explanatory variables. Define $F_t = [(y'_s, x'_s)' | s \leq t]$,

⁴See McFadden (1989) or Chib et Greenberg (1998) who proposed the simulation-based Bayesian and non-Bayesian estimation by MCMC of correlated binary data using the multivariate probit model.

as the information set available at time t . Assume that conditional on F_{t-1} , y_t has an M -variate Bernoulli distribution with probability p_t .

$$y_t|F_{t-1} \sim B(p_t). \quad (5.1)$$

Let us consider a multivariate conditional probit specification by assuming that the elements y_t are generated by

$$y_{m,t} = \mathbb{1}(y_{m,t}^* > 0), \quad (5.2)$$

with $\mathbb{1}(\cdot)$ being the indicator function and

$$y_{m,t}^* = \pi_{m,t} + \epsilon_{m,t}, \quad (5.3)$$

where $\pi_{m,t}$ represents the expected value of $y_{m,t}^*$ that may depend on covariates $x_{m,t}$ that may vary across markets, countries and time. Using vector notation, equation 5.3 becomes:

$$y_t^* = \pi_t + \epsilon_t, \quad (5.4)$$

with $\epsilon_t|F_{t-1}$ being i.i.d. multivariate normally distributed

$$\epsilon_t|F_{t-1} \sim IIN(0, \Omega). \quad (5.5)$$

A typical element $\omega_{m,m'}$ of Ω denotes the conditional covariance between $\epsilon_{m,t}$ and $\epsilon_{m',t}$, given the information set F_{t-1} .

The $M \times 1$ vector y_t^* is related to the conditional probability p_t through the common cumulative distribution function of ϵ_t , $\Phi(\cdot)$:

$$p_t = \Pr(y_t^* \geq 0|F_{t-1}) = \Pr(-\epsilon_t \leq \pi_t|F_{t-1}) = \Phi(\pi_t). \quad (5.6)$$

Recall, however, that our objective is to scrutinize the potential crisis mutation among three particular markets, namely currency, banking and sovereign debt. Accordingly, without any loss of generality, henceforth we restrict our attention to the trivariate form of the model. In this case $y_{m,t}$ denotes the type of crisis, $m \in \{1, 2, 3\}$, with $M = 3$.

Hence, crises can be modeled assuming that π_t in 5.4 is determined as follows:

$$\pi_t = \alpha + Bx_{t-1} + \Delta y_{t-1} + \Gamma \pi_{t-1}. \quad (5.7)$$

$y_t = (y_{1,t}, y_{2,t}, y_{3,t})'$, $\pi_t = (\pi_{1,t}, \pi_{2,t}, \pi_{3,t})'$, and $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$. B is a $3 \times k$ matrix and Δ and Γ are 3×3 matrices. The exogenous variables are specific to the type of crisis so there is no common cause among them. Then $x_{t-1} = (x'_{1,t-1}, x'_{2,t-1}, x'_{3,t-1})'$, where $x_{m,t-1}$ is a $(k_m \times 1)$ vector of explanatory variables corresponding to the m^{th} dependent

variable at time $t - 1$ and B is block-diagonal with the typical block on the diagonal, *i.e.* the row vector b'_m of slope coefficients corresponding to $x_{m,t-1}$. Similarly, we could also allow the dynamics to be crisis-specific by assuming that Δ and Γ are diagonal matrices. Obviously, when there are covariates in common for some crises, the number of variables in x_{t-1} , K , will be smaller than $\sum_{m=1}^3 k_m$, and B will not be block-diagonal.

Besides, $\theta_m = (\alpha_m; b_m; \delta_m; \gamma_m)'$ is the vector of parameters for equation m , with $\theta = (\theta'_1; \theta'_2; \theta'_3)'$. Finally, we assume that the error term has a trivariate normal distribution. The disturbances ε are i.i.d., with a trivariate normal distribution of zero mean and covariance matrix $V(\varepsilon) = I_T \otimes \Omega$, where Ω is a covariance matrix given by:

$$\Omega = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}, \quad (5.8)$$

where $\rho_{m,m-1}$ represent the correlation coefficients.

In this multivariate framework several ways to perceive the mutation mechanism can be designed.

1. Unobserved common factors can be taken into account through the contemporaneous inter-market dependence of the innovation terms ($\omega_{mm'} \neq 0$).
2. The unobservable latent variable $y_{m,t}^*$ depends on past values of other markets $y_{m',t-s}^*$, themselves unobservable, for $m \neq m'$. It implies that $\Delta = 0$ and $\Gamma = 0$. In such a case, $\pi_{m,t}$ depends only on the latent variables $y_{m',t-s}^*$, $s > 0$, which can be interpreted as a mutation phenomenon.
3. The unobservable latent variable $y_{m,t}^*$ on a specific market may depend on past crisis/calm periods on other markets. Formally, the pressure index $y_{m,t}^*$ depends on past values of the observable variable $y_{m',t}$, where $m \neq m'$ and thus the column m' of Δ is different from 0.
4. It is possible to combine the two previous cases, assuming that $y_{m,t}^*$ depends on both the latent variable $y_{m',t}^*$, and past crisis/calm periods, $y_{m',t}$, on other markets.
5. Crises dynamics can be modeled by considering that the latent variable $y_{m,t}^*$ depends on its lagged values, *i.e.* $\pi_{m,t}$ depends on $\pi_{m,t-1}$ via the matrix Γ .

Model 5.7 is a multivariate extension of the dynamic probit model recently proposed by Kauppi and Saikkonen (2008). This new specification enables us to compute not only marginal crisis probabilities, $\Pr(y_m = 1|y_m^*) = \Phi(y_m^*)$, as it is usually done in the literature, but also joint and conditional probabilities, *i.e.* $\Pr(y_1 = 1, y_2 = 1, y_3 = 1|y^*) =$

$\Phi_3(y^*)$, and $\Pr(y_m = 1|y_{m'}^*) = \Phi_3(y^*)/\Phi_2(y_{m'}^*)$, for $m \in \{1, 2, 3\}$, where Φ , Φ_2 and Φ_3 represent the univariate, bivariate and trivariate normal cumulative distribution functions respectively.

Remark:

Matrices Δ and Γ provide useful information about both crises persistence and mutation (causality). On the one hand, the diagonal terms of Γ specify the persistence of each crisis. These parameters correspond to a first order autoregressive representation of the latent variable. An increase in the pressure index during a certain period is always transmitted to the next period, hence linearly increasing the probability of a turmoil. The closer they are to 1, the more persistent the crisis episode will be. It is noticeable that the diagonal elements of this matrix are constrained to be strictly inferior to 1. We exclude the case where the latent variable $y_{m,t}^*$ follows a random walk, which would be empirically counter-intuitive as financial crises cannot be apprehended as persistent events.

At the same time, the diagonal terms of Δ also deliver information about crises persistence but that is somewhat different from that inferred from Γ . Indeed, they indicate to what extent the probability of occurrence of a crisis depends on the regime prevailing the period before. In this situation we observe the existence of threshold effects, as a tumultuous period lasts more than one spell only if the pressure index soars sufficiently to exceed a threshold which sets off a crisis in the previous period.

Altogether, we can distinguish between a linear crisis persistence captured through the diagonal terms of Γ , and a non-linear, threshold-based one, apprehended by the diagonal terms of Δ .

On the other hand, mutation is taken into account in the off-diagonal elements of the two matrices Γ and Δ . These Granger-causal effects between the three crises play a key role in the mutation of crises. As in the analysis of crises persistence, both a linear and a non-linear, threshold-effect transmission can be identified. A significant off-diagonal γ element shows that no sooner the pressure index on a specific market rockets than the index on another market rises. By contrast, a δ term reveals the presence of crisis transmission only if the corresponding pressure index is high enough to generate a crisis on the other market. In other words, if $\gamma_{1,2} > 0$ a high money market pressure at time $t-1$ increases the probability of a currency turmoil at t , whereas if $\delta_{1,2} > 0$ the probability of a currency turmoil increases at t as a consequence of a banking crisis that occurred the period before.

5.3 Exact Maximum Likelihood Estimation

The exact maximum likelihood estimator for the multivariate dynamic probit model cannot be obtained as a simple extension from the univariate model. For this reason, the

simulated maximum likelihood method is generally considered. Nevertheless, Huguenin, Pelgrin and Holly (2009) prove that it leads to a bias in the estimation of the correlation coefficients as well as in their standard deviations. Therefore, they advocate the exact maximum likelihood estimation. Since the correlations between the crisis binary variables, *i.e.* the contemporaneous transmission channels from one crisis to another one, constitute our main focus, asymptotic unbiased estimation of the correlations is of importance here and it calls for an explicit form of the likelihood. This section deals with this objective.

5.3.1 The Maximum Likelihood

Let us first notice that to identify the slope and covariance parameters, we impose that the diagonal elements of Ω to be standardized, *i.e.* equal to one. Following Greene (2002), the full information maximum-likelihood (FIML) estimates are obtained by maximizing the log-likelihood $\text{LogL}(Y|Z; \theta, \Omega)$, where θ is the vector of identified parameters and Ω is the covariance matrix. Under the usual regularity conditions⁵ (Lesaffre and Kauffmann, 1992), the likelihood is given by the joint density of observed outcomes:

$$L(y|z, \theta; \Omega) = \prod_{t=1}^T L_t(y_t|z_{t-1}, \theta; \Omega), \quad (5.9)$$

where $y_t = (y_{1,t}, y_{2,t}, y_{3,t})'$ and $y = [y_1, \dots, y_T]$. The individual likelihood $L_t(\cdot)$ is given in Lemma 1 as it is a well known result in the literature.

Lemma 1. *The likelihood of observation t is the cumulative density function, evaluated at the vector w_t of a 3-variate standardized normal vector with a covariance matrix $Q_t \Omega Q_t$:*

$$L_t(y_t|z_{t-1}, \theta; \Omega) = \Pr(y_1 = y_{1,t}, y_2 = y_{2,t}, y_3 = y_{3,t}) = \Phi_{3, \varepsilon_t}(w_t; Q_t \Omega Q_t), \quad (5.10)$$

where Q_t is a diagonal matrix whose main diagonal elements are $q_{m,t} = 2y_{m,t} - 1$ and thus depends on the realization or not of the events ($q_{m,t} = 1$ if $y_{m,t} = 1$ and $q_{m,t} = -1$ if $y_{m,t} = 0$, $\forall m \in \{c, b, s\}$). Besides, the elements of the vector $w_t = [w_{1,t}, \dots, w_{3,t}]$ are given by $w_{m,t} = q_{m,t} \pi_{m,t}$ (for a complete proof of Lemma 1, see Appendix 4.6.1).

Thus, the FIML estimates are obtained by maximizing the log-likelihood:

$$\text{LogL}(y|z, \theta; \Omega) = \sum_t^T \text{Log} \Phi_{3, \varepsilon}(w_t; Q_t \Omega Q_t) \quad (5.11)$$

with respect to θ and Ω ⁶.

⁵If the parameters θ are estimated while the correlation coefficients are assumed constant, the log-likelihood function is concave. In this case, the MLE exists and it is unique. Nevertheless, when θ and ρ are jointly estimated (as in our model), the likelihood function is not (strictly) log-concave as a function of ρ . Thus, the MLE exists only if the log-likelihood is not identically $-\infty$ and $E(z^T z|\rho)$ is upper semi-continuous finite and not identically 0. Furthermore, if no $\theta \neq 0$ fulfills the first order conditions for a maximum, the MLE of (θ, ρ) for the multivariate probit model exists and for each covariance matrix not on the boundary of the definition interval, the MLE is unique.

⁶Besides, we tackle the autocorrelation problem induced by the dependent binary crisis variables by considering a Gallant correction for the covariance matrix.

5.3.2 The Empirical Procedure

The main problem with FIML is that it requires the evaluation of high-order multivariate normal integrals, while existing results are not sufficient to allow accurate and efficient evaluation for more than two variables (see Greene, 2002, page 714). Indeed, Greene (2002) argues that the existing quadrature methods to approximate trivariate or higher-order integrals are far from being exact. To tackle this problem in the case of a static probit, Huguenin, Pelgrin and Holly (2009) decompose the triple integral into simple and double integrals, leading to an Exact Maximum Likelihood Estimation (EML) that requires computing double integrals. Most importantly, they prove that the EML increases the numerical accuracy of both the slope and covariance parameters estimates, which outperform the maximum simulated likelihood method (McFadden, 1989) which is generally used for the estimation of multivariate probit models. Therefore, we extend the decomposition proposed by Huguenin, Pelgrin and Holly, (2009) in the case of our multivariate dynamic model so as to obtain a direct approximation of the trivariate normal cumulative distribution function.

The EML log-likelihood function is given by:

$$\text{LogL}(y|z, \theta; \Omega) = \sum_{t=1}^T \text{Log} \left[\prod_{m=1}^3 \Phi(w_{m,t}) + G \right], \quad (5.12)$$

where $\Phi(w_t)$ is the univariate normal cumulative distribution function of w_t . Indeed, the log-likelihood function depends on the product of the marginal distributions (w_t) and the correction term G which captures the dependence between the m events analyzed.

The maximum likelihood estimators $\{\hat{\theta}; \hat{\Omega}\}_{EML}$ are the values of θ and Ω which maximize eq. 5.12:

$$\{\hat{\theta}; \hat{\Omega}\}_{EML} = \text{Arg max}_{\theta; \Omega} \sum_{m=1}^3 \text{LogL}(.), \quad (5.13)$$

with $L(.)$ given in (5.11).

Under the regularity conditions of Lesaffre and Kaufman (1992), the EML estimator of a multivariate probit model exists and is unique. Besides, the estimates $\{\hat{\theta}; \hat{\Omega}\}_{EML}$ are consistent and efficient estimators of the slope and covariance parameters and are asymptotically normally distributed. It is worth noting that in a correctly specified model for which the error terms are independent across the m equations the EML function corresponds to $\sum_{t=1}^T \prod_{m=1}^3 \Phi(w_{m,t})$, since the probability correction term G in eq. 5.12 tends toward zero.

We present here only the results for a bivariate and a trivariate model:

$$\Phi_2(w_t; Q_t \Omega Q_t) = \Phi(w_{1,t}) \Phi(w_{2,t}) \frac{1}{2\pi} \int_0^{\rho_{12}} \exp\left(-\frac{1}{2} \frac{w_{1,t}^2 + w_{2,t}^2 - 2w_{1,t}w_{2,t}}{1 - \lambda_{12}^2}\right) \frac{d\lambda_{12}}{\sqrt{1 - \lambda_{12}^2}} \quad (5.14)$$

for a bivariate model and

$$\begin{aligned} \Phi_3(w_t; Q_t \Omega Q_t) &= \prod_{m=1}^3 \Phi(w_{m,t}) + G \\ &= \Phi(w_{1,t}) \Phi(w_{2,t}) \Phi(w_{3,t}) \\ &\quad + \Phi(w_{3,t}) \int_0^{\rho_{12}} \phi_2(w_{1,t}, w_{2,t}; \lambda_{12}) d\lambda_{12} \\ &\quad + \Phi(w_{2,t}) \int_0^{\rho_{13}} \phi_2(w_{1,t}, w_{3,t}; \lambda_{13}) d\lambda_{13} \\ &\quad + \Phi(w_{1,t}) \int_0^{\rho_{23}} \phi_2(w_{2,t}, w_{3,t}; \lambda_{23}) d\lambda_{23} \\ &\quad + \int_0^{\rho_{12}} \int_0^{\rho_{13}} \frac{\partial \phi_3(w_t; \lambda_{12}, \lambda_{13}, 0)}{\partial w_{1,t}} d\lambda_{12} d\lambda_{13} \\ &\quad + \int_0^{\rho_{12}} \int_0^{\rho_{23}} \frac{\partial \phi_3(\dot{w}_t; \lambda_{12}, 0, \lambda_{23})}{\partial w_{2,t}} d\lambda_{12} d\lambda_{23} \\ &\quad + \int_0^{\rho_{13}} \int_0^{\rho_{23}} \frac{\partial \phi_3(\ddot{w}_t; 0, \lambda_{13}, \lambda_{23})}{\partial w_{3,t}} d\lambda_{13} d\lambda_{23} \\ &\quad + \int_0^{\rho_{12}} \int_0^{\rho_{13}} \int_0^{\rho_{23}} \frac{\partial^3 \phi_3(\dot{\dot{w}}_t; \lambda_{12}, \lambda_{13}, \lambda_{23})}{\partial w_{1,t} \partial w_{2,t} \partial w_{3,t}} d\lambda_{12} d\lambda_{13} d\lambda_{23} \end{aligned} \quad (5.15)$$

for a trivariate model, where ρ are the non-diagonal elements of the $Q_t \Omega Q_t$ matrix and λ are the non-diagonal elements of a theoretical 2×2 matrix and respectively a 3×3 matrix in which one of the correlation coefficients is null. Moreover, $\dot{w}_t$ is a vector of indices obtained by changing the order of the elements to $(w_{2,t}, w_{3,t}, w_{1,t})$. Similarly $\ddot{w}_t$ corresponds to a vector of indices of the form $(w_{3,t}, w_{1,t}, w_{2,t})$. Finally, $\dot{\dot{w}}_t$ corresponds to $w_t, \dot{w}_t$ or $\ddot{w}_t$ respectively, depending on the way the last integral is decomposed. The computation of the last term is not trivial. However, this integral can be decomposed in a non-unique way, as follows:

$$\begin{aligned}
 & \int_0^{\rho_{12}} \int_0^{\rho_{13}} \int_0^{\bar{\rho}_{23}} \frac{\partial^3 \phi_3(\dot{w}_t; \lambda_{12}, \lambda_{13}, \lambda_{23})}{\partial w_{1,t} \partial w_{2,t} \partial w_{3,t}} d\lambda_{12} d\lambda_{13} d\lambda_{23} \\
 &= \int_0^{\rho_{13}} \int_0^{\rho_{23}} \frac{\partial \phi_3(\ddot{w}_t; \lambda_{12}, \lambda_{13}, \lambda_{23})}{\partial w_{3,t}} d\lambda_{13} d\lambda_{23} - \int_0^{\rho_{13}} \int_0^{\rho_{23}} \frac{\partial \phi_3(\ddot{w}_t; 0, \lambda_{13}, \lambda_{23})}{\partial w_{3,t}} d\lambda_{13} d\lambda_{23} \\
 &= \int_0^{\rho_{12}} \int_0^{\rho_{23}} \frac{\partial \phi_3(\dot{w}_t; \lambda_{12}, \lambda_{13}, \lambda_{23})}{\partial w_{2,t}} d\lambda_{12} d\bar{\rho}_{23} - \int_0^{\rho_{12}} \int_0^{\rho_{23}} \frac{\partial \phi_3(\dot{w}_t; \lambda_{12}, 0, \lambda_{23})}{\partial w_{2,t}} d\lambda_{12} d\lambda_{23} \\
 &= \int_0^{\lambda_{12}} \int_0^{\lambda_{13}} \frac{\partial \phi_3(w_t; \lambda_{12}, \lambda_{13}, \lambda_{23})}{\partial w_{1,t}} d\lambda_{12} d\lambda_{13} - \int_0^{\rho_{12}} \int_0^{\rho_{13}} \frac{\partial \phi_3(w_t; \lambda_{12}, \lambda_{13}, 0)}{\partial w_{1,t}} d\lambda_{12} d\lambda_{13}.
 \end{aligned} \tag{5.16}$$

These finite-range multiple integrals are numerically evaluated by using a Gauss-Legendre Quadrature rule⁷ over bounded intervals. In such a context, two possibilities can be considered: whether the likelihood function is directly maximized, or the first order conditions⁸ are derived so as to obtain an exact score vector. As stressed by Huguenin, Pelgrin and Holly (2009), the two methods may not lead to the same results if the objective function is not sufficiently smooth.

5.4 Empirical Application

This section aims at implementing the multivariate dynamic probit methodology presented above to a system composed by three types of crises, *i.e.* currency, banking and sovereign debt crises. We thus evaluate the probability of mutation of one type of crisis into another one. After a short data description and the presentation of the criteria implemented to detect the three types of crises, we estimate bivariate models by excluding sovereign debt crises. This constitutes a benchmark for the second part where the sovereign debt crises are included in the system.

5.4.1 Dating the crises

5.4.1.1 The Database

Monthly macroeconomic indicators expressed in US dollars covering the period from January 1985 to June 2010 have been extracted for 17 emerging countries⁹ from the IMF-IFS database as well as from the national banks data of the countries under analysis via Datastream.¹⁰ The government bond returns are obtained via the JPMorgan EMDB database. More exactly, we have selected the main leading indicators used in the literature

⁷Details about this quadrature are available in Appendix 4.6.2.

⁸The score vector of the trivariate probit model is presented in Appendix 4.6.3.

⁹Argentina, Brazil, Chile, Colombia, Ecuador, Egypt, El Salvador, Indonesia, Lebanon, Malaysia, Mexico, Panama, Peru, Philippines, South Africa, Turkey and Venezuela.

¹⁰We choose not to include any European country, as 1) only few of them have suffered from the three types of crises and 2) if this is the case it corresponds to a single episode: the recent turmoil.

for the three types of crises that we analyze (see Candelon et al., 2009, 2012; Lestano et al., 2003; Glick and Hutchison, 1999; Hagen and Ho, 2004; Pescatori and Sy, 2007), namely, the one-year growth rate of international reserves, the growth rate of M2 to reserves ratio, one-year growth of domestic credit over GDP ratio, one-year growth of domestic credit, one-year growth of GDP, government deficit, debt service ratio and external debt ratio.

5.4.1.2 Dating the Crisis Periods

1. The Currency Crises

Currency crises are generally identified using the market pressure index (MPI), which is a linear combination between exchange rate and foreign reserves changes. Hence, if the pressure index exceeds a predetermined threshold,¹¹ a crisis period is identified. As in Lestano and Jacobs (2004) and Candelon et al. (2009, 2012), a modified version of the pressure index proposed by Kaminski et al. (1998) which also incorporates the interest rate is used. It is denoted by (KLRm) and takes the form of a weighted average of the percentage of the exchange rate, of the foreign reserves and of the change in the domestic interest rate:

$$\text{KLRm}_{i,t} = \frac{\Delta e_{i,t}}{e_{i,t}} - \frac{\sigma_e}{\sigma_{res}} \frac{\Delta res_{i,t}}{res_{i,t}} + \frac{\sigma_e}{\sigma_r} \Delta r_{i,t}, \quad (5.17)$$

where $e_{i,t}$ denotes the exchange rate (*i.e.*, units of country i 's currency per US dollar in period t), $res_{i,t}$ represents the foreign reserves of country i in period t (expressed in US\$), while $r_{i,t}$ is the interest rate in country i at time t . σ_x denotes the standard deviation of the relative changes in the variable $\Delta x_{i,t}/x_{i,t}$, where x denotes each variable separately, including the exchange rate and the foreign reserves. For the interest rate, we consider the absolute changes $\Delta x_{i,t} = x_{i,t} - x_{i,t-6}$.¹² For both subsamples, the currency crisis ($CC_{i,t}$) threshold equals 1.5 standard deviations above the mean:

$$CC_{i,t} = \begin{cases} 1, & \text{if } \text{KLRm}_{i,t} > 1.5\sigma_{\text{KLRm}_{i,t}} + \mu_{\text{KLRm}_{i,t}} \\ 0, & \text{otherwise.} \end{cases} \quad (5.18)$$

2. The Banking Crises

Banking crises are most commonly identified using the banking sector balance sheet, policy responses to bank runs and bank failures on a yearly basis (see the recent dating of Leaven and Valencia (2008)). Nevertheless, our crisis dating requires a monthly frequency. Moreover, Eichengreen (1996, 1998) notices that banking crises are not

¹¹Usually fixed to 2 or 3 times the sample's standard deviation as in Kaminski et al. (1998).

¹²Additionally, we take into account the existence of higher volatility in periods of high inflation, and consequently the sample is split into high and low inflation periods. The cut-off corresponds to a six-month inflation rate being higher than 50%.

always associated with a visible policy intervention. Indeed, some interventions may take place in the absence of a crisis in order to solve structural economic problems and perhaps to prevent a crisis. Besides, some measures can be taken only when the crisis has spread to the whole economy. Thus, Hagen and Ho (2004) propose a money market pressure index, accounting for the increasing demand for central bank reserves, to identify banking crises. Thus, it resembles a banking pressure index (BPI), available at monthly frequency:

$$BPI_{i,t} = \frac{\Delta \gamma_{i,t}}{\sigma_{\Delta \gamma}} + \frac{\Delta r_{i,t}}{\sigma_{\Delta r}}, \quad (5.19)$$

where $\gamma_{i,t}$ is the ratio of reserves to bank deposits in country i at time t , r is the real interest rate, Δ operator represents the six-month difference operator, and $\sigma_{\Delta \gamma}$ and $\sigma_{\Delta r}$ are the standard deviations of the two components. Sharp increases in the indicator (greater than the 90th percentile denoted as $P_{BPI,90}$) signal a banking crisis:

$$BC_{i,t} = \begin{cases} 1, & \text{if } IMP_{i,t} > P_{BPI,90,i} \\ 0, & \text{otherwise.} \end{cases} \quad (5.20)$$

3. The Sovereign Debt Crises

Countries' 'default' does not constitute an adequate measure to characterize a sovereign debt crisis. Indeed, a country may face debt-servicing difficulties or problems to re-finance its debt on the international capital markets without being in default. In order to overcome this problem, Pescatori and Sy (2007) consider a market-oriented measure of debt-servicing difficulties based on sovereign bond spreads.

In the line of this study, we consider that a sovereign debt crisis ($SC_{n,t}$) occurs if the CDS spreads exceed a critical threshold estimated by using kernel density estimation. More precisely, the existence of a mode around high spread values can be used to define crisis and calm periods, since whenever spreads are close to a limit that cannot be passed smoothly, the observations will concentrate around it until the limit is finally broken or the increasing pressure is reduced. Additionally, as expected, this estimated threshold corresponds to a percentile between the 90th and the 99th percentiles, depending on the country (the number of crisis periods varies from one country to another), since crises are extreme events:

$$SC_{i,t} = \begin{cases} 1, & \text{if } CDS_{spread}_{i,t} > Kernel\ Threshold_i \\ 0, & \text{otherwise.} \end{cases} \quad (5.21)$$

It is worth noting that most of the crisis periods we have identified by using the three aforementioned methods correspond to the ones reported in the literature on financial crises, *e.g.* Reinhart and Rogoff (2008).

5.4.1.3 Remarks

1. As in Kumar (2003), we dampen the magnitude of every variable using the formula : $f(x_t) = \text{sign}(x_t)\log(1 + |x_t|)$ so as to reduce the impact of extreme values.¹³ The model with transformed variables should hence have a normal error term.
2. It should also be noted that the entire sample is used for the identification of currency and banking crises, while the identification of debt crises is realized by using data from December 1997 (see Table 5.1) since the CDS spread series used for the identification of sovereign debt crises are not available before 1997 in the JPMorgan EMDB database. Consequently, our empirical analysis will consist of two parts, the first one analyzing the case of twin crises (currency and banking) for which the entire database can be used, while the second part focuses on the interactions between the three types of crises and is thus based on data from 1997 onwards. The data sample actually used for each of the 17 countries and the two types of analyses is available in Table 5.1.
3. We only retain the countries for which the percentage of crisis periods is superior to 5% (see Table 5.2).¹⁴
4. As mentioned in section 5.2, there are three dynamic multivariate specifications that can be used. However, as shown by Candelon et al. (2010), the dynamic model including the lagged binary crisis variables seems to be the best choice according to model selection using the Akaike information criterion. However, since we cannot expect a crisis to have a certain impact on the probability of emergence of another type of crisis from one month to another, which would justify the notation $y_{m,t-1}$ from the theoretical part, in the empirical application we consider a response lag l of 3, 6 and respectively 12 months for the bivariate models and one of 3 or 6 months for the trivariate models¹⁵. Therefore, for each type of crisis we build a lagged variable $y_{m,t-k}$ which takes the value of one if there was crisis in the past k periods or at time t , and the value of 0 otherwise:

¹³Missing values throughout the series are replaced by cubic spline interpolation.

¹⁴Argentina, Chile, Ecuador, Egypt, Indonesia, Lebanon, Mexico, South Africa and Venezuela are included in the bivariate analysis, whereas a trivariate model is specified for Ecuador and South Africa. Since the threshold has been arbitrarily set to 5%, we have also checked the borderline countries, like Colombia or Turkey in the bivariate analysis and Egypt in the trivariate analysis respectively, and similar results have been obtained.

¹⁵A 12-months lag is not used in the case of trivariate models since it would significantly reduce the already small number of observations we have at our disposal.

$$y_{m,t-k} = \begin{cases} 1, & \text{if } \sum_{j=0}^k y_{m,t-j} > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (5.22)$$

5. The significance of the parameters of each model is tested by using simple t-statistics based on robust estimates of standard-errors (which rely on a Gallant kernel, as in Kauppi and Saikkonen, 2008). A special attention is given to the interpretation of cross-effects which stand for the transmission channels of the shocks/crisis. Besides, the joint nullity of the contemporaneous correlations between shocks is tested using a log-likelihood ratio test for the trivariate models.

Table 5.1: Database

Country	Bivariate model	Trivariate model
Argentina	February 1988 - May 2010	December 1997 - May 2010
Brazil	September 1990 - May 2010	December 1997 - May 2010
Chile	January 1989 - May 2009	May 1999 - May 2010
Colombia	February 1986 - August 2009	December 1997 - August 2009
Ecuador	January 1994 - November 2007	December 1997 - November 2007
Egypt	February 1986 - June 2009	July 2001 - June 2009
El Salvador	January 1991 - November 2008	April 2002 - November 2008
Indonesia	January 1989 - August 2009	May 2004 - August 2009
Lebanon	January 1989 - April 2010	April 1998 - April 2010
Malaysia	January 1988 - March 2010	December 1997 - March 2010
Mexico	January 1988 - May 2010	December 1997 - May 2010
Peru	January 1990 - May 2010	December 1997 - May 2010
Philippines	January 1995 - February 2008	December 1997 - February 2008
South Africa	January 1988 - August 2009	December 1997 - August 2009
Turkey	January 1988 - May 2010	December 1997 - May 2010
Venezuela	February 1986 - November 2009	December 1997 - November 2009

Note: Data sample.

5.4.2 Bivariate Analysis

Along the lines of Kaminsky et al. (1998) it is possible to find a large number of explanatory variables that may signal the occurrence of a crisis. Nevertheless, Candelon et al. (2010) showed that a univariate dynamic probit model presents the advantage of yielding plausible results while being fairly parsimoniously parametrized. Indeed, a large part of the information is integrated either in the past state variable or in the lagged index and thus only a few explanatory variables turn out to be significant. In this context, we expect their multivariate (bivariate or trivariate) extension to be even more parsimonious. Therefore, we consider the four explanatory variables which are significant

Table 5.2: Percentage of crisis periods

	Bivariate model		Trivariate model		
	Currency crisis	Banking crisis	Currency crisis	Banking crisis	Debt crisis
Argentina	5.13	8.90	4.00	6.67	10.0
Brazil	3.77	7.19	0.00	3.33	2.67
Chile	6.07	10.0	5.79	5.79	3.31
Colombia	4.95	9.90	9.22	12.8	0.00
Ecuador	5.73	9.93	6.67	10.8	6.67
Egypt	6.76	9.96	4.17	7.30	7.30
El Salvador	3.65	9.85	0.00	0.00	2.50
Indonesia	5.30	9.90	0.00	14.0	6.25
Lebanon	9.62	9.96	1.38	8.97	2.76
Malaysia	3.10	10.0	4.05	6.08	4.73
Mexico	6.50	9.93	0.00	9.33	0.00
Panama	0.00	9.89	0.00	6.38	0.00
Peru	4.45	8.22	0.00	10.7	0.00
Phillipines	4.90	9.80	5.69	6.50	3.25
South Africa	6.71	9.89	7.09	7.80	4.26
Turkey	4.80	8.56	4.00	6.67	0.00
Venezuela	7.33	10.1	4.17	7.64	2.78

Note: The entries represent the proportion of crises period over the whole sample. It is indicated in bold as it exceeds 5%.

in Candelon et al. (2010), *i.e.* one-year growth of international reserves, one-year growth of M2 to reserves for currency crises as well as one-year growth of domestic credit over GDP and one-year growth of domestic credit for banking crises, resulting in four different specifications including one explanatory variable for each type of crisis. Moreover, three different lags (3 months, 6 months and 12 months) are considered for the lagged binary variable $y_{m,t-k}$. The dynamic probit model is estimated country-by-country using the Exact maximum likelihood.¹⁶ It is indeed a simplification as contagion (or spill-over) from one country to another is not taken account. A panel version of the model would lead to several problems. First, as shown by Berg et al. (2008), heterogeneity due to country specificities would have to be accounted for. Second, the estimation of a fixed effect panel would be biased without a correction of the score vector.¹⁷ Third, in a country-by-country analysis contagion has to be ignored. For all these reasons, we consider this extension to be beyond the scope of this paper and leave it for future research.

Each model is estimated via maximum-likelihood, the bivariate normal cumulative distribution function being approximated using the Gauss-Legendre quadrature, as proposed by Huguenin, Pelgrin and Holly, (2009). However, the quadrature specified in Matlab by default, *i.e.* the adaptive Simpson quadrature, has been considered as a benchmark.

¹⁶Initial conditions are introduced as given by the univariate static probit.

¹⁷See Candelon et al., (2010) for a discussion about this point.

Information criteria, namely AIC and SBC, are used to identify the best model for each country. The specification including the lagged binary variable turns out to be preferred. Optimal lag lengths are determined similarly. It is nevertheless worth stressing that the results are generally robust to the choice of explanatory variables and even to the choice of lags.

A summary of the results for the selected models is given in Table 5.3.

Table 5.3: Bivariate Analysis

Country		3 months		6 months		12 months	
		Δ	Ω	Δ	Ω	Δ	Ω
Argentina	currency	$\begin{bmatrix} + & + \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$
	banking	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$
Chile	currency	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$	$\begin{bmatrix} . & - \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$
	banking	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$
Ecuador	currency	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$
	banking	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$
Egypt	currency	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & - \end{bmatrix}$	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & - \end{bmatrix}$
	banking	$\begin{bmatrix} - & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$	$\begin{bmatrix} - & + \end{bmatrix}$	$\begin{bmatrix} - & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} - & 1 \end{bmatrix}$
Lebanon	currency	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$
	banking	$\begin{bmatrix} - & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$
Mexico	currency	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$
	banking	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$
South Africa	currency	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} 1 & . \end{bmatrix}$
	banking	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} . & 1 \end{bmatrix}$
Venezuela	currency	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$	$\begin{bmatrix} + & . \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$	$\begin{bmatrix} . & . \end{bmatrix}$	$\begin{bmatrix} 1 & + \end{bmatrix}$
	banking	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$	$\begin{bmatrix} . & + \end{bmatrix}$	$\begin{bmatrix} + & 1 \end{bmatrix}$

Note: Three different lags of the dependent variable are used, namely 3, 6 and 12 months. ' Δ ' stands for the parameters of the lagged crisis variables, while Ω represents the covariance matrix. A '+'/'-' sign means that the coefficient is significant and positive/ negative, while a '.' indicates its non-significance. For example, in the case of Argentina, 3 months, all the parameters are positive and significant except for the impact of a currency crisis on the probability of occurrence of banking crises. Similarly, the correlation coefficient between currency and banking crises is significant.

First of all, it seems that most of the models exhibit dynamics, whatever the lag used to construct the 'past crisis' variable. This result confirms the findings of Candelon et al. (2010) and Bussière (2007), showing that crises exhibit a regime dependence: if the country is proven to be more vulnerable than investors had initially thought, investors will start withdrawing their investments, thus increasing the probability of a new crisis. More precisely, most of the countries are found to have experienced banking and currency crises, with a significant autoregressive coefficient, i.e the crisis variable depends on its

own past, *e.g.* Argentina, Egypt, Lebanon, Mexico, South Africa, Venezuela. Besides, only for a small number of cases, only one of the two types of crises is best reproduced by a dynamic model (currency crises in Chile (3 and 12 months), Mexico (6 and 12 months); banking crises in Argentina (6 and 12 months), Ecuador, Lebanon (6 months), South Africa (12 months) and Venezuela (12 months)). Actually, in Chile a past currency crisis has only a short term positive impact on the emergence of another currency crisis, whereas a banking crisis has just a long term effect on the probability of occurrence of another banking crisis. Mexico, however, seems to be more prone to recurring currency crises than banking crises as the former type of crisis has a long-term impact on the probability of experiencing a new crisis, whereas the latter has a positive effect only in the short run. On the contrary, for Argentina, South Africa and Venezuela the impact of past banking crises on currency crises is longer (up to one year) as opposed to that of past currency crises on banking ones (up to three and six months, respectively).

Second, for the majority of these countries (Argentina, Chile, Lebanon, Mexico and Venezuela), currency and banking crises are interconnected. This link between crises can take two forms. On the one hand, a certain type of crisis increases (or diminishes) the probability of occurrence of the other type of crisis. This strong link between banking to currency crisis was emphasized by Glick and Hutchinson (1999) within a panel framework. Nevertheless, there is no reason for the transmission of shocks to be symmetric. Indeed, our country per country analysis reveals that for some countries like Argentina (3 and 6 months) a banking crisis in the past months increased the probability of a currency crisis at time t . At the same time, a banking crisis in Chile in the last 12 months reduced the probability of experiencing a currency crisis. Conversely, a currency crisis in Egypt and in Lebanon (3 months) diminished the probability of a banking crisis.

On the other hand, crisis shocks can be contemporaneously positively correlated. This feature seems to be very stable across models (independent of the lag used). The only exceptions are Egypt and Lebanon, for which there is no instantaneous correlation in the model with 3-months lagged binary variables, and Mexico, for which such a correlation appears only for the 12-months lag. To sum up, but for Egypt, all countries are characterized by a positive instantaneous correlation between shocks of currency and banking crises variables, corroborating the previous findings of Glick and Hutchinson (1999).

Third, the macroeconomic variables are rarely significant.¹⁸ These results go along the lines of our previous findings (see Candelon et al. 2010) that the dynamics of crises captures most of the information explaining the emergence of such phenomena. Furthermore, when these coefficients are significant, they have the expected sign (an increase in

¹⁸These results are available upon request.

the growth of international reserves diminishes the probability of a crisis, while a surprise in the rest of indicators soars the probability of a crisis).

To summarize, these results confirm the presence of interaction between the banking and currency crises. The twin crisis phenomenon is thus confirmed empirically. Besides, our findings are robust to the quadrature choice and the lags considered when constructing the dynamic binary variables.

5.4.3 Trivariate Analysis

But is it really enough to look at two crises only? This subsection extends the previous analysis to the trivariate case by modeling simultaneously the occurrence of currency, banking and debt crises. However, only two countries experienced these three events during a sufficiently long period. Ecuador presents for our sample an *ex-post* probability larger than 5% for whatever the type of crisis. Such a result is not surprising if one remembers that Ecuador faced a strong financial turmoil in the late 1990, affecting first the banking sector,¹⁹ then the Sucre²⁰, and the government budget. Jacone (2004) showed that institutional weaknesses, rigidities in public finances, and high financial dollarization have amplified this crisis. South Africa constitutes a borderline case as the sovereign debt crisis probability is slightly below 5%.

Each of the models is estimated for these countries using both the methodology proposed by Huguenin et al. (2009) based on the Gauss-Legendre quadrature and the direct approximation of a triple integral based on the adaptive Simpson quadrature that Matlab uses by default. Similar results are obtained for the two methods.²¹ However, the latter implies a significant gain in time without any loss in accuracy proving that recently developed quadrature methods are good approximations of the normal cumulative distribution function. Besides, 6 and 12 month-lags of the dynamic crisis variable are considered.

In the case of Ecuador, the results corroborate our bivariate findings: the banking crises are persistent, while currency crises are not. Nevertheless, it is clear that the bivariate model is misspecified, since it cannot capture the impact of a banking crisis on the occurrence of a currency crisis when using the 6-months lagged binary variables to account for the dynamics of these phenomena (see Table 5.4).

Moreover, the trivariate model turns out to be more parsimonious in terms of parameters to be estimated since the past debt crisis indicator has a positive effect on the probability of occurrence of both currency and debt crises. Therefore it supports the implementation of a trivariate crisis model whenever it is feasible. We also observe that the contemporaneous correlation matrix is diagonal, ruling out common shocks. Crises in

¹⁹16 out of the 40 banks existing in 1997 faced liquidity problems.

²⁰The Ecuadorian currency has been replaced by the U.S. dollar on March, 13, 2000.

²¹The results for Ecuador when considering a 6-months lag have been obtained with Matlab's quadrature since the model based on the Gauss-Legendre Quadrature did not converge.

Table 5.4: Trivariate Analysis

Country		3 months		6 months	
		Δ	Ω	Δ	Ω
Ecuador	currency	$\begin{bmatrix} . & . & + \end{bmatrix}$	$\begin{bmatrix} 1 & . & . \end{bmatrix}$	$\begin{bmatrix} . & + & + \end{bmatrix}$	$\begin{bmatrix} 1 & . & . \end{bmatrix}$
	banking	$\begin{bmatrix} . & + & . \end{bmatrix}$	$\begin{bmatrix} . & 1 & . \end{bmatrix}$	$\begin{bmatrix} . & + & . \end{bmatrix}$	$\begin{bmatrix} . & 1 & . \end{bmatrix}$
	sovereign	$\begin{bmatrix} . & . & + \end{bmatrix}$	$\begin{bmatrix} . & . & 1 \end{bmatrix}$	$\begin{bmatrix} . & . & + \end{bmatrix}$	$\begin{bmatrix} . & . & 1 \end{bmatrix}$
South Africa	currency	$\begin{bmatrix} + & . & . \end{bmatrix}$	$\begin{bmatrix} 1 & . & + \end{bmatrix}$	$\begin{bmatrix} + & . & . \end{bmatrix}$	$\begin{bmatrix} 1 & . & + \end{bmatrix}$
	banking	$\begin{bmatrix} . & . & . \end{bmatrix}$	$\begin{bmatrix} . & 1 & . \end{bmatrix}$	$\begin{bmatrix} . & . & . \end{bmatrix}$	$\begin{bmatrix} . & 1 & . \end{bmatrix}$
	sovereign	$\begin{bmatrix} . & . & + \end{bmatrix}$	$\begin{bmatrix} + & . & 1 \end{bmatrix}$	$\begin{bmatrix} . & . & + \end{bmatrix}$	$\begin{bmatrix} + & . & 1 \end{bmatrix}$

Note: Two different lags of the dependent variable are used, namely 3 and 6 months. ' Δ ' stands for the parameters of the lagged crisis variables, while Ω represents the covariance matrix. A '+'/'-' sign means that the coefficient is significant and positive/negative, while a '.' indicates its non-significance. For example, in the case of Ecuador, 3 months, sovereign debt crises have a positive and significant impact on the probability of occurrence of currency crises.

Ecuador turn out to be exclusively driven by transmission channels, as in the late 1990, when the banking distress was diffused to the currency and the government budget.

In the case of South-Africa, both currency and debt crises are dynamic. There is no evidence of causality between the different types of crises, but significant contemporaneous correlation. It highlights the fact that contrary to Ecuador, South African crises did not mutate but they originated from a common shock. It is worth noting that the results are found to be robust to the sensitivity analyses performed, namely to the choice of macroeconomic variables and the use of different lags for the past crisis variables.

5.4.4 Further results

To grasp better the properties of the models estimated and selected, a conditional probability analysis as well as an Impulse Response Functions (IRF) analysis are provided. For sake of space, we only report the results obtained for Ecuador.²²

First, Figure 5.1 reports the conditional probabilities for each type of crisis obtained from both the bi- and trivariate models considering a forecast horizon of 3 and 6 months. To allow a fair comparison, both models are estimated on the same sample, *i.e.* from 1997 onwards. It goes without saying that the bivariate model does not provide any conditional probabilities for sovereign debt crisis.

It turns out that the trivariate model outperforms the bivariate one whatever the forecast horizon is, *i.e.* the conditional probabilities issued from the trivariate model are higher than those obtained from the bivariate model during observed crisis periods, while they appear to be similar for calm periods. Such results corroborate hence our previous findings, stressing that a crisis model should take into account the whole sequence of

²²For South Africa, crisis mutation is exclusively driven by the contemporaneous correlation matrix as indicated in Table 5.4. Otherwise we can see that currency and sovereign debt crises are more persistent than banking ones. All figures are available from the authors upon request.

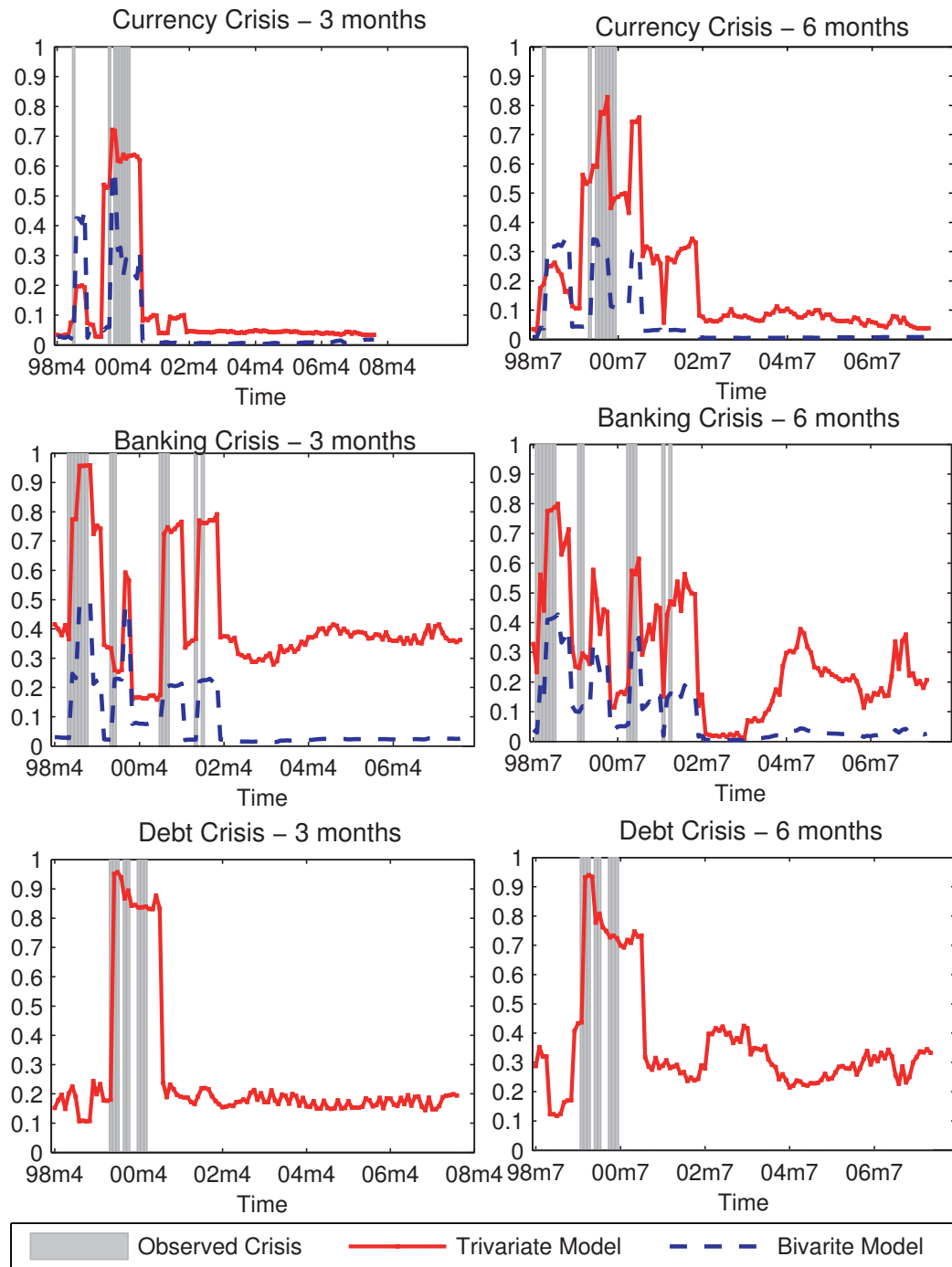


Figure 5.1: Conditional crisis probabilities - Ecuador

Note: Probabilities at time t are calculated including observed information prior 3 or 6 months.

crises to be accurate. Besides, the conditional probabilities obtained from the trivariate model do not immediately collapse after the occurrence of the crisis, which is the case for the bivariate model. It stresses hence the vulnerability of the economy after the exit from a turmoil, in particular if it affects the foreign exchange market.

Second, to evaluate the effect of a crisis, considered here as a shock, an IRF analysis is performed for the trivariate model. As the order of the variables has been shown to be crucial, we consider the historical sequence of crises observed in Ecuador, *i.e.* banking crises (the most exogenous ones), debt crises and currency crises (the most endogenous ones). Orthogonal impulse response functions are considered on the latent variable for a 3-month horizon. The exogenous variables are fixed to the unconditional mean ($\bar{x}_{m,t}$). Departing from eq. 5.7, we express the IRF in terms of the latent model, *i.e.* the probability of being in a crisis state at time t and the binary crisis/calm variable as follows:

$$\begin{aligned} y_{m,t}^* &= \hat{\alpha}_m + \bar{x}_{m,t-1}\hat{\beta}_m + \Delta_{m,m'}\tilde{y}_{m',t-1} + \hat{\varepsilon}_{m,t}, \\ \Pr(\hat{y}_{m,t} = 1) &= \Phi_3(y_{m,t}^*), \\ \hat{y}_{m,t} &= \mathbf{1}(y_{m,t}^* > 0) = \mathbf{1}(\Pr(\hat{y}_{m,t} = 1) > 0.5), \end{aligned} \tag{5.23}$$

where $\hat{\alpha}$, $\hat{\beta}$, $\hat{\Delta}$ are obtained from the estimated trivariate model for Ecuador and the correlated residuals $\hat{\varepsilon}_{m,t}$ are transformed into orthogonal ones via a Choleski decomposition of the covariance matrix $\hat{\Omega}$. Therefore, a crisis is to arise at time t , if $\hat{y}_{m,t} = 1$.

Additionally, as in any non-linear model, the IRFs are calculated for two initial states: a tranquil one, $\tilde{y}_{t-1} = 0$, *i.e.* “no type of crisis is observed at time $t = 0$ or in the previous 3 months” and a turmoil regime, $\tilde{y}_{t-1} = 1$, *i.e.* “all types of crisis are observed in $t = 0$ ”. Confidence intervals are built taking the 2.5% and 97.5% percentiles of IRF’s distribution obtained from 10,000 simulations of the model. The magnitude of the shock²³ is fixed to 5, allowing for a potential mutation of the crisis.

It is important to distinguish between a significant IRF and a significant shift from a calm to a crisis period. First, IRFs are demeaned, so that they are significant if the corresponding confidence interval does not include the value of 0. Second, the shift probability from calm to crisis or the probability of remaining in a crisis period is significantly different from zero at time t if the confidence interval associated to the IRF_t remains in the grey area, *i.e.* the centered IRF_t is significantly lying above its unconditional mean ($\hat{\alpha} + \bar{x}_{m,t-1}\hat{\beta}_m$). While the first analysis is common to all vector autoregressive (VAR) models, the second one is specific to non-linear (threshold) time series models.

Figures 5.2 to 5.4 report the diffusion of banking, currency and debt crises respectively through the system.

²³Results for shocks of magnitude 10 are available upon request.

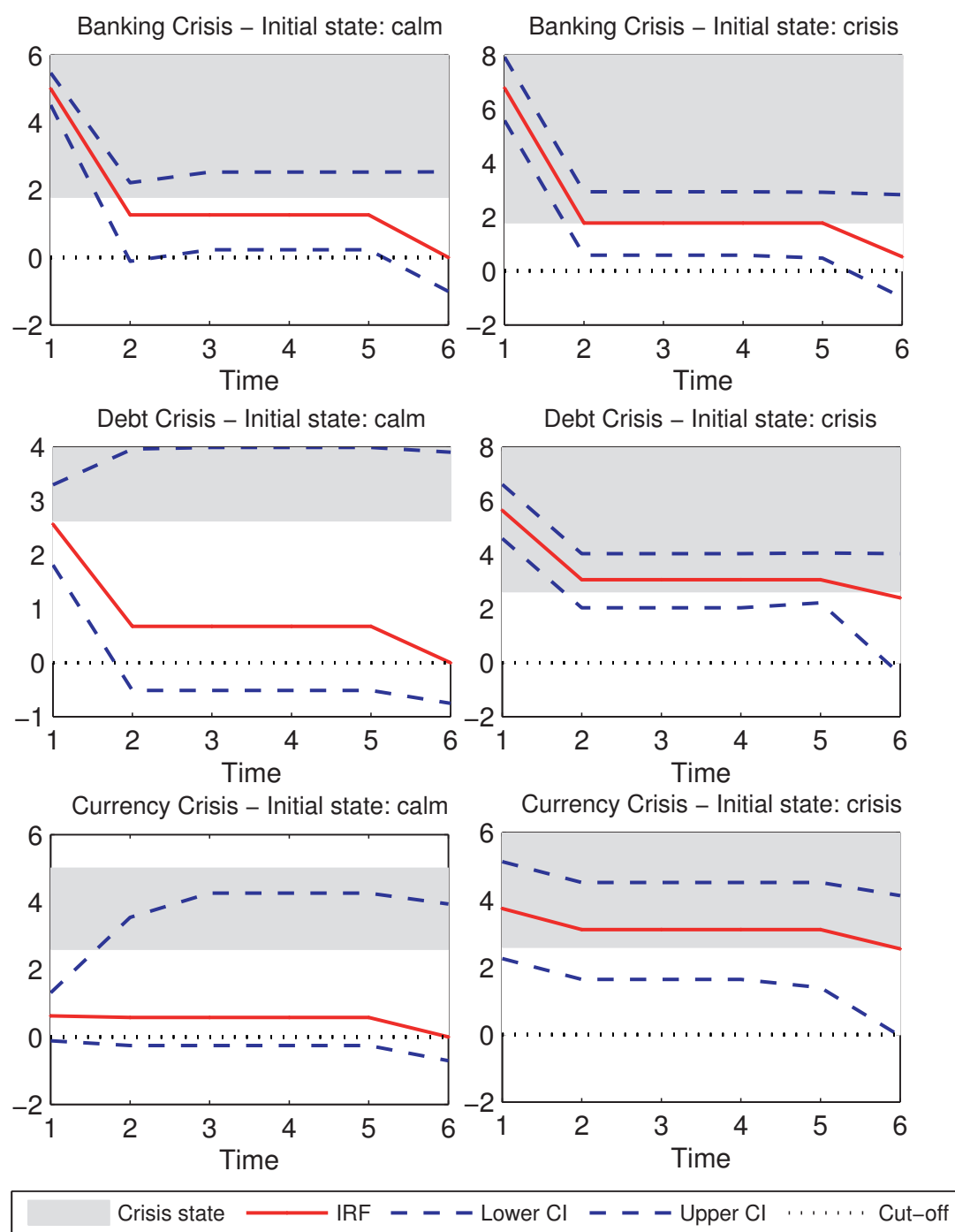


Figure 5.2: IRF after a banking crisis shock - Ecuador 3 months

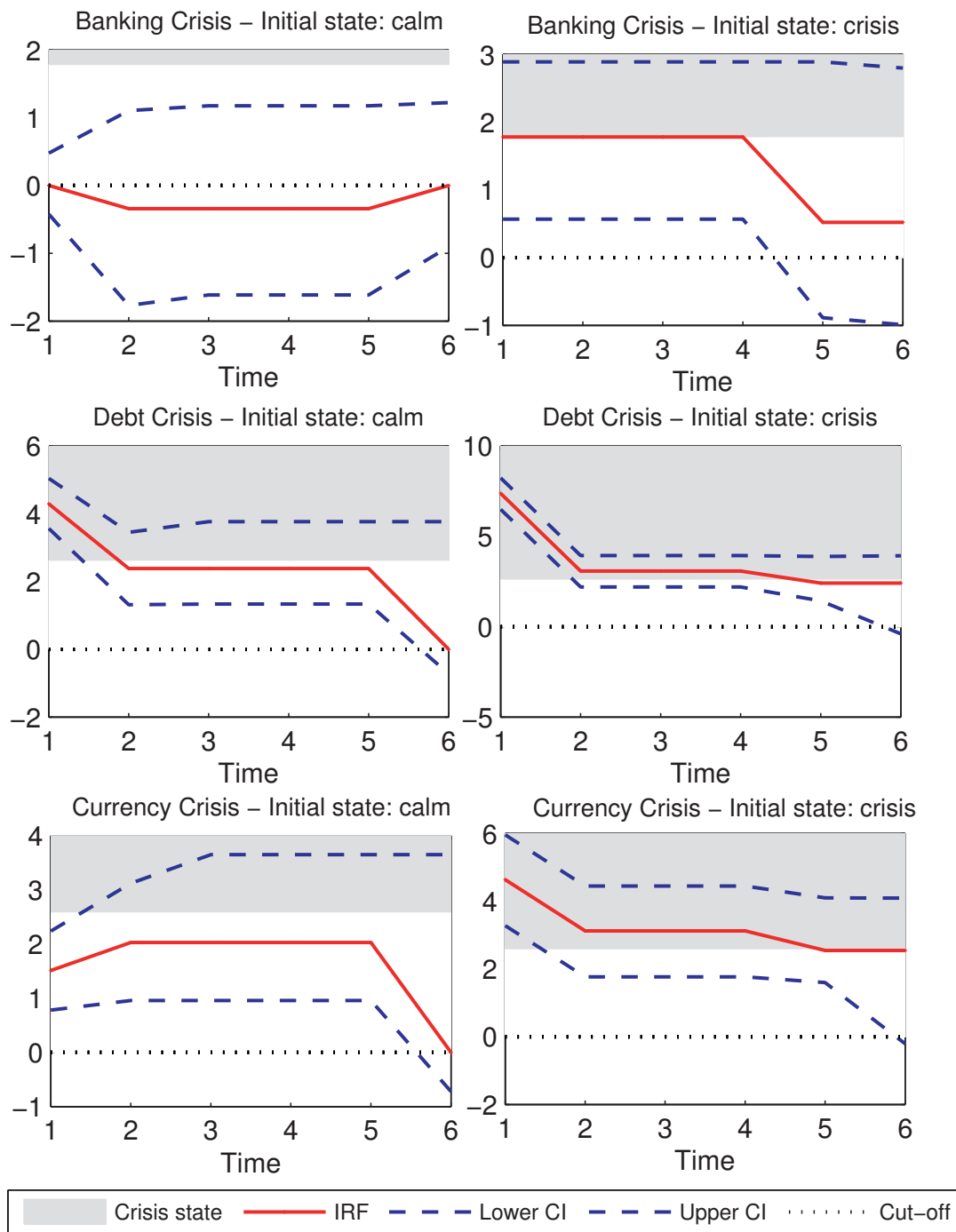


Figure 5.3: IRF after a debt crisis shock - Ecuador 3 months

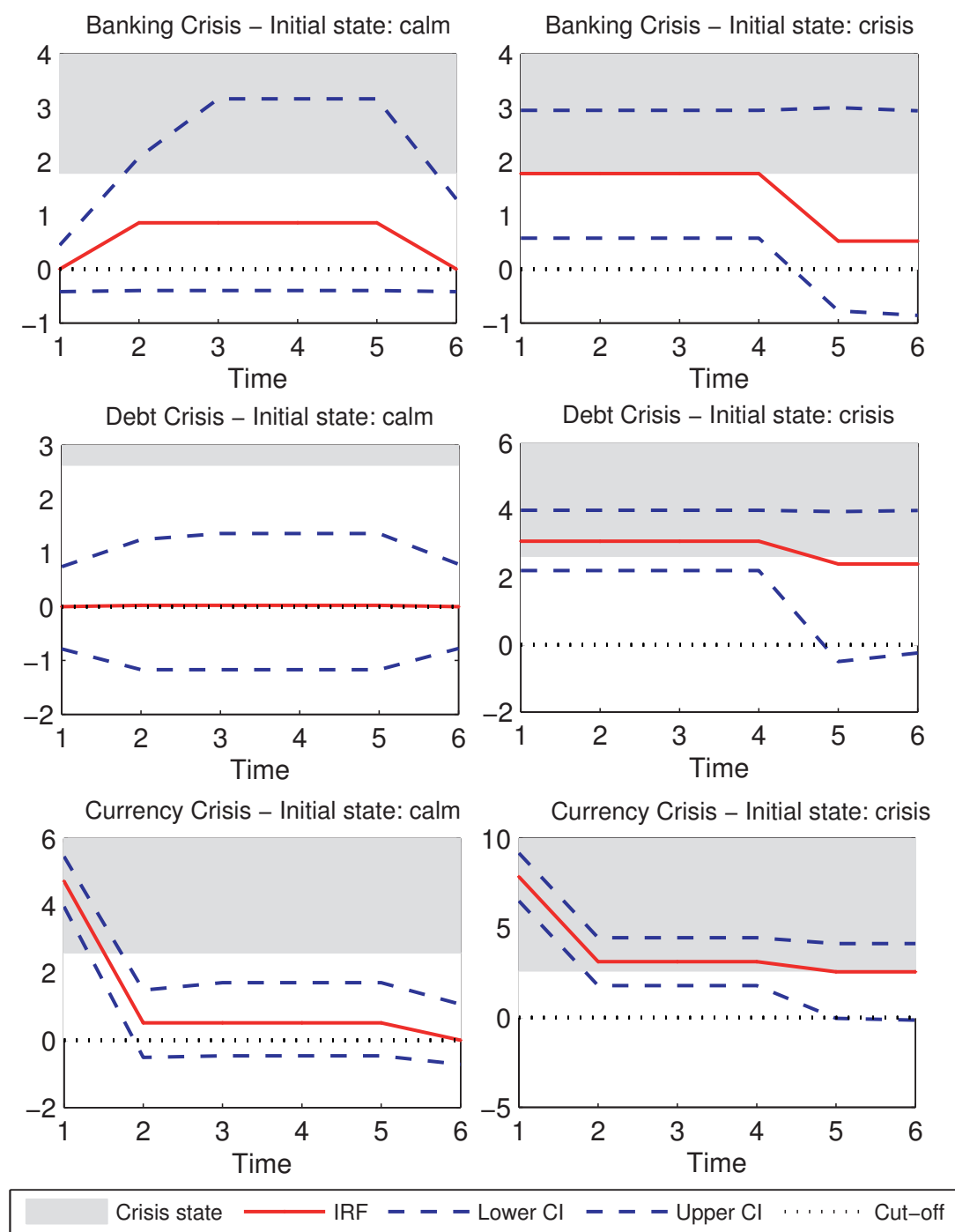


Figure 5.4: IRF after a currency crisis shock - Ecuador 3 months

First, it appears in Figure 5.2 that banking crisis shock has almost no persistence in a calm initial state, as the IRF function reverses to mean after a single period. On the contrary, the persistence jumps to 5 months for an initial crisis state. Similarly, the diffusion of a banking crisis shock to another type of turmoil is exclusively observed in a crisis initial state. Besides, the shift probability from calm to crisis period is significant only for the banking crisis and up to the second period (see the left part of Figure 5.2), whereas the probability of remaining in a crisis period is significant for all three types of crises until $t = 2$ (see the right part of Figure 5.2). This underlines the uncertainty surrounding the duration of a crisis beyond one month after the shock. Overall, these first results clearly correspond to the path exhibited by the crisis sequence faced by Ecuador in the late 90's. Figure 5.3 reports the response of the three latent variables to a debt crisis shock. In such a case, the impact of the shock on the banking and currency crises vanishes almost instantaneously in the case of a calm initial state, while it disappears after 4 or 5 months, if the economy is facing initially a joint crisis. As for the debt crisis, the impact of the shock lasts at least 5 months even though we are certain of being in a crisis period only during the first two periods (the confidence interval is in the grey area at that time). Finally, Figure 5.4 presents the IRF after currency crisis shock. As in the previous cases, the impact on the banking crisis is not important if we depart from a calm situation, while it becomes significant during 4 periods for an initial crisis period. At the same time, the response of the debt crisis is slowly dampened towards the baseline for a calm initial state, whereas it is significant during the first 4 periods if the shock occurs while being in a crisis state. It seems that the persistence of the effect of this shock is around two months for a calm initial period while it dies away only after 5 months in the alternative situation.

Overall, the conditional probability and the IRF analyses stress the superiority of the trivariate model to scrutinize the diffusion mechanisms that occurred in Ecuador after the banking crisis in 1998. Strong interactions between the three types of crises are clearly present in particular between banking and other crises.

5.5 Conclusion

This paper is the first attempt to simultaneously model the three types of crises (currency, banking and sovereign debt), thus allowing to investigate the potential mutations between not only the currency and the banking crises but also the sovereign debt one. It is actually an extension of previous papers which investigate the twin crises phenomenon (in particular Glick and Hutchinson, 1999). To achieve this objective, a methodological novelty has been introduced consisting in an exact maximum likelihood approach to estimate the multivariate dynamic probit model, extending hence the Huguenin, Pelgrin

and Holly (2009)'s method to dynamic models. Applied to a large sample of emerging countries, we find that in the bivariate case causality from banking to currency (and vice-versa) are quite common. More importantly, for the two countries, Ecuador and South Africa, which underwent the 3 types of crises, the trivariate model turns out to be the best performing one in term of conditional probabilities and comprehension of the reasons why a specific crisis mutates to another one: this can be due to common shocks (as in South Africa) or to a strong causal structure (as in Ecuador). More generally, this paper advocates the use of trivariate probit crisis models whenever it is possible, so as to have a better insight on the financial turmoils. Finally, the work in this paper can be extended in several directions. First, a panel data approach could be adopted to jointly estimate the models for a set of countries, thereby appropriately accounting for heterogeneity across countries and imposing parameter heterogeneity wherever it is supported by the data. Provided the specifications across countries are sufficiently homogenous, it might be possible to also include countries which have experienced some but not all types of crises into the joint analysis and to make probability statements on the occurrence in the future of as yet unobserved crises types for these countries.

Second, a comparison of the multivariate probit models with alternative models such as multivariate extensions of the Markov switching model of Hamilton (1989) might give further insights into the dynamics of the generation and transmission of crises.

Third, it would be interesting to investigate the interaction between the method of construction of the crises indicator variables and the nature of the DGP for the indicator variables from which they have been derived, along lines proposed by Harding and Pagan (2011). Fourth, another extension could be to extend the model by specifying the process for the explanatory variables, for instance along the lines of Dueker (2005) who considers a univariate probit, whereby the underlying latent variable and the set of its explanatory variables are generated by a VAR-model. Fifth, an extension should deal with the forecasting properties of the proposed models to find out whether they provide an accurate early warning against an imminent crisis.

5.6 Appendix

5.6.1 Appendix: Proof of lemma 1

By definition, the likelihood of observation t is given by:

$$\begin{aligned}
 L_t(y_t|z_{t-1}, \theta; \Omega) &= \Pr((-q_{1,t}y_{1,t}^* \leq 0), \dots, (-q_{M,t}y_{M,t}^* \leq 0)) \\
 &= \Pr(-q_{1,t}\varepsilon_{1,t} \leq q_{1,t}\pi_{1,t}, \dots, -q_{M,t}\pi_{M,t} \leq q_{M,t}\pi_{M,t}) \\
 &= \Phi_{M,-Q_t\varepsilon_t}(w_t|0_M; \Omega) \\
 &= \int_{-\infty}^{w_{M,t}} \dots \int_{-\infty}^{w_{1,t}} \phi_{M,-Q_t\varepsilon_t}(Q_t\varepsilon_t, \Omega) \prod_{m=1}^M d\varepsilon_{m,t}.
 \end{aligned}$$

Since each $q_{m,t}$ takes only the values $\{-1, 1\}$, it is straightforward to show that $Q_t = Q_t^{-1}$ and $|Q_t\Omega Q_t| = |\Omega|$. Moreover, the density of an M -variate standardized normal vector $-Q_t\varepsilon_t$ with covariance matrix Ω may be re-written as the density of an M -variate standardized normal vector ε_t with covariance matrix $Q_t\Omega Q_t$:

$$\begin{aligned}
 \phi_{M,-Q_t\varepsilon_t}(Q_t\varepsilon_t; \Omega) &= |2\pi\Omega|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} (-Q_t\varepsilon_t)' \Omega^{-1} (-Q_t\varepsilon_t) \right\} \\
 &= |2\pi(Q_t\Omega Q_t)|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \varepsilon_t' (Q_t\Omega Q_t)^{-1} \varepsilon_t \right\} \\
 &= \phi_{M,\varepsilon_t}(\varepsilon_t; Q_t\Omega Q_t).
 \end{aligned}$$

Therefore, the likelihood of observation t is given by:

$$\begin{aligned}
 L_t(y_t|Z_{t-1}, \theta; \Omega) &= \int_{-\infty}^{q_{M,t}\pi_{M,t}} \dots \int_{-\infty}^{q_{1,t}\pi_{1,t}} \phi_{M,\varepsilon_t}(\varepsilon_t; Q_t\Omega Q_t) \prod_{m=1}^M d\varepsilon_{m,t} \\
 &= \Phi_{M,\varepsilon_t}(Q_t\pi_t; Q_t\Omega Q_t).
 \end{aligned}$$

5.6.2 Appendix: The Gauss-Legendre Quadrature rule

The goal of the Gauss-Legendre Quadrature rule is to provide an approximation of the following integral:

$$\int_a^b f(x) dx. \tag{5.24}$$

In a first step, the bounds of the integral must be changed from $[a, b]$ to $[-1, 1]$ before applying the Gaussian Quadrature rule:

$$\int_a^b f(x) dx = \frac{b-a}{2} \int_{-1}^1 f(z) dz, \tag{5.25}$$

where $z_i = \frac{b-a}{2}abs_i + \frac{b+a}{2}$ and the nodes abs_i , $i \in \{1, 2, \dots, p\}$ are zeros of the Legendre polynomial $P_p(abs)$.

Definition 3. Then, the standard p -point Gauss-Legendre quadrature rule over a bounded arbitrary interval $[a, b]$ is given by the following approximation:

$$\int_a^b f(x)dx \approx \frac{b-a}{2} \sum_{i=1}^p v_i f(z_i) + R_p, \quad (5.26)$$

where v_i are the corresponding weights, $v_i = \frac{2}{(1-abs_i^2) \left(\frac{\partial P_p(abs)}{\partial abs} \Big|_{abs_i} \right)^2}$, $\sum_{i=1}^p v_i = 2$, and R_p is the error term, $R_p = Q_p f^{(2p)}(\xi) = \frac{(b-a)^{2p+1} (p!)^4}{(2p+1)(2p!)^3} f^{2p}(\xi)$, with $\xi \in (a, b)$.

5.6.3 Appendix: The EML score vector for a trivariate dynamic probit model

For ease of notation, let us denote by $\rho_{i,j}$, $i, j = \{1, 2, 3\}$, $i \neq j$ the correlation coefficients associated to the Ω matrix. The likelihood of observation t may be written as:

$$\begin{aligned} P_t &= \Phi_3(q_1\pi_{1,t}, q_2\pi_{2,t}, q_3\pi_{3,t}, q_1q_2\rho_{12}, q_1q_3\rho_{13}, q_2q_3\rho_{23}) \\ &= \Phi(q_1\pi_{1,t})\Phi(q_2\pi_{2,t})\Phi(q_3\pi_{3,t}) \\ &\quad + q_1q_2\Phi(q_3\pi_{3,t})\Psi_2(\pi_{1,t}, \pi_{2,t}, \rho_{12}) \\ &\quad + q_1q_3\Phi(q_2\pi_{2,t})\Psi_2(\pi_{1,t}, \pi_{3,t}, \rho_{13}) \\ &\quad + q_2q_3\Phi(q_1\pi_{1,t})\Psi_2(\pi_{2,t}, \pi_{3,t}, \rho_{23}) \\ &\quad + q_1q_2q_3\Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) \\ &\quad + q_1q_2q_3\Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) \\ &\quad + q_1q_2q_3\Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}), \end{aligned} \quad (5.27)$$

where

$$\Psi_2(\pi_{1,t}, \pi_{2,t}, \rho_{12}) = \int_0^{\rho_{12}} \psi_2(\pi_{1,t}, \pi_{2,t}, \lambda_{12}) d\lambda_{12}$$

$$\Psi_2(\pi_{1,t}, \pi_{3,t}, \rho_{13}) = \int_0^{\rho_{13}} \psi_2(\pi_{1,t}, \pi_{3,t}, \lambda_{13}) d\lambda_{13}$$

$$\Psi_2(\pi_{2,t}, \pi_{3,t}, \rho_{23}) = \int_0^{\rho_{23}} \psi_2(\pi_{2,t}, \pi_{3,t}, \lambda_{23}) d\lambda_{23},$$

and

$$\Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) = \int_0^{\rho_{13}} \int_0^{\rho_{23}} \frac{-\pi_{3,t} + \lambda_{13}\pi_{1,t} + \lambda_{23}\pi_{2,t}}{1 - \lambda_{13}^2 - \lambda_{23}^2} \psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \lambda_{13}, \lambda_{23}, 0) d\lambda_{13} d\lambda_{23}$$

$$\Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) = \int_0^{\rho_{23}} \int_0^{\rho_{12}} \frac{-\pi_{2,t} + \lambda_{23}\pi_{3,t} + \lambda_{12}\pi_{1,t}}{1 - \lambda_{23}^2 - \lambda_{12}^2} \psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \lambda_{23}, \lambda_{12}, 0) d\lambda_{23} d\lambda_{12}$$

$$\begin{aligned} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}) = & \int_0^{\rho_{12}} \int_0^{\rho_{13}} \frac{-(1 - \rho_{23}^2)\pi_{1,t} + (\lambda_{12} - \lambda_{13}\rho_{23})\pi_{2,t} + (\lambda_{13} - \lambda_{12}\rho_{23})\pi_{3,t}}{1 - \lambda_{12}^2 - \lambda_{13}^2 - \rho_{23}^2 + 2\lambda_{12}\lambda_{13}\rho_{23}} \\ & \times \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \lambda_{13}, \rho_{23}) d\lambda_{12} d\lambda_{23}. \end{aligned}$$

Therefore, the first order partial derivatives can be obtained as follows:

$$\begin{aligned} \frac{\partial}{\partial \pi_1} P_t = & q_1 \psi(\pi_{1,t}) \Phi(q_2 \pi_{2,t}) \Phi(q_3 \pi_{3,t}) \\ & + q_1 q_2 \Phi(q_3 \pi_{3,t}) \frac{\partial}{\partial \pi_1} \Psi_2(\pi_{1,t}, \pi_{2,t}, \rho_{12}) \\ & + q_1 q_3 \Phi(q_2 \pi_{2,t}) \frac{\partial}{\partial \pi_1} \Psi_2(\pi_{1,t}, \pi_{3,t}, \rho_{13}) \\ & + q_1 q_2 q_3 \psi(\pi_{1,t}) \Psi_2(\pi_{2,t}, \pi_{3,t}, \rho_{23}) \\ & + q_1 q_2 q_3 \frac{\partial}{\partial \pi_1} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) \\ & + q_1 q_2 q_3 \frac{\partial}{\partial \pi_1} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) \\ & + q_1 q_2 q_3 \frac{\partial}{\partial \pi_1} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}), \end{aligned} \tag{5.28}$$

$$\begin{aligned}
 \frac{\partial}{\partial \pi_2} P_t &= q_2 \psi(\pi_{2,t}) \Phi(q_1 \pi_{1,t}) \Phi(q_3 \pi_{3,t}) \\
 &\quad + q_1 q_2 \Phi(q_3 \pi_{3,t}) \frac{\partial}{\partial \pi_2} \Psi_2(\pi_{1,t}, \pi_{2,t}, \rho_{12}) \\
 &\quad + q_1 q_2 q_3 \psi(\pi_{2,t}) \Psi_2(\pi_{1,t}, \pi_{3,t}, \rho_{13}) \\
 &\quad + q_2 q_3 \Phi(q_1 \pi_{1,t}) \frac{\partial}{\partial \pi_2} \Psi_2(\pi_{2,t}, \pi_{3,t}, \rho_{23}) \\
 &\quad + q_1 q_2 q_3 \frac{\partial}{\partial \pi_2} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) \\
 &\quad + q_1 q_2 q_3 \frac{\partial}{\partial \pi_2} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) \\
 &\quad + q_1 q_2 q_3 \frac{\partial}{\partial \pi_2} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}),
 \end{aligned} \tag{5.29}$$

$$\begin{aligned}
 \frac{\partial}{\partial \pi_3} P_t &= q_1 \psi(\pi_{3,t}) \Phi(q_1 \pi_{1,t}) \Phi(q_2 \pi_{2,t}) \\
 &\quad + q_1 q_2 q_3 \psi(\pi_{3,t}) \Psi_2(\pi_{1,t}, \pi_{2,t}, \rho_{12}) \\
 &\quad + q_1 q_3 \Phi(q_2 \pi_{2,t}) \frac{\partial}{\partial \pi_3} \Psi_2(\pi_{1,t}, \pi_{3,t}, \rho_{13}) \\
 &\quad + q_2 q_3 \Phi(q_1 \pi_{1,t}) \frac{\partial}{\partial \pi_3} \Psi_2(\pi_{2,t}, \pi_{3,t}, \rho_{23}) \\
 &\quad + q_1 q_2 q_3 \frac{\partial}{\partial \pi_3} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) \\
 &\quad + q_1 q_2 q_3 \frac{\partial}{\partial \pi_3} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) \\
 &\quad + q_1 q_2 q_3 \frac{\partial}{\partial \pi_3} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}),
 \end{aligned} \tag{5.30}$$

$$\begin{aligned}
 \frac{\partial}{\partial \rho_{12}} P_t &= q_1 q_2 \Phi(q_3 \pi_{3,t}) \frac{\partial}{\partial \rho_{12}} \Psi_2(\pi_{1,t}, \pi_{2,t}, \rho_{12}) \\
 &\quad + q_1 q_2 q_3 \frac{\partial}{\partial \rho_{12}} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) \\
 &\quad + q_1 q_2 q_3 \frac{\partial}{\partial \rho_{12}} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}),
 \end{aligned} \tag{5.31}$$

$$\begin{aligned}
\frac{\partial}{\partial \rho_{13}} P_t &= q_1 q_3 \Phi(q_2 \pi_{2,t}) \frac{\partial}{\partial \rho_{13}} \Psi_2(\pi_{1,t}, \pi_{3,t}, \rho_{13}) \\
&+ q_1 q_2 q_3 \frac{\partial}{\partial \rho_{13}} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) \\
&+ q_1 q_2 q_3 \frac{\partial}{\partial \rho_{13}} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}),
\end{aligned} \tag{5.32}$$

$$\begin{aligned}
\frac{\partial}{\partial \rho_{23}} P_t &= q_2 q_3 \Phi(q_1 \pi_{1,t}) \frac{\partial}{\partial \rho_{23}} \Psi_2(\pi_{2,t}, \pi_{3,t}, \rho_{23}) \\
&+ q_1 q_2 q_3 \frac{\partial}{\partial \rho_{23}} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) \\
&+ q_1 q_2 q_3 \frac{\partial}{\partial \rho_{23}} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) \\
&+ q_1 q_2 q_3 \frac{\partial}{\partial \rho_{23}} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}),
\end{aligned} \tag{5.33}$$

where

$$\begin{aligned}
\frac{\partial}{\partial \pi_1} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) &= \int_0^{\rho_{23}} \int_0^{\rho_{13}} \frac{\partial}{\partial \lambda_{13}} \psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \lambda_{13}, \lambda_{23}, 0) d\lambda_{13} \lambda_{23} \\
&= \int_0^{\rho_{23}} \psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \lambda_{23}, 0) d\lambda_{23},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \pi_2} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) &= \int_0^{\rho_{13}} \int_0^{\rho_{23}} \frac{\partial}{\partial \lambda_{23}} \psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \lambda_{13}, \lambda_{23}, 0) d\lambda_{23} \lambda_{13} \\
&= \int_0^{\rho_{13}} \psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \lambda_{13}, \rho_{23}, 0) d\lambda_{13},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \pi_3} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) &= \\
&\int_0^{\rho_{13}} \int_0^{\rho_{23}} [(\pi_{3,t} - \lambda_{13} \pi_{1,t} - \lambda_{23} \pi_{2,t})^2 - (1 - \lambda_{13}^2 - \lambda_{23}^2)] \\
&\times \frac{1}{(1 - \lambda_{13}^2 - \lambda_{23}^2)^2} \psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \lambda_{13}, \lambda_{23}, 0) d\lambda_{13} d\lambda_{23},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \rho_{13}} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) &= \\
&\int_0^{\rho_{23}} \frac{-\pi_{3,t} + \rho_{13} \pi_{3,t} + \lambda_{23} \pi_{2,t}}{1 - \rho_{13}^2 - \lambda_{23}^2} \psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \lambda_{23}, 0) d\lambda_{23},
\end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \rho_{23}} \Psi_3(\pi_{3,t}, \pi_{1,t}, \pi_{2,t}, \rho_{13}, \rho_{23}, 0) = \\ \int_0^{\rho_{13}} \frac{-\pi_{3,t} + \lambda_{13}\pi_{3,t} + \rho_{23}\pi_{2,t}}{1 - \lambda_{13}^2 - \rho_{23}^2} \psi_3(\pi_3, \pi_1, \pi_2, \lambda_{13}, \rho_{23}, 0) d\lambda_{13}, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \pi_1} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) &= \int_0^{\rho_{23}} \int_0^{\rho_{12}} \frac{\partial}{\partial \lambda_{12}} \psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \lambda_{23}, \lambda_{12}, 0) d\lambda_{12} \lambda_{23} \\ &= \int_0^{\rho_{23}} \psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \lambda_{23}, \rho_{12}, 0) d\lambda_{23}, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \pi_2} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) = \\ \int_0^{\rho_{23}} \int_0^{\rho_{12}} [(\pi_{2,t} - \lambda_{23}\pi_{3,t} - \lambda_{12}\pi_{1,t})^2 - (1 - \lambda_{23}^2 - \lambda_{12}^2)] \\ \times \frac{1}{(1 - \lambda_{23}^2 - \lambda_{12}^2)^2} \psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \lambda_{23}, \lambda_{12}, 0) d\lambda_{23} d\lambda_{12}, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \pi_3} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) &= \int_0^{\rho_{12}} \int_0^{\rho_{23}} \frac{\partial}{\partial \lambda_{23}} \psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \lambda_{23}, \lambda_{12}, 0) d\lambda_{23} \lambda_{12} \\ &= \int_0^{\rho_{12}} \psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \lambda_{12}, 0) d\lambda_{12}, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \rho_{12}} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) &= \int_0^{\rho_{23}} \frac{-\pi_{2,t} + \lambda_{23}\pi_{3,t} + \rho_{12}\pi_{1,t}}{1 - \lambda_{23}^2 - \rho_{12}^2} \times \\ &\quad \psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \lambda_{23}, \rho_{12}, 0) d\lambda_{23}, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \rho_{23}} \Psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \rho_{12}, 0) &= \int_0^{\rho_{12}} \frac{-\pi_{2,t} + \rho_{23}\pi_{3,t} + \lambda_{12}\pi_{1,t}}{1 - \rho_{23}^2 - \lambda_{12}^2} \times \\ &\quad \psi_3(\pi_{2,t}, \pi_{3,t}, \pi_{1,t}, \rho_{23}, \lambda_{12}, 0) d\lambda_{12}, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \pi_1} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}) &= \int_0^{\rho_{12}} \int_0^{\rho_{13}} \{[(1 - \rho_{23}^2)\pi_{1,t} - (\lambda_{12} - \lambda_{13}\rho_{23})\pi_{2,t} \\ &\quad - (\lambda_{13} - \lambda_{12}\lambda_{23})\pi_{3,t}]^2 \\ &\quad - (1 - \rho_{23}^2)(1 - \lambda_{12}^2 - \lambda_{13}^2 - \rho_{23}^2 + 2\lambda_{12}\lambda_{13}\rho_{23})\} \times \\ &\quad \frac{1}{(1 - \lambda_{12}^2 - \lambda_{13}^2 - \rho_{23}^2 + 2\lambda_{12}\lambda_{13}\rho_{23})^2} \times \\ &\quad \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \lambda_{13}, \rho_{23}) d\lambda_{12} d\lambda_{13}, \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \pi_2} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}) &= \\
&\int_0^{\rho_{13}} \int_0^{\rho_{12}} \frac{\partial}{\partial \lambda_{12}} \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \lambda_{13}, \rho_{23}) d\lambda_{12} \lambda_{13} \\
&= \int_0^{\rho_{13}} \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \lambda_{13}, \rho_{23}) d\lambda_{13},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \pi_3} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}) &= \\
&\int_0^{\rho_{12}} \int_0^{\rho_{13}} \frac{\partial}{\partial \lambda_{13}} \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \lambda_{13}, \rho_{23}) d\lambda_{13} \lambda_{12} \\
&= \int_0^{\rho_{12}} \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \rho_{13}, \rho_{23}) d\lambda_{12},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \rho_{12}} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}) &= \\
&\int_0^{\rho_{13}} \frac{(1 - \rho_{23}^2)\pi_{1,t} + (\rho_{12} - \lambda_{13}\rho_{23})\pi_{2,t} + (\lambda_{13} - \rho_{12}\rho_{23})\pi_{3,t}}{1 - \rho_{12}^2 - \lambda_{13}^2 - \rho_{23}^2 + 2\rho_{12}\lambda_{13}\rho_{23}} \\
&\times \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \lambda_{13}, \rho_{23}) d\lambda_{13},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \rho_{13}} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}) &= \\
&\int_0^{\rho_{12}} \frac{(1 - \rho_{23}^2)\pi_{1,t} + (\lambda_{12} - \rho_{13}\rho_{23})\pi_{2,t} + (\rho_{13} - \lambda_{12}\rho_{23})\pi_{3,t}}{1 - \lambda_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\lambda_{12}\rho_{13}\rho_{23}} \\
&\times \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \rho_{13}, \rho_{23}) d\lambda_{12},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \rho_{23}} \Psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \rho_{13}, \rho_{23}) &= \\
&\int_0^{\rho_{12}} \int_0^{\rho_{13}} \frac{\partial^2}{\partial \pi_{2,t} \partial \lambda_{13}} \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \lambda_{13}, \rho_{23}) d\lambda_{12} d\lambda_{13} \\
&= \int_0^{\rho_{12}} \frac{-(1 - \rho_{13}^2)\pi_{2,t} + (\lambda_{12} - \rho_{13}\rho_{23})\pi_{1,t} + (\rho_{23} - \lambda_{12}\rho_{13})\pi_{3,t}}{1 - \lambda_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\lambda_{12}\rho_{13}\rho_{23}} \\
&\times \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \rho_{13}, \rho_{23}) d\lambda_{12} \\
&= \int_0^{\rho_{12}} \int_0^{\rho_{13}} \frac{\partial^2}{\partial \pi_{3,t} \partial \lambda_{12}} \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \lambda_{12}, \lambda_{13}, \rho_{23}) d\lambda_{12} d\lambda_{13} \\
&= \int_0^{\rho_{13}} \frac{-(1 - \rho_{12}^2)\pi_{3,t} + (\lambda_{13} - \rho_{12}\rho_{23})\pi_{1,t} + (\rho_{23} - \rho_{12}\lambda_{13})\pi_{2,t}}{1 - \rho_{12}^2 - \lambda_{13}^2 - \rho_{23}^2 + 2\rho_{12}\lambda_{13}\rho_{23}} \\
&\times \psi_3(\pi_{1,t}, \pi_{2,t}, \pi_{3,t}, \rho_{12}, \lambda_{13}, \rho_{23}) d\lambda_{13}.
\end{aligned}$$

Chapter 6

Conclusions

This dissertation proposes four essays which contribute to the financial crises Early Warning Systems literature in several ways. In the wake of the global financial crisis, numerous questions with respect to EWS forecasting abilities have been raised, as very few signals were drawn prior to the starting of the turmoil. Two research topics are relevant to this literature, namely the specification of an EWS and the evaluation of these models. Most of the papers in the literature focus on the first issue. They scrutinize the causes of the crises, *i.e.* the leading indicators considered in EWS, and/or propose new methodologies to approach the link between the crisis indicator and the macroeconomic and financial variables. By contrast, the literature on EWS evaluation is scarce. To our knowledge, no formal evaluation methodology for the forecasting abilities of an EWS has been proposed hitherto. This dissertation contributes to both research directions.

The main objectives of this thesis are hence (i) to propose forecast evaluation methods for non-linear models (in particular, a unified evaluation methodology for EWS), as well as (ii) to introduce methodological novelties based on recent econometric developments in the specification of EWS, hence improving their forecasting abilities. Chapter 2 and chapter 3 scrutinize the first objective, while chapter 4 and chapter 5 explore the second one.

EWS evaluation is vital for crises prevention. Numerous EWS specifications have been proposed in the literature and various conclusions were drawn with respect to the importance of certain leading indicators for different crises. An accurate comparison of the relative forecasting abilities of these models and a clear analysis of their absolute forecasting performance appears hence compulsory. In chapters 2 and 3 we fill this gap in the literature by proposing two forecast validation methodologies: one designed for interval forecasts, and a toolbox evaluating any model yielding probabilities to observe an event (crisis, recession, etc.).

In chapter 2 we propose a very general evaluation test for interval forecasts and High Density Regions issued from any type of model. Our model free test based on the GMM approach proposed by Bontemps (2006) and Bontemps and Meddahi (2005, 2011) uses simple J-statistics base on particular moments defined by orthonormal polynomials associated with the binomial distribution. It relies on an original approach that transforms the violation series into a series of sums of violations defined for H blocks of size N . Our GMM-test performs well in finite samples of relatively small size since the finite-sample distribution of the J -statistic is close to its chi-squared distribution regardless of the block size chosen. It allows us to independently test the three hypotheses of conditional coverage, independence and unconditional coverage. Most importantly, it has very good power properties for realistic sample sizes.

Chapter 3 introduces an original, two-step model-free evaluation methodology specifically designed for EWS models for any type of crises. First, an original method to identify the cut-off, *i.e.* the threshold best discriminating between crisis and calm periods is proposed. Second, we introduce several comparison tests to assess the relative forecasting abilities of EWS. We show that an adequate EWS evaluation should take into account the cut-off both in the optimal crisis forecast step (by relying on the Credit-Scoring or Accuracy-Measures optimal cut-off) and in the model comparison step (by using the Area Under the ROC Curve test). Our methods to identify this cut-off perform better than existing ones, identifying on average more than 2/3 of the crisis and calm periods, in contrast with the NSR one, that correctly forecasts all the calm periods at the expense of most of the crisis ones. Besides, we argue that the use of statistical inference is compulsory to accurately grasp the importance of a leading indicator in an EWS. Indeed, we are able to show that the yield spread is an important indicator of currency crises only for South-Asian countries, and not for all the countries as it seems at a first glance.

To improve the predictive abilities of EWS models, different characteristics of financial crises not scrutinized before in the literature must be taken into account (*e.g.* crises persistence and possibility to spill-over to other markets). We thus contribute to the literature on EWS specification by proposing new methodologies, which incorporate information with respect to the regime prevailing in the previous period and/or on other markets, *i.e.* crisis dynamics and mutation. These improvements of existing binary EWS are tackled in chapters 4 and 5.

Chapter 4 shows that crisis dynamics should be taken into account when forecasting such events. Based on an exact maximum-likelihood method (proposed by Kauppi and Saikonen, 2008) we estimate several dynamic binary-choice specifications. Subsequently, we adapt this methodology to panel by drawing on the works of Carro (2007). We hence show that in the case of currency crises dynamic models perform better than

Markov-switching and static logit models both in-sample and out-of-sample. Besides, dynamic EWS deliver good forecasting probabilities, beyond the optimal cut-off in the crisis periods and underneath this threshold the rest of the time. The crises persistence hence seems to be a key indicator in financial crises analysis.

In chapter 5 a multivariate dynamic probit EWS is proposed. It is an extension of the multivariate model of Huguenin et al. (2009) that relies on the evaluation of higher-order integrals and on the use of quadrature-rules. Encompassing the 3 main types of financial crises, i.e. banking, currency and sovereign debt, it allows us to investigate the potential causality between them: this can be due to common shocks (as in South Africa) or to a strong causal structure (as in Ecuador). The trivariate model outperforms the bivariate one for countries that underwent the three types of crises, supporting the implementation of such models whenever it is feasible.

Finally, I briefly present some promising avenues of research to extend present work. Since EWS are generally non-linear models, more adapted forecasting methods (*e.g.* yielding interval-forecasts or HDR) could be considered instead of simple point forecasts. The test proposed in chapter 2 could then be implemented to evaluate the validity of these forecasts.

Second, the forecasting properties of the proposed EWS models could be evaluated with respect to the recent crisis, so as to find out whether they provide an accurate early warning against an imminent crisis. Along these lines, a comparison of the multivariate probit models with alternative models such as multivariate extensions of the Markov switching model of Hamilton (1989) might give further insights into the dynamics of the generation and transmission of crises.

Third, the multivariate dynamic model could be extended by specifying the process for the explanatory variables (*e.g.* along the lines of Dueker, 2005, who considers a univariate probit whereby the underlying latent variable and the set of its explanatory variables are generated by a VAR-model). Besides, a panel data approach could be adopted to jointly estimate the multivariate dynamic models for a set of countries, while accounting for heterogeneity across countries.

References

- [1] Abiad, A., 2003, Early Warning Systems: A Survey and a Regime Switching Approach, IMF Working Paper 32, International Monetary Fund, Washington.
- [2] Alessi, L., Detcken, C., 2011, 'Real Time' Early Warning Indicators for Costly Asset Price Boom/Bust cycles: a Role for Global Liquidity, ECB Working Paper 1039.
- [3] Arias, G., and Erlandsson, G., 2005, Improving Early Warning Systems with a Markov Switching Model - an Application to South-East Asian Crises, Lund University, Working Paper.
- [4] Bai, J., and Ng, S., 2001, A New Look at Panel Testing of Stationarity and the PPP Hypothesis, Boston College Working Papers in Economics 518, Boston College Department of Economics.
- [5] Bao, Y., Lee, T.H., and Saltoglu, B., 2004, A Test for Density Forecast Comparison with Applications to Risk Management, University of California, Riverside.
- [6] Bao, Y., Lee, T.H., and Saltoglu, B., 2007, Comparing Density Forecast Models, Journal of Forecasting 26, 203-225.
- [7] Basel Committee on Banking Supervision, 2005, Studies on the Validation of Internal Rating Systems, Working Paper no. 14, Bank for International Settlements.
- [8] Berg, A., and Pattillo, C., 1999, Predicting Currency Crises: The Indicators Approach and an Alternative, Journal of International Money and Finance 18, 561-586.
- [9] Berg, J.B., Candelon, B., and Urbain, J.P., 2008, A Cautious Note on the Use of Panel Models to Predict Financial Crises, Economics Letters 101, issue 1, 80-83.
- [10] Berg, A., and Cooke, R., 2004, Autocorrelation Corrected Standard Errors in Panel Probits: an Application to Currency Crisis Prediction, IMF Working Paper 39, International Monetary Fund, Washington.
- [11] Berkowitz, J., Christoffersen, P., and Pelletier, D., 2010, Evaluating Value-at-Risk Models with Desk-Level Data, forthcoming in Management Science.

- [12] Bontemps, C., 2006, Moment-Based Tests for Discrete Distributions, Working Paper Toulouse School of Economics.
- [13] Bontemps, C., and Meddahi, N., 2005, Testing Normality: a GMM Approach, *Journal of Econometrics* 124, 149-186.
- [14] Bontemps, C., and Meddahi, N., 2011, Testing Distributional Assumptions: A GMM Approach, *Journal of Applied Econometrics*, forthcoming.
- [15] Bordo, M.D., Eichengreen, B., Klingebiel, D., and Martinez-Peria, M.S., 2001a, Is the Crisis Problem Growing more Severe?, *Economic Policy* 32, 51-82.
- [16] Bordo, M.D., Eichengreen, B., Klingebiel, D., and Martinez-Peria, M.S., 2001b, Financial Crises: Lessons from the Last 120 Years, *Economic Policy*.
- [17] Bruinshoofd, A., Candelon, B., and Raabe, K., 2010, Banking Sector Fragility and the Transmission of Currency Crises, *Open Economies Review* 21, issue 2, 263-292.
- [18] Bussière, M., and Fratzscher, M., 2006, Towards a New Early Warning System of Financial Crises, *Journal of International Money and Finance* 25, issue 6, 953-973.
- [19] Bussière, M., 2007, Balance of Payment Crises in Emerging Markets. How Early Were the "Early" Warning Signals?, ECB Working Paper 713.
- [20] Candelon, B., Dumitrescu, E.I., and Hurlin, C., 2009, Towards an Unified Framework to Evaluate Financial Crises Early Warning Systems, METEOR Research memorandum RM/10/046.
- [21] Candelon, B., Dumitrescu, E.I., and Hurlin, C., 2012, How to Evaluate an Early Warning System? Towards a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods, *IMF Economic Review* 60, issue 1.
- [22] Candelon, B., and Palm, F., 2010, Banking and Debt Crises in Europe: the Dangerous Liaisons?, *De Economist* 158, issue 1, 81-99.
- [23] Candelon, B., Dumitrescu, E.I., and Hurlin, C., 2010, Currency Crises Early Warning Systems: Why They Should Be Dynamic, METEOR Research memorandum RM/10/047.
- [24] Candelon, B., Dumitrescu, E.I., and Hurlin, C., and Palm, F., 2011, Modeling Financial Crisis Mutation, DR LEO 2011-17.
- [25] Candelon, B., Colletaz, G., and Hurlin, C., and Tokpavi, S., 2011, Backtesting Value-at-Risk: a GMM Duration-Based Test, *Journal of Financial Econometrics* 9, issue 2, 314-343.

-
- [26] Caprio, G. and Klingebiel, D., 2006, Bank Insolvencies: Cross Country Experience, World Bank Publications 1620.
- [27] Carro, J.M., 2007, Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects, *Journal of Econometrics* 140, 503-528.
- [28] Ciarlone, A., and Trebeschi, G., 2005, Designing an Early Warning System for Debt Crises, *Emerging Markets Review* 6, issue 4, 376-395.
- [29] Chatfield, C., 1993, Calculating Interval Forecasts, *Journal of Business and Economic Statistics* 11, issue 2, 121-135.
- [30] Chib, S., and Greenberg, E., 1998, Analysis of Multivariate Probit Models, *Biometrika* 85, issue 2, 37-361.
- [31] Christoffersen, F.P., 1998, Evaluating Interval Forecasts, *International Economic Review* 39, 841-862.
- [32] Clark, T.E., and McCracken, M.W., 2001, Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics* 105, issue 1, 85-110.
- [33] Clark, T.E., and McCracken, M.W., 2011, Advances in Forecast Evaluation, Working Papers 2011-025, Federal Reserve Bank of St. Louis.
- [34] Clark, T.E., and West, K.D., 2007, Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics* 138, issue 1, 291-311.
- [35] Clements, M.P., and Taylor, N., 2002, Evaluating Interval Forecasts of High-Frequency Financial Data, *Journal of Applied Econometrics* 18, issue 4, 445-456.
- [36] Colletaz, G., and Hurlin, C., 2005, Modèles Non-Linéaires et Prévision, Rapport Institut CDC pour la recherche.
- [37] Corradi, V., and Swanson, N.R., 2006a, Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification, *Journal of Econometrics* 133, 779-806.
- [38] Corradi, V., and Swanson, N.R., 2006b, Predictive Density Evaluation, in *Handbook of Economic Forecasting*, C.W.J. Granger, G. Elliott and A. Timmermann, editors, Elsevier: Amsterdam, 197-284.
- [39] Corsetti, G., Pesenti, P., and Roubini, N., 1999, What Caused the Asian Currency and Financial Crisis?, *Japan and the World Economy* 11, issue 3, 305-373.
- [40] Dabrowski, M., and Jakubiak, M., 2003, The Sources of Economic Growth in Ukraine after 1998 Currency Crisis and the Country's Prospects, CASE Network Reports 0055, CASE-Center for Social and Economic Research.

- [41] Davis, E.P., and Karim, D., 2008, Could Early Warning Systems Have Helped to Predict the Sub-Prime Crisis? *National Institute Economic Review* 206, issue 1, 35-47.
- [42] DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L., 1988, "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach", *Biometrics* 44, issue 3, 837-845.
- [43] Demirguc-Kunt, A., and Detragiache, E., 1998, The Determinants of Banking Crises in Developed and Developing Countries, *IMF Staff Paper* 45, issue 1, 81-109.
- [44] Detragiache, E., Spilimbergo, A., 2001, Crises and Liquidity: Causes and Interpretation, *IMF Working Paper* 02, International Monetary Fund, Washington.
- [45] Diebold, F. X., and Mariano, S., 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, American Statistical Association 13, issue 3, 253-263.
- [46] Diebold, F.X., and Rudebusch, G.D., 1989, Scoring the Leading Indicators, *The Journal of Business* 62, issue 3, 369-391.
- [47] Dueker, M., 2005, Dynamic Forecasts of Qualitative Variables: A Qual VAR Model of U.S. Recessions, *Journal of Business and Economic Statistics* 23, 96-104.
- [48] Dufour, J.-M., 2006, Monte Carlo Tests with Nuisance Parameter: a General Approach to Finite Sample Inference and Nonstandard Asymptotics, *Journal of Econometrics* 127, issue 2, 443-477.
- [49] Dumitrescu, E.I., Hurlin, C., Madkour, J., 2011, Testing Interval Forecasts: a GMM-Based Approach, forthcoming in *Journal of Forecasting*.
- [50] Eichengreen, B., Rose, A., and Wyplosz, C., 1996, Contagious Currency Crises: First Test, *Journal of Monetary Economics* 52, issue 1, 65-68.
- [51] Eichengreen, B., and Rose, A., 1998, Staying Afloat When the Wind Shifts: External Factors and Emerging Market Banking Crises, *NBER Working Paper* 6370.
- [52] Engelmann, B., Hayden, E., and Tasche, D., 2003, Testing Rating Accuracy?, *Risk* 16, 82-86.
- [53] Estrella, A., and Hardouvelis, G.A., 1991, The Term Structure as a Predictor of Real Economic Activity, *The Journal of Finance* 46, issue 2, 555-576.
- [54] Estrella, A., and Mishkin, F.S., 1996, The Yield Curve as a Predictor of US Recessions, *Current Issues in Economics and Finance* 2, issue 7, 1-5.

-
- [55] Estrella, A., and Trubin, M.R., 2006, The Yield Curve as a Leading Indicator: Some Practical Issues, *Current Issues in Economics and Finance* 12, issue 5, 1-7.
- [56] Falcetti, E., and Tudela, M., 2006, Modelling Currency Crises in Emerging Markets: A Dynamic Probit Model with Unobserved Heterogeneity and Autocorrelated Errors, *Oxford Bulletin of Economics and Statistics* 68, issue 4, 445-471.
- [57] Flood, R., and Garber, P., 1984, Collapsing Exchange Rate Regimes: some Linear Examples, *Journal of International Economics* 17, 1-13.
- [58] Frankel, J., and Saravelos, G., 2010, Are Leading Indicators of Financial Crises Useful for Assessing Country Vulnerability? Evidence from the 2008-09 Global Crisis, NBER Working Paper 16047.
- [59] Fratzscher, M., 2003, On Currency Crises and Contagion, *International Journal of Finance and Economics* 8, issue 2, 109-129.
- [60] Fuertes, A.-M., and Kalotychou, E., 2007, Optimal Design of Early Warning Systems for Sovereign Debt Crises, *International Journal of Forecasting* 23, issue 1, 85-100.
- [61] Gallant, A.R., 1987, *Nonlinear Statistical Models*, John Wiley and Sons, New York.
- [62] Glick, R. and Hutchison, M., 1999, Banking and Currency Crises: How Common Are The Twins?, Working Papers 012000, Hong Kong Institute for Monetary Research.
- [63] Greene, W.H., 2002, *Econometric Analysis*, 5th ed., Prentice Hall, New Jersey.
- [64] Gould, W., Pitblado, J., Sribney, W., 2005, *Maximum Likelihood Estimation with Stata*, Stata Press.
- [65] Hagen, T.K., and Ho, J., 2004, Money Market Pressure and the Determinants of Banking Crises, CEPR 4651.
- [66] Hamilton, J.D., 1989, A New Approach of the Economic Analysis of Non-Stationary Time Series and the Business Cycle, *Econometrica* 57, 357-384.
- [67] Hansen, L.P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, 1029-1054.
- [68] Harding, D., and Pagan, A., 2006, The Econometric Analysis of Constructed Binary Time Series, Working Papers Series 963, The University of Melbourne.
- [69] Harding, D., and Pagan, A., 2009, An Econometric Analysis of Some Models for Constructed Binary Time Series, NCER Working Paper 39.

- [70] Harding, D., and Pagan A., 2011, An Econometric Analysis of Some Models of Constructed Binary Random Variables, *Journal of Business and Economic Statistics* 29, 86-95.
- [71] Harvey, D. I., and Leybourne, S. J., 2007, Testing for Time Series Linearity, *Econometrics Journal* 10, 149-165.
- [72] Hoeffding, W., 1948, A Class of Statistics with Asymptotically Normal Distributions, *Annals of Statistics* 19, 293-325.
- [73] Hoggarth, G., Reis, R., and Saporta, V., 2002, Costs of Banking System Instability: Some Empirical Evidence, *Journal of Banking and Finance* 26, 825-855.
- [74] Huguenin, J., Pelgrin, F., and Holly, A., 2009, Estimation of Multivariate Probit Models by Exact Maximum Likelihood, Working Paper 09-02.
- [75] Hyndman, R.J., 1995, Highest-Density Forecast Regions for Non-linear and Non-normal Time-Series Models, *Journal of Forecasting* 14, 431-441.
- [76] Im, K.S., Pesaran, M.H., and Shin, Y., 2003, Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics* 115, issue 1, 53-74.
- [77] Jacobs, J.P.A.M., Kuper, G.H., and Lestano, 2004, Financial Crisis Identification: a survey, Working Paper, University of Groningen.
- [78] Jacobs, J.P.A.M., Kuper, G.H., and Lestano, 2008, Currency Crises in Asia: A Multivariate Logit Approach, in *International Finance Review Asia Pacific Financial Markets: Integration, Innovation and Challenges* 8, 157-173, edited by S.J. Kim and M. McKenzie.
- [79] Jacone, L.I., 2004, The Late 1990s Financial Crisis in Ecuador: Institutional Weaknesses, Fiscal Rigidities, and Financial Dollarization at Work, IMF Working Paper 12, International Monetary Fund, Washington.
- [80] Jorda, O., Schularick M., and Taylor, A.M., 2011, Financial Crises, Credit Booms, and External Imbalances: 140 Years of Lessons, *IMF Economic Review* 59, issue 2, 340-378.
- [81] Kamin, S. B., Schindler, J., and Samuel, S., 2007, The contribution of domestic and external factors to emerging market currency crises: An early warning systems approach, *International Journal of Finance and Economics* 12, issue 3, 317-336.
- [82] Kaminsky, G., Lizondo, S., and Reinhart, C., 1998, Leading Indicators of Currency Crises, *IMF Staff Papers* 45, issue 1, 1-48.

-
- [83] Kaminsky, G., and Reinhart, C., 1999, The Twin Crises: the Causes of Banking and Balance of-Payments Problems. *American Economic Review* 89, 473-500.
- [84] Kaminsky, G.L., 2003, Varieties of Currency Crises, NBER Working Paper No. 10193.
- [85] Kapetanios, G., 2003, Determining the Poolability of Individual Series in Panel Datasets, Working Paper 499.
- [86] Kauppi, H., and Saikkonen, P., 2008, Predicting U.S. Recessions with Dynamic Binary Response Models, *The Review of Economics and Statistics* 90, issue 4, 777-791.
- [87] Kindleberger, C.P., 2000, *Comparative Political Economy, a Retrospective*, The MIT Press, Massachusetts.
- [88] Krugman, P., 1979, A Model of Balance of Payments Crises, *Journal of Money, Credit and Banking* 11, 311-25.
- [89] Krugman, P., 1997, Currency Crises. NBER conference.
- [90] Krugman, P., 2002, Crises: the Next Generation, in *Economic Policy in the International economy: Essays in Honor of Assaf Razin*, Assaf Razin, Elhanan Helpman, and Efraim Sadka, editors, Cambridge.
- [91] Kumar, M., Moorthy, U., and Perraudin, W., 2003, Predicting Emerging Market Currency Crashes, *Journal of Empirical Finance* 10, 427-454.
- [92] Lambert, J., and Lipkovich, I., 2008, A Macro for Getting More out of Your ROC Curve, SAS Global forum, paper 231.
- [93] Leaven, L., and Valencia, F., 2008, Systemic Banking Crises: A New Database, IMF Working paper 224, International Monetary Fund, Washington.
- [94] Lesaffre, E., and Kauffmann, H., 1992, Existence and Uniqueness of the Maximum Likelihood Estimator for a Multivariate Probit Model, *Journal of the American Statistical Association* 87, 419, 805-811.
- [95] Lestano, Jacobs, J.P.A.M., and Kuper, G. H., 2003, Indicators of Financial Crises Do Work! An early-warning system for six Asian countries, Working Paper.
- [96] Lestano and Jacobs, J.P.A.M., 2004, A Comparison of Currency Crisis Dating Methods: East Asia 1970-2002, CCSO Working Papers 200412, CCSO Centre for Economic Research.

- [97] Maddala, G.S., and Wu, S., 1999, A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, Oxford Bulletin of Economics and Statistics, special issue, 631-652.
- [98] Martinez-Peria, M.S., 2002, A Regime-Switching Approach to the Study of Speculative Attacks: A Focus on EMS Crises, Empirical Economics 27, issue 2, 299-334.
- [99] Masson, P. R., 1998, Contagion-Monsoonal Effects, Spillovers, and Jumps Between Multiple Equilibria, IMF Working Paper 142, International Monetary Fund, Washington.
- [100] McFadden, D., 1989, A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration?, Econometrica 57, 995-1026.
- [101] Obstfeld, M., 1994, The Logic of Currency Crises, Cahiers Economiques et Monetaires 43, 189-213.
- [102] Patton, A., 2011, Volatility Forecast Comparison using Imperfect Volatility Proxies, Journal of Econometrics, 160, issue 1, 246-256.
- [103] Peltonen, T., 2006, Are Emerging Market Currency Crises Predictable? - A Test, ECB Working Paper 571.
- [104] Pesaran, H., 2007, A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence, Journal of Applied Econometrics 22, issue 2, 265-312.
- [105] Pescatori, A., and Sy, A.N.R., 2007, Are Debt Crises Adequately Defined?, IMF Staff Papers 54, issue 2, 306-337.
- [106] Phillips, P.C.B., and Yu, J., 2011, Dating the Timeline of Financial Bubbles During the Subprime Crisis, Quantitative Economics 2, issue 3, 455-491.
- [107] Politis, D.N., Romano, J.P., and Wolf, M., 1999, *Subsampling*, Springer-Verlag, New-York.
- [108] Reinhart, C.M., and Rogoff, K., 2008, *This Time Is Different: Eight Centuries of Financial Folly*, Princeton University Press.
- [109] Reinhart, C.M., Rogoff, K., Qian, R., 2010, *On Graduation from Default, Inflation and Banking Crises: Elusive or Illusion?*, in Daron Acemoglu and Michael Woodford editors, Chicago: University of Chicago Press, 2011 forthcoming.
- [110] Reinhart, C.M., and Rogoff, K., 2011, From Financial Crash to Debt Crisis, American Economic Review 101, issue 5, 1676-1706.

-
- [111] Renault, O., and De Servigny, A., 2004, *The Standard & Poor's Guide to Measuring and Managing Credit Risk*, 1st ed. McGraw-Hill, 2004.
- [112] Rose, A.K., and Spiegel, M.M., 2010, Cross-Country Causes and Consequences of The 2008 Crisis: International Linkages And American Exposure, *Pacific Economic Review* 15, issue 3, 340-363.
- [113] Rose, A.K. and Spiegel, M.M., 2011. Cross-Country Causes and Consequences of the crisis: An update, *European Economic Review* 55, issue 3, 309-324.
- [114] Rosenberg, C., Halikias, I., House, B., Keller, C., Pitt, A., and Setser, B., 2005, "Debt-Related Vulnerabilities and Financial Crises An Application of the Balance Sheet Approach to Emerging Market Countries", IMF Occasional Paper 240, International Monetary Fund, Washington.
- [115] Schneider, M., and Tornell, A., 2000, Balance Sheet Effects, Bailout Guarantees and Financial Crises, *Review of Economic Studies* 71, 883-913.
- [116] Stein, R.M., 2005, The Relationship between Default Prediction and Lending Profits: Integrating ROC Analysis and Loan Pricing, *Journal of Banking & Finance* 29, 1213-1236.
- [117] Teräsvirta, T., 2006, Forecasting Economic Variables with Non Linear Models, in *Handbook of Economic Forecasting*, G. Elliott, C.W.J. Granger and A. Timmermann editors, Elsevier, volume 1, chapter 8, 413-457.
- [118] Tudela, M., 2004, Explaining Currency Crises: a Duration Model Approach, *Journal of International Money and Finance* 23, issue 5, 799-816.
- [119] Wallis, K.F., 2003, Chi-squared Tests of Interval and Density Forecasts, and the Bank of England's Fan Charts, *International Journal of Forecasting* 19, 165-175.
- [120] West, K.D., 2006, Forecast Evaluation, in *Handbook of Economic Forecasting*, G. Elliott and C. Granger and A. Timmermann editors, Elsevier, edition 1, volume 1, number 1.
- [121] Williams, R., 2004, A Note on Robust Variance Estimation for Cluster-Correlated Data", *Biometrics* 56, issue 2, 645-646.
- [122] Zhang, Z., 2001, Speculative Attacks in the Asian Crisis, IMF Working Paper 189, International Monetary Fund, Washington.

Résumé en Français

Y-a-t-il une chance que le monde ne connaisse plus des crises financières?

Dans une perspective historique, nul ne doute que les crises financières apparaissent plutôt comme la règle et non l'exception (Reinhart et al., 2010; Bordo et al., 2001b). A tel point qu'il serait sans doute fastidieux et quelque peu illusoire de vouloir dresser un simple inventaire historique exhaustif des crises financières à l'échelle internationale allant de l'emblématique crise des tulipes (1636), à la Bulle des Mers du Sud (1720), au krach boursier de 1929, jusqu'aux innombrables crises bancaires, de changes et de dette souveraine qui ont secoué l'économie mondiale à la fin du XXème siècle et au début de XXIème siècle (voir Kindleberger, 2000, pour un essai en la matière).

L'objet de cette thèse sera par conséquent plus limité et aura pour ambition de proposer une réflexion sur la faisabilité technique (au sens économétrique) des systèmes avancés de prévision de certains types de crises. Le terme "crise financière" fait en effet référence à une diversité de situations et de mécanismes économiques suivant les marchés ou les institutions qu'elles frappent. Dans ce chapitre introductif, nous considérerons trois principaux types de crises financières, *i.e.* les crises de change, les crises bancaires et les crises de dette (Reinhart et Rogoff, 2008), en insistant tout particulièrement sur les crises de change qui représentent l'objet d'étude principal des applications empiriques proposées dans les chapitres suivants.

Rappelons au préalable brièvement la définition de ces trois types de crises. Dans un régime de change fixe, une crise de change se traduit par l'abandon forcé de l'ancrage de la devise, de sorte que la monnaie est réalignée ou que le régime de change fixe est complètement abandonné. Dans un régime de change flexible, une crise de change se traduit par une forte dépréciation de la monnaie à court terme (Bordo et al., 2001a). Caprio et Klingebiel (1996) définissent les crises bancaires comme des situations où malgré l'insolvabilité d'une partie importante du secteur bancaire, celui-ci reste ouvert. Enfin, une crise de dette correspond à un défaut de paiement (total ou partiel) des obligations de dette, au reniement ou à la restructuration de la dette dans des termes moins favorables que ceux initialement prévus (Reinhart et Rogoff, 2011).

Deux courants de recherche, parfois distincts, parfois liés, ont reçu une attention particulière dans la littérature sur les crises financières depuis les années 90. Le premier courant s'est attaché à expliciter les mécanismes économiques favorisant l'émergence et la diffusion des crises (généralement passées). L'exemple emblématique de ces travaux de recherche, essentiellement théoriques, est celui des différentes générations de modèles de crises de change apparus depuis la fin des années 80. Parallèlement, un deuxième courant à la fois théorique et empirique, s'est intéressé à la prévision des crises et à la construction de systèmes avancés de détection de crises ou *Early Warning Systems* (EWS). Chaque crise engendrant un regain d'intérêt des autorités de régulation (au sens large) pour les modèles de prévision de crises, cela se traduit généralement par le développement de nouveaux modèles EWS.

Un observateur averti ou candide (suivant les avis) observerait aisément une évolution parallèle entre le nombre de crises financières observées à l'échelle internationales et le nombre d'EWS proposés par la littérature académique et utilisés par les autorités de régulation. A tel point qu'aujourd'hui, certains chercheurs académiques doutent de l'utilité même de ces modèles (cf. Rose et Spiegel, 2010, 2011). D'autres chercheurs au contraire (cf. Frankel et Saravelos, 2011), faisant fi de cette corrélation positive, pensent au contraire qu'il existe fondamentalement un ensemble d'indicateurs avancés pour les crises financières et que l'effort de recherche dans ce domaine doit être maintenu.

Mais ce débat académique sur l'existence même des EWS ne remet pas en cause leur utilité pour les organisations internationales et les autres acteurs de la vie économique. Des institutions si différentes que le Fond Monétaire International, la Réserve Fédérale Américaine, le Crédit Suisse, la Deutsche Bank, la Commission Bancaire Européenne ont besoin de ces modèles pour étayer leur processus de décision, aussi imparfaites soient les prévisions de crises qui en découlent. A cela une raison : il n'existe pas d'alternative objective, en dehors des avis d'experts, pour quantifier et justifier certaines politiques ou prises de décision. Pire encore, la mise en place des EWS est parfois la conséquence de véritables injonctions politiques. Ainsi en 2011, le G20 a décidé que le Fond Monétaire International devrait surveiller les niveaux de dette, de déficits budgétaires et de balances commerciales des pays représentant plus de 5% de l'output cumulé du G20 afin de réduire la probabilité d'apparition d'une autre crise financière globale.

Les EWS, que l'on croit ou non à la possibilité de prévoir les crises financières, sont appelés à jouer un rôle déterminant dans la définition des politiques économiques au niveau microéconomique, de même qu'au niveau macroéconomique et international. A l'extrême le fait de ne pas croire à la possibilité de prévoir les crises financières peut ne pas être antinomique avec le fait de vouloir améliorer les propriétés des EWS. On peut faire ici le parallèle avec la recherche en économétrie théorique sur les prévisions issues

des modèles non linéaires : certains tests récents permettent de comparer des prévisions issues de modèles qui peuvent être tous mal spécifiés et de déterminer le “moins mauvais” modèle (Corrandi et Svansson 2006b). Or, dans le domaine des EWS on constate par exemple qu’aucune méthodologie statistique de comparaison des EWS et d’identification du meilleur modèle (ou du moins mauvais suivant les points de vue) n’a pas été proposée à ce jour dans cette littérature.

Ainsi la problématique de notre thèse repose sur deux grandes questions : comment améliorer la spécification des EWS ? et comment déterminer le meilleur modèle prédictif ? Cette thèse essentiellement économétrique s’intéressera successivement à ces deux problématiques en s’appuyant sur les progrès récents dans le domaine de l’économétrie de la prévision, de l’économétrie de panel et de l’économétrie des modèles qualitatifs. Toutefois, avant d’exposer les contributions de cette thèse, nous présenterons un aperçu des principaux avancements théoriques et empiriques en matière d’analyse et prévision de crises financières.

I. Peut-on prévoir les crises financières?

L’approche la plus naturelle de la prévision des crises financières consiste à comprendre les causes des crises passées pour ensuite en déduire une spécification économétrique estimable (sous forme réduite ou non) destinée à prévoir la survenue de ces événements dans le futur. Cette approche sensiblement comparable à celle de la Cowles Commission s’applique par exemple dans le domaine des modèles de prévision de crises de change.

Les modèles dits de “première génération” avaient pour vocation de comprendre l’émergence des crises de changes apparues notamment dans les pays de l’Amérique Latine dans les années 80. Le conflit entre les politiques domestiques (politique monétaire expansionniste, déséquilibres budgétaires persistants) et l’arrimage du taux de change est considéré à l’origine de ces crises. Quand le montant des réserves atteint un niveau critique, les investisseurs lancent une attaque spéculative, ce qui entraîne la consommation rapide des réserves et finalement la modification des taux pivots ou l’abandon de la parité. Dans ce modèle, les choix erronés de politique macroéconomique conduisent à l’accroissement de la vulnérabilité du pays et à l’abandon de l’ancrage de la devise (Krugman, 1979; Flood et Garber, 1984). Les EWS correspondants sont donc fondés sur un ensemble de variables explicatives macroéconomiques.

Malgré les qualités de ces modèles, certains chercheurs considèrent qu’ils ne sont pas une représentation fidèle de la réalité des crises (Obstfeld, 1994; Krugman, 1997). La “deuxième génération” de modèles de crise est ainsi née en réponse à la crise ayant surpris le mécanisme de taux de change européen en 1991 et 1992. Fondés sur une représentation en équilibres multiples, les anticipations des agents dans ces modèles ne

dépendent pas des fondamentaux. Bien que l'ancrage de la devise semble soutenable et que la politique macroéconomique apparaît saine, des crises auto-réalisatrices peuvent apparaître dès que la confiance des marchés est entamée. Toutefois, ces théories n'ont pas pu expliquer la crise économique mexicaine (crise Tequila) de 1994-1995 et la crise financière asiatique de 1997-1998. Dans ce contexte, une "troisième génération" de crises qui explore les liens entre les crises de change, les crises bancaires et le secteur financier a émergé. On peut donc observer des déséquilibres de balance de paiement engendrés par des crises provenant d'un autre pays. Ainsi il apparaît que la détérioration de fondamentaux macroéconomiques ne constitue pas un facteur automatique de crise (Masson, 1998). Ce phénomène est ultérieurement connu dans la littérature sous le nom de contagion transfrontière (Bruinshoofd, Candelon et Raabe, 2008). Étant donné l'instabilité d'une multitude de mécanismes financiers et bancaires, un large éventail de facteurs peuvent entraîner le déclenchement d'une crise : des booms des prix d'actifs, des crises de bilan (Schneider et Tornell, 2000); l'accumulation excessive de dette externe (Corsetti, Pesenti et Roubini, 1999), etc. En particulier le cercle vicieux du désendettement peut entraîner des coûts très élevés pour l'économie réelle (Krugman, 2002). Ces modèles essayent par ailleurs d'expliquer la survenue de crises jumelles : crises de change et bancaires (Kaminski, 1999).

Cette succession de modèles de crises et d'explications théoriques donne le sentiment que les modèles "courent" après les crises et pose fondamentalement la question de la prévisibilité des crises de change. L'analyse des trois générations de modèles fait ainsi ressortir leur impuissance à expliquer l'origine des crises futures. Rose et Spiegel (2010, 2011) contestent ainsi l'idée même d'EWS. Pour autant, les coûts sociaux et économiques de ces crises récentes appellent sans équivoque à s'efforcer de construire de tels modèles, si imparfaits soient-ils.

De nombreux articles ont montré que les crises financières engendrent des coûts de sauvetage immenses et qu'elles entraînent de sérieuses conséquences en termes de pertes de production et de détérioration des conditions sociales et de l'emploi. Caprio et Klingebiel (1996) estiment que les coûts de sauvetage représentent en moyenne 10% du PIB, certaines crises étant plus coûteuses que d'autres : la crise mexicaine (1994) a coûté 20% du PIB, alors que la crise du Jamaïca (1996) a coûté 37% du PIB. Des coûts supplémentaires, *i.e.* des pertes de production économique (à travers des réductions de l'investissement et de la consommation notamment) sont dues à la fois au rationnement du crédit et à l'incertitude. Un rapport du FMI (1998) estimait que les pays émergents connaissent une perte cumulative de la production réelle d'approximativement 8% lors d'une crise de change sévère. Hoggarth et al. (2002) estiment quant à eux que les crises bancaires coûtent en moyenne 5.6% du PIB et les crises jumelles coûtent environ 29.9% du PIB. En

ce qui concerne les pertes cumulatives de production des crises jumelles, celles enregistrées par les pays de l'OCDE (23.8 % du PIB) dépassent les pertes des pays émergents (13.9% du PIB). Dans le cadre de la crise financière globale, le FMI notait en 2009 : “On estime que le PIB mondial (annuel) a baissé de 5% au quatrième trimestre, le déclin le plus significatif ayant été enregistré pour les pays développés (environ 7%)”. De tels coûts rendent caduques les arguments de Rose et Spiegel. Les coûts de ces crises sont incommensurablement plus élevés que les éventuels coûts engendrés par des modèles EWS mal spécifiés qui pourraient engendrer par exemple de faux signaux d'alarme. Les autorités de régulation au sens large ont besoin de tels modèles EWS, car la prévision correcte d'au moins une partie des crises financières mondiales, se traduirait certainement par des gains énormes.

Mais au-delà de cet argument d'autorité, revenons à la question de la prévisibilité des crises et au design d'un EWS “optimal”. Tout d'abord, force est de constater que l'objet prévu (la crise de change dans notre exemple) n'est pas invariant aux modifications de la politique et de l'environnement économique. En cela, la question de la prévisibilité des crises rejoint la critique de Lucas. Contrairement à la littérature microéconomique (par exemple sur le risque de défaut individuel ou credit scoring), où la survenue d'une crise (défaillance individuelle) est invariante à l'action menée par la banque, la survenue des crises macroéconomiques peut être affectée par les mesures de politique économique. Ainsi la réalisation ex-post de l'événement (crise ou pas crise dans un schéma binaire) peut ne pas correspondre à la prévision issue d'un EWS simplement en raison des actions de politique économique. Poussons le raisonnement plus loin : un EWS “optimal” est nécessairement un modèle faux au sens de l'erreur de type I puisque la vocation de ce modèle est de prévenir les crises. En cas de prévision de crise, des mesures de politique économique doivent être menées pour l'éviter ou en amoindrir les effets, et ainsi on observera ex-post une erreur de prévision si l'on suppose que la politique économique est efficace. La comparaison ex-post des indicateurs de crises et des probabilités de crise construites ex-ante par les EWS (ne prenant pas en compte l'intervention des régulateurs) doit donc être menée de façon particulièrement prudente.

Par ailleurs, la question se pose de l'horizon de prévision “optimal” des crises. Un EWS n'a de sens que s'il prévient la survenue des crises avec un délai suffisant long pour que les autorités puissent agir pour empêcher la crise et / ou en limiter les effets. Un EWS qui prévoit une crise de change dans la 1/2 heure n'a que peu d'intérêt eu égard aux délais de réaction des autorités monétaires. C'est pourquoi dans cette littérature les horizons de prévision sont généralement plus longs (de l'ordre de 6 mois à 2 ans) que dans la littérature économétrique classique consacrée aux problématiques de prévision.

Enfin, dans un contexte d'interdépendance de institutions financières et des marchés, où des nombreuses interconnexions entre les fondamentaux macroéconomiques et la mise en oeuvre de nouvelles mesures de politique économique existent, la question de l'amélioration des EWS et de leurs capacités prédictives devient cruciale. Cette question est d'autant plus importante que la fragilité financière se transmet facilement entre les pays, générant des crises globales, comme celle que l'on vient de vivre.

II. Early Warning Systems

Deux problématiques émergent dans le cadre de la littérature sur les EWS : la spécification des EWS et leur évaluation. Si la spécification des EWS, au sens économétrique, consiste à définir une fonction de lien entre différents indicateurs avancés et l'observation des crises financières, l'évaluation requiert des procédures de validation des capacités prédictives de ces modèles.

A. Spécification des EWS

Sur le plan technique la spécification d'un EWS comporte trois éléments : (i) une datation des crises, (ii) un ensemble de variables explicatives et (iii) une fonction de lien entre ces indicateurs et les probabilités de crise.

La littérature consacrée aux crises financières est caractérisée par l'absence de datation officielle des crises contrairement par exemple aux cycles économiques (avec notamment la datation "officielle" proposée par le National Bureau of Economic Research pour les Etats-Unis). Des lors, de très nombreuses méthodes ont été proposées pour identifier les dates d'entrée et de sortie de crise. Dans la plupart des cas, quelle que soit la crise considérée, ces méthodes sont basées sur des indices de pression du marché (voir Jacobs et al., 2004 pour une revue de la littérature). Ceci a des conséquences directes pour la modélisation des EWS. En effet, la variable dépendante intégrée dans un modèle de prévision de crises est généralement constituée par l'output d'un autre modèle, indépendant du modèle de prévision, que l'on qualifiera de modèle de datation. Cet indicateur construit est souvent caractérisé par une dépendance temporelle de type Markovienne souvent négligée en pratique dans la phase de modélisation de l'EWS, en dépit de son importance pour l'inférence statistique (Harding et Pagan, 2011).

Le choix des variables explicatives considérées dans un EWS peut être guidé par la théorie économique (voir la section) ou par une approche de type datamining. Jacobs et al. (2004) recensent ainsi le choix d'indicateurs avancés pour les crises de change, bancaires et de dette souveraine et leurs significativité dans un large nombre d'applications empiriques proposées dans la littérature. Plus récemment, Frankel et Saravelos (2010) s'intéressent aux indicateurs avancés pertinents pour la crise du 2008-2009 alors qu'Alessi

et Detken (2011) caractérisent les facteurs déterminants des cycles de prix d'actifs en temps réel.

Enfin, la spécification d'un EWS est fondée sur une fonction de lien entre ces indicateurs et les probabilités de crise. Un des premiers EWS, inspiré de la théorie du signal, a été proposé par Kaminski et al. (1998) pour les crises jumelles. Ils définissent un critère "noise-to-signal ratio (NSR)" d'identification du seuil au-delà duquel les indicateurs macroéconomiques signalent une crise. Par conséquent, les capacités prédictives de leur EWS dépendent de ce seuil puisque l'on prévoit l'apparition d'une crise lorsque ce seuil a été dépassé. Depuis ces travaux pionniers, toute une variété de spécifications d'EWS a été proposée dans la littérature. Toutefois, les EWS utilisant des modèles de régression qualitatives (de type logit / probit) restent les plus répandus non seulement dans la littérature académique consacrée aux EWS mais aussi dans la pratique (cf. IMF, Federal Reserve, Deutsche Bank, French Banking Commission, Asian Development Bank). Utilisés pour la première fois dans la littérature sur la prévision des crises de change par Berg et Patilo (1999), il est reconnu que ces modèles sont généralement meilleurs que ceux fondés sur l'approche du signal. Kumar et al. (2002) considèrent ainsi des modèles logit, Bussiere et Fratzscher (2006) proposent un logit multinomial pour capter le biais de l'après-crise et Kamin et al. (2007) fait appel à une approche de type probit. Signalons enfin les travaux de Tudela (2004) basés sur un modèle de durée pour analyser les déterminants des crises de change et de Peltonen (2006) qui considère des réseaux neuronaux. D'autres spécifications fondés notamment sur des modèles de type Markov-Switching ont été proposés (cf. Martinez-Peria, 2002; Abiad, 2003; et Fratzcher, 2003).

Les EWS fondés sur des modèles de régressions qualitatives sont également appréciés dans le cadre de la prévision d'autres types de crises financières. Dans cette perspective, Demirgüç et Detragiache (1998), Eichengreen et Rose (1998), Davis et Karim (2008) les appliquent à la prévision des crises bancaires, tandis que Detragiache et Spilimbergo (2001), Ciarlone et Trebeschi (2005), Fuertes et Kalotychou (2007) proposent ce type d'EWS pour les crises de dette souveraine. Notons encore que l'analyse de la prévisibilité des crises financières s'est étendue à la suite des événements du 2008-2009. Davis et Karim (2008) analysent des EWS du type logit et arbres binaires dans le contexte de la crise des subprimes. Phillips et Yu (2011), proposent un test diagnostique de bulle spéculative et Jorda et al. (2011), vérifie l'utilité des indicateurs de déséquilibres extérieurs pour prévoir les crises financières au travers de modèles logit.

B. Evaluation des EWS

Force est de constater que très peu d'études ont été consacrées à la problématique de l'évaluation des EWS. En effet, à notre connaissance, mis à part le critère NSR de

Kaminski et al. (1998), *i.e.* seuil qui discrimine les périodes de crise et de calme, aucune méthode statistique n'a été proposée pour évaluer les capacités prédictives de ces modèles. Or, le critère NSR ne peut pas être considéré comme satisfaisant car il ne prend pas en compte les deux types d'erreurs (de type I et II). Par ailleurs, aucun test statistique n'a été proposé pour comparer les prévisions issues de deux EWS concurrents.

Cette observation est d'autant plus étonnante qu'un très grand nombre de travaux d'économétrie ont été consacrés à la problématique générale de l'évaluation des prévisions. Les nombreuses avancées économétriques au cours de la dernière décennie, notamment sur la modélisation non-linéaire ont donné lieu à d'innombrables travaux consacrés à la construction et à l'évaluation des prévisions ponctuelles, par intervalle de confiance ou par densité (voir Terasvirta, 2006; West, 2006; Clark et McCracken, 2011, pour quelques revues de la littérature). On distingue les méthodes d'évaluation absolue des méthodes dites relatives ou de comparaison. Les méthodes usuelles de comparaison des prévisions ponctuelles issues de deux modèles concurrents (non-emboîtés ou emboîtés) sont généralement basées sur une fonction de perte associée à la séquence de prévisions (cf. Diebold et Mariano, 1995; Harvey et al., 1997; Clark et McCracken, 2001; Clark et West, 2007). L'évaluation absolue est alors plutôt fondée sur des critères (de type MSFE, MAPE etc.). Mais récemment, toute une littérature s'est développée autour de l'évaluation absolue et relative des prévisions par intervalle de confiance et par densité. Les travaux pionniers de Christoffersen (1998) ont introduit différentes définitions de la validité des intervalles de confiance ainsi que des stratégies de tests de type LR. Bao et al. (2004, 2007), Corradi et Swanson, (2006b) ont quant à eux proposé différents tests de spécification correcte et de comparaison pour les densités de prévision. Notons que ces tests permettent notamment de comparer des densités de prévisions issues de modèles potentiellement mal spécifiés.

Comment expliquer cet écart entre le peu d'intérêt consacré à l'évaluation des prévisions des EWS en particulier et la littérature abondante consacrée à l'évaluation des prévisions issue des modèles (notamment non linéaires) en général ? En quoi l'évaluation des EWS est-elle spécifique ? Techniquement, on peut identifier plusieurs différences entre les deux littératures. Tout d'abord, dans le cas des EWS, l'événement à prévoir, *i.e.* l'apparition des crises, n'est pas directement observable mais est le résultat d'un modèle de datation. Ensuite, les modèles EWS ont pour output des probabilités de survenues de crise. Mais ces deux différences ne peuvent expliquer à elles seules ce relatif manque d'intérêt pour l'évaluation des EWS. En effet, les EWS ne sont pas les seuls modèles de prévision pour lesquels la variable à prévoir est inobservable : c'est par exemple le cas aussi en économétrie financière avec les prévisions de VaR ou de volatilité. Pour autant, des méthodes d'évaluation robustes ont été proposées (Patton, 2011). De la même

façon, d'autres modèles, comme par exemple les modèles Markov Switching, génèrent des probabilités sans que cela empêche l'évaluation des prévisions.

Ce manque d'intérêt est d'autant plus paradoxal que des méthodes d'évaluation des prévisions de probabilités de crise existent, comme par exemple dans le cas des risques individuels (credit-scoring par exemple). Leur transposition à la validation des EWS doit permettre tant une évaluation absolue (par mesure des écarts entre les prévisions et les réalisations) qu'une évaluation relative des modèles (par comparaison de prévisions issues de spécifications alternatives). Mais cette transposition nécessite que l'on adopte une démarche rigoureuse de spécification du modèle et que l'on insiste tout particulièrement sur l'estimation d'un seuil optimal permettant d'identifier de façon optimale les périodes de crise et calme prévues.

III. Contributions

Les objectifs principaux de cette thèse d'économétrie appliquée sont de proposer (i) une méthode d'évaluation systématique des capacités prédictives des EWS et (ii) de nouvelles spécifications d'EWS visant à améliorer leurs capacités prédictives. Ce travail comporte quatre parties.

Les deux premières, correspondant au deuxième et troisième chapitre, relèvent de la problématique générale de l'évaluation des prévisions. Le deuxième chapitre propose un test original de validation des intervalles de prévision applicable notamment aux prévisions issues de modèles non-linéaires (HDR). Le troisième chapitre propose une méthode d'évaluation spécifique aux modèles de type EWS. Notons toutefois le caractère général de cette méthode qui peut être appliquée à la validation des prévisions des modèles de cycle économique et de façon plus générale à l'évaluation de tout modèle dont l'output est une séquence de probabilités.

Les deux chapitres suivants proposent des améliorations des spécifications des EWS pour les crises de change. Le quatrième chapitre insiste sur l'apport de la prise en compte de la dynamique dans la construction des EWS et le cinquième chapitre développe des EWS multivariés qui permettent d'analyser conjointement plusieurs types de crise.

Chapitre 2 : Testing Interval Forecasts: a GMM-Based Approach

Le deuxième chapitre intitulé "Testing Interval Forecasts: a GMM-Based Approach",¹ à paraître dans la revue *Journal of Forecasting*, propose un test original d'évaluation des prévisions par intervalles de confiances ou High Density Regions (HDR). En effet, bien que les intervalles de confiance constituent la méthode la plus généralement utilisée par les économistes pour rendre compte de l'incertitude sur les prévisions, très peu d'études

¹Ce chapitre de la thèse est issu de l'article Dumitrescu, Hurlin et Madkour (2011), "Testing Interval Forecasts: a GMM-Based Approach", *Journal of Forecasting* (à paraître).

ont été consacrées à l'évaluation de ce type de prévisions. La seule exception notable étant l'article fondateur de Christoffersen (1998).

Dans ce contexte, nous développons un test original basé sur une statistique de type J (Hansen, 1982) obtenue à partir de moments définis par les polynômes ortho-normaux associés à la distribution binomiale (Bontemps, 2006 ; Bontemps et Meddahi, 2005, 2011). Ce test est de type model-free et peut être appliquée à des prévisions par intervalles de confiance ou HDR issues de n'importe quel type de modèle linéaire ou non.

Le test est fondé sur la notion de violations (Christoffersen, 1998). Une violation est définie comme une situation dans laquelle la valeur réalisée ex-post se situe en dehors de l'intervalle de confiance ou de l'HDR prévu ex-ante. Notre approche originale s'appuie sur la transformation de la série de violations dans une série de sommes de violations définies pour H blocs de taille N . Sous l'hypothèse nulle de validité des intervalles de confiance ces sommes sont distribuées selon une loi binomiale, de sorte que l'analyse de la validité de ces intervalles revient à tester l'hypothèse de distribution binomiale pour le processus de violations. Nous dérivons la distribution asymptotique du test sous l'hypothèse nulle de validité du modèle de prévision.

Notre test GMM présente de très bonnes propriétés à distance finie dans des échantillons de petite taille, puisque la distribution en échantillon fini de la statistique du test est très proche de la distribution du chi-deux indépendamment de la taille de blocs choisie. Plus précisément, si la taille des blocs est petite, le nombre de blocs est important et la statistique converge alors vers sa distribution asymptotique du chi-deux. En revanche, si la taille des blocs est grande, le nombre de blocs est petit et cette convergence en loi ne peut pas être obtenue. Mais dans ce cas, chaque somme de violations tend en distribution vers une loi normale du fait des propriétés de la loi binomiale. Sous l'hypothèse d'indépendance, la statistique du test équivalente à la somme des carrés de variables distribuées selon une loi normale, tend vers une distribution du chi-deux. On retrouve alors la même distribution que la distribution asymptotique. Notre test présente plusieurs avantages. Premièrement, il fournit un cadre unifié qui permet de tester de manière indépendante les hypothèses de couverture conditionnelle, d'indépendance et de couverture non-conditionnelle. Deuxièmement, aucune restriction n'est pas imposée sous l'hypothèse alternative. Troisièmement, notre test d'évaluation est toujours faisable et très simple à implémenter. Par ailleurs, les expériences Monte-Carlo montrent que pour des tailles d'échantillon typiques d'applications sur données réelles, notre test GMM a des très bonnes propriétés notamment en termes de puissance.

Chapitre 3 : How to Evaluate an Early Warning System?

Le troisième chapitre intitulé “How to Evaluate an Early Warning System? Towards a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods”,² publié dans la revue IMF Economic Review, propose une méthodologie originale et unifiée d’évaluation des capacités predictive des EWS. Comme nous l’avons dit précédemment, la validation des prévisions des EWS a effectivement fait l’objet de très peu de travaux dans la littérature. La plupart des études ont jusqu’ici utilisé le critère QPS pour évaluer les modèles et se sont basées sur le seuil déterminé par la méthode NSR de Kaminiski et al. (1998) pour discriminer entre les périodes prévues de crise et de calme. De plus, l’identification du modèle optimale ne fait généralement l’objet d’aucune inférence statistique.

Dans ce contexte, nous proposons une méthodologie d’évaluation originale, de type model-free, qui peut être appliquée à tout type d’EWS, pour toutes les crises financières, que ce soit pour des échantillons d’estimation ou de validation. Cette procédure se décompose en deux étapes. Dans la première étape, on identifie un cut-off optimal, *i.e.* le seuil qui discrimine le mieux entre les périodes prévues de crise et de calme. Cette identification du cut-off optimal est fondée sur les concepts de sensibilité et spécificité. Nous montrons que cette méthode d’identification améliore sensiblement les prévisions comparativement aux méthodes existantes telles que la méthode NSR et / ou le recours à des seuils arbitrairement choisis sans prise en compte des deux types d’erreurs.

Dans la deuxième étape, nous procédons à une évaluation comparative de modèles alternatifs sur la base des critères et de tests statistiques de validation. Nous proposons notamment différents critères de validation (ROC curve) et différents tests (tests d’aire sous la ROC curve) directement inspirés de la littérature sur les modèles de prévision des risques individuels (credit-scoring). Nos analyses, à la fois théoriques et empiriques révèlent l’importance capitale de ces tests (le test de l’aire sous la courbe ROC en particulier) pour l’identification d’un EWS optimal. Nous montrons qu’une évaluation correcte des EWS doit être fondée sur la prise en compte du cut-off optimal dans les deux étapes. Notre méthodologie permet alors d’améliorer sensiblement le diagnostic de performance de différentes spécifications de type EWS.

Nous illustrons cette méthodologie en nous intéressant à la question de la prise en compte du spread de taux dans la prévision des crises pour 12 pays émergents. Il en ressort principalement que le spread est un indicateur notable de crise de change uniquement pour la moitié des pays, si le test AUC est utilisé, alors que les critères usuels (de type QPS) laissaient à penser que ce facteur était important pour tous les pays. De plus, le

²Ce chapitre est issu de l’article Candelon, Dumitrescu et Hurlin (2012), “How to Evaluate an Early Warning System? Towards a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods”, IMF Economic Review 60(1).

cut-off optimal identifie correctement plus de deux tiers des périodes de crise et de calme, contrairement au seuil NSR qui n'identifie pas la plupart des crises.

Chapitre 4 : Currency Crises Early Warning Systems: Why They Should Be Dynamic?

Le quatrième chapitre, “Currency Crises Early Warning Systems: Why They Should Be Dynamic?”, ³ met en évidence l'intérêt de la prise en compte de la dynamique, *i.e.* de la persistance des crises, dans la prévision des crises.

La dépendance temporelle des indicateurs de crises construits par les modèles de datation a notamment été observée par Berg et Coke (2004). C'est pourquoi dans ce chapitre nous proposons une nouvelle génération d'EWS qui tient compte de la dynamique du phénomène analysé ainsi que du caractère binaire de l'indicateur de crise. Cette dynamique endogène des crises peut, en effet, être intégrée dans le modèle de différentes façons. Une première façon de le faire consiste à considérer l'indicateur binaire retardé de crise dans le modèle. En l'occurrence, la crise se transmet de manière non-linéaire d'une période à l'autre, puisque l'index doit dépasser un seuil afin de déclencher une crise. Un autre moyen est de considérer un modèle autorégressif pour l'index de crise. Finalement, ces deux types de dynamique peuvent être englobés dans une même et seule spécification d'un modèle dichotomique dynamique.

Dans ce contexte, nous proposons le premier EWS fondé sur une spécification dynamique d'un modèle dichotomique estimé par une méthode exacte de maximum de vraisemblance (Kauppi et Saikonen, 2008) et nous testons ces spécifications dans un cadre unifié. Dans la seconde partie de l'article, nous étendons cette approche en panel en nous appuyant notamment sur les travaux récents de Carro (2007). Notre EWS est facile à implémenter pour tout type de crise et peut intégrer des indicateurs macroéconomiques avancés, source de persistance exogène de crise.

Une application empirique sur 15 pays émergents nous permet de comparer les capacités prédictives de notre EWS à celles de deux modèles concurrents : un modèle de type Markov-Switching et un modèle de type logit statique. Pour ce faire, nous utilisons la procédure d'évaluation développée dans le chapitre précédent. Les résultats montrent que les modèles logit dynamiques présentent de bien meilleures capacités prédictives que les modèles statiques et que les modèles de type MS non seulement pour l'échantillon d'estimation, mais aussi pour celui de validation. L'EWS dynamique présente aussi des très bonnes propriétés quant à l'identification des périodes de crise et de calme. Ces résultats montrent que la dynamique des crises doit être considérée dans la spécification

³Ce chapitre est issu de l'article Candelon, Dumitrescu et Hurlin (2010), “Currency Crises Early Warning Systems: Why They Should Be Dynamic?”, METEOR Research memorandum RM/10/047.

des EWS, car elle améliore sensiblement la qualité du signal nécessaire pour prévoir les crises.

Chapitre 5 : Modelling Financial Crises Mutation

Le cinquième chapitre, “Modelling Financial Crises Mutation”,⁴ propose un modèle dynamique multivarié qui analyse conjointement les trois principaux types de crises financières, *i.e.* les crises bancaires, les crises de change et les crises de dette souveraine. Ce modèle permet par ailleurs d’analyser les schémas de causalité intervenant entre ces différentes crises.

Cet article constitue une extension multivariée des EWS qui s’intéressent seulement au cas de crises jumelles (Glick et Hutchinson, 1999). Pour ce faire, on considère une approche méthodologique originale fondée sur l’estimation d’un modèle probit dynamique multivarié par la méthode du maximum de vraisemblance exact. Ainsi, nous étendons la spécification de Huguenin et al. (2009) aux modèles dynamiques et les spécifications de Kauppi et Saikonen (2008) aux modèles multivariés.

L’illustration empirique montre clairement que le modèle trivarié améliore le modèle bivarié pour les pays qui connaissent les trois types de crises, alors qu’en bivarié la causalité entre les crises bancaires et les crises de change (et à l’inverse) est fréquente. L’avantage essentiel de cette méthodologie consiste à identifier les schémas de contagion d’un type de crise à l’autre: soit c’est le résultat des chocs communs (comme en Afrique du Sud), soit c’est due à une structure causale forte (comme en Equateur). Les graphiques des probabilités conditionnelles et l’analyse des fonctions de réponse confirment nos résultats et mettent en évidence les mécanismes de diffusion des trois types de crise. La possibilité que les crises se transforment devrait alors être prise en compte plus souvent dans la spécification des EWS, afin de les rendre plus efficaces.

Finalement, le sixième chapitre résume les principaux résultats de cette thèse et expose plusieurs pistes de développements.

⁴Ce chapitre est issu de l’article Candelon, Dumitrescu, Hurlin et Palm (2011), “Modelling Financial Crises Mutation”, DR LEO 2011-17.

Nederlandse samenvatting

Dit proefschrift presenteert vier artikelen die op een verschillende manier bijdragen aan de literatuur over de Early Warning Systems. In het kielzog van de wereldwijde financiële crisis, zijn er vele vragen gerezen met betrekking tot het voorspellende vermogen van EWS omdat zeer weinig signalen zijn getrokken tot aan het begin van het oproer. Twee onderzoeksonderwerpen zijn relevant voor deze literatuur, namelijk de specificatie van EWS en de evaluatie van deze modellen. De meeste artikelen in de literatuur focussen op het eerste onderwerp. Ze bekritisieren de oorzaken van de crisis, i.e. de voornaamste indicatoren in EWS, en/of stellen nieuwe methodologiën voor om de link tussen de crisis indicator en de macro-economische en financiële variabelen te benaderen. In tegenstelling, de literatuur over EWS evaluatie is schaars. Voor zover wij weten, is er tot nu toe nog geen formele evaluatie methodologie voor het voorspellende vermogen van een EWS. Dit proefschrift draagt bij aan beide onderzoeksrichtingen.

De voornaamste doelen van dit proefschrift zijn daarom het voorstellen van voorspellende evaluatie methoden voor non-lineaire modellen (in het bijzonder, een verenigde evaluatie methodologie voor EWS), als ook het introduceren van methodologische noviteiten gebaseerd op recente econometrische ontwikkelingen in de specificatie van EWS, zodoende een verbetering van hun voorspellende vermogen. Hoofdstuk 2 en hoofdstuk 3 onderzoeken het eerste doel, terwijl hoofdstuk 4 en hoofdstuk 5 het tweede doel analyseren.

EWS evaluatie is essentieel voor de preventie van crisissen. Vele EWS specificaties zijn voorgesteld in de literatuur en verschillende conclusies zijn getrokken met betrekking tot het belang van bepaalde toonaangevende indicatoren voor verschillende crisissen. Een accurate vergelijking van het relatieve voorspellende vermogen van deze modellen en een duidelijke analyse van hun absolute voorspellende vermogen lijkt daarom noodzakelijk. In hoofdstuk 2 en 3 vullen we deze opening in de literatuur door twee voorspellende validatie methodologiën voor te stellen: een ontworpen voor interval voorspellingen, en een toolbox die ieder model evalueert en de kans voortbrengt om een gebeurtenis te observeren (crisis, recessie, etc.). Ten eerste, stellen we een model vrije test voor, voor interval voorspellingen en High Density Regions gebaseerd op de GMM methode voorgesteld door Bontemps (2006) en Bontemps en Meddahi (2005, 2011). Onze test heeft hele goede power eigen-

schappen voor realistische steekproefgrootten en stelt ons in staat om onafhankelijk de drie hypothesen van conditionele dekking, onafhankelijkheid en onconditionele dekking te testen. Ten tweede, introduceren wij een origineel, tweestaps model vrije evaluatie methodologie speciaal ontworpen voor EWS modellen voor ieder type crisis. Een originele methode om de grens te identificeren, i.e. de drempel die het beste onderscheid maakt tussen crisissen en kalme perioden is voorgesteld. Daarnaast worden verschillende vergelijkings testen gebruikt om het relatieve voorspellende vermogen te analyseren van EWS. Hierdoor zijn we in staat om te laten zien dat alleen voor Zuid-Aziatische landen het renteverskil een belangrijke indicator van valuta crisissen is en niet voor alle landen zoals in eerste instantie werd gedacht.

Om het voorspellende vermogen van EWS modellen te verbeteren moet rekening gehouden worden met verschillende karakteristieken van financiële crisissen die niet eerder in de literatuur onderzocht zijn (bijvoorbeeld het voortduren van crisissen en de mogelijkheid tot spill-over naar andere markten). We dragen daarom bij aan de literatuur over EWS specificatie door nieuwe methodologiën voor te stellen, die informatie bevatten met betrekking tot het overheersende regime in de vorige periode en/of in andere markten, i.e. crisis dynamiek en mutatie. Deze verbeteringen van bestaande binaire EWS worden besproken in hoofdstuk 4 en 5. We laten zien dat in het geval van valuta crisissen dynamische modellen beter presteren dan Markov-switching en statische logit modellen zowel in de steekproef als buiten de steekproef. Bovendien, een multivariabel model dat de drie belangrijkste types van financiële crisissen bevat, i.e. banken, valuta en staatsschuld laat ons de mogelijke causaliteit tussen crisissen onderzoeken: dit kan veroorzaakt worden door gemeenschappelijke schokken (zoals in Zuid Afrika) of door een sterk causaal verband (zoals in Ecuador). Spaarzamer dan bivariabele modellen, zou het wanneer mogelijk geïmplementeerd moeten worden.

Curriculum Vitae

Elena-Ivona Dumitrescu was born on November 28, 1984 in Tirgu-Jiu, Romania. After having completed a Bachelor's degree in Statistics and Economic Forecasting with distinction in 2007, she obtained an Erasmus scholarship and enrolled at the University of Orléans, France, where she completed a Master in Econometrics and Applied Statistics with honors in 2009.

After graduation, Elena started her Ph.D. research at the University of Orléans and Maastricht University, under the joint supervision of Prof. Christophe Hurlin and Prof. Bertrand Candelon. Her work, tackling financial crises forecasting methods and other econometric topics, has been presented at top European conferences (European Economic Association, Glasgow, 2010; Econometric Society World Congress, Shanghai, 2010; IN-FINITI, Dublin, 2010; Eurostat Colloquium, Luxembourg, 2010; European Meeting of the Econometric Society, Oslo, 2011), congresses in France (AFFI, 2010; AFSE, 2010, 2011) and in IMF seminars (2009, 2011). Some of these articles are forthcoming in international peer reviewed academic journals (*e.g.* “How to Evaluate an Early Warning System?”, *IMF Economic Review* 60 (1); “Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests”, *Finance*; “Testing for Granger Non-causality in Heterogeneous Panels”, *Economic Modelling*; “Testing Interval Forecasts: a GMM-Based Approach”, *Journal of Forecasting*).

In 2010 she earned an EOLE grant, which allowed her to conduct more of her research at Maastricht University. Besides, in the fall 2011 she was a visiting Ph.D. at the European University Institute, where she will start a Postdoctorate in September 2012. For more details, you can visit her homepage at <https://sites.google.com/site/ivonadumitrescu/>.

Elena-Ivona DUMITRESCU

Méthodes Econométriques pour les Crises Financières

Résumé :

Connus sous le nom de Systèmes d'Alerte Avancés, ou Early Warning Systems (EWS), les modèles de prévision des crises financières sont appelés à jouer un rôle déterminant dans l'orientation des politiques économiques tant au niveau microéconomique qu'au niveau macroéconomique et international. Or, dans le sillage de la crise financière mondiale, des questions majeures se posent sur leur réelle capacité prédictive. Deux principales problématiques émergent dans le cadre de cette littérature : comment évaluer les capacités prédictives des EWS et comment les améliorer ?

Cette thèse d'économétrie appliquée vise à proposer (i) une méthode d'évaluation systématique des capacités prédictives des EWS et (ii) de nouvelles spécifications d'EWS visant à améliorer leurs performances. Ce travail comporte quatre chapitres. Le premier propose un test original d'évaluation des prévisions par intervalles de confiance fondé sur l'hypothèse de distribution binomiale du processus de violations. Le deuxième chapitre propose une stratégie d'évaluation économétrique des capacités prédictives des EWS. Nous montrons que cette évaluation doit être fondée sur la détermination d'un seuil optimal sur les probabilités prévues d'apparition des crises ainsi que sur la comparaison des modèles. Le troisième chapitre révèle que la dynamique des crises (la persistance) est un élément essentiel de la spécification économétrique des EWS. Les résultats montrent en particulier que les modèles de type logit dynamiques présentent de bien meilleures capacités prédictives que les modèles statiques et que les modèles de type Markoviens. Enfin, dans le quatrième chapitre nous proposons un modèle original de type probit dynamique multivarié qui permet d'analyser les schémas de causalité intervenant entre différents types de crises (bancaires, de change et de dette). L'illustration empirique montre clairement que le passage à une modélisation trivariée améliore sensiblement les prévisions pour les pays qui connaissent les trois types de crises.

Mots clés : Crise financière, Early Warning System, Prévision par Intervalle de confiance, High Density Regions, Evaluation des EWS, Modèle Dynamique, Modèles Probit Dynamiques Multivariés, GMM, Credit-Scoring.

Econometric Methods for Financial Crises

Abstract:

Known as *Early Warning Systems (EWS)*, financial crises forecasting models play a key role in defining economic policies at microeconomic, macroeconomic and international level. However, in the wake of the global financial crisis, numerous questions with respect to their forecasting abilities have been raised, as very few signals were drawn prior to the starting of the turmoil. Two questions arise in this context: how to evaluate EWS forecasting abilities and how to improve them?

The broad goal of this applied econometrics dissertation is hence (i) to propose a systematic model-free evaluation methodology for the forecasting abilities of EWS as well as (ii) to introduce new EWS specifications with improved out-of-sample performance. This work has been concretized in four chapters. The first chapter introduces a new approach to evaluate interval forecasts which relies on the binomial distributional assumption of the violations series. The second chapter proposes an econometric evaluation methodology of the forecasting abilities of an EWS. We show that adequate evaluation must take into account the cut-off both in the optimal crisis forecast step and in the model comparison step. The third chapter points out that crisis dynamics (persistence) is essential for the econometric specification of an EWS. Indeed, dynamic logit models lead to better out-of-sample forecasting probabilities than those of their main competitors (static model and Markov-switching one). Finally, a multivariate dynamic probit EWS is proposed in the fourth chapter to take into account the causality between different types of crises (banking, currency, sovereign debt). The empirical application shows that the trivariate model improves forecasts for countries that underwent the three types of crises.

Keywords : Financial Crisis, Early Warning System, Interval Forecasts, High Density Regions, EWS Evaluation, Dynamic Model, Multivariate Dynamic Probit Models, GMM, Credit-Scoring.



Laboratoire d'Economie d'Orléans; UMR CNRS 7322
Rue de Blois - BP 6739
45067 ORLEANS Cedex2

