Identification and Estimation in an Incoherent Model of Contagion*

Daniele Massacci
Cambridge University

October 20, 2007

Abstract

This paper deals with the issues of identification and estimation in the canonical model of contagion advanced in Pesaran and Pick (2007). The model is a two-equation nonlinear simultaneous equations system with endogenous dummy variables; it also represents an extension of univariate threshold autoregressive (TAR) models to a simultaneous equations framework. For a range of economic fundamentals, the model produces multiple (i.e. two) equilibria, and the choice of the equilibrium is modelled as being driven by a Bernoulli process; further, the presence of multiple equilibria leads to an incoherent econometric specification. The coherency issue is then reflected in the analytical expression for the likelihood function derived in the paper. It is proved that neither identification nor Full Information Maximum Likelihood (FIML) estimation of the model require knowledge of the Bernoulli process driving the solution choice in the multiple equilibria region. Monte Carlo experiments show that the FIML estimator performs better than the GIVE estimators proposed in Pesaran and Pick (2007). Finally, an empirical illustration based on stock market returns is provided.

JEL classification: C10, C13, C15, C32, G10, G15

Keywords: Contagion, Identification, Estimation, Coherent Models, Threshold Models

*I would like to thank Hashem Pesaran for his continuous advice and comments. This paper also benefits from conversations with Mardi Dungey, Andreas Pick, Alessio Sancetta and Takashi Yamagata. Constructive comments from seminar participants at the Faculty of Economics and CIMF, University of Cambridge, have been most helpful. Financial support from the following institutions is gratefully acknowledged: Corpus Christi College Cambridge; the Economic and Social Research Council; European Trust and Faculty of Economics, University of Cambridge; Tudor Investment Corporation. Errors and omissions are my own responsibility.

1 Introduction

Market crashes have not been unusual in the financial world. Three of the biggest historical crashes that have been witnessed are: the Tulipmania, which invested the Republic of the Netherlands during the 17th century; the South Sea Bubble, which derives its name from the South Sea Company, whose stock's value plunged at the end of May 1720; the 1929 US stock market crash, which started on Black Tuesday, October 29. More recent crisis episodes that deserve to be mentioned are: the 1987 US stock market crash, which affected other major world markets, such as Honk Kong, London, Tokyo and Singapore; the 1992 collapse of the Exchange Rate Mechanism of the European Monetary system, which started with the speculative attacks to the Italian Lira and the British Pound, and then spread to other countries as well; the December 1994 Mexican crisis, which also affected countries like Argentina, Brazil and the Philippines and generated the well known Tequila effect; the 1997 Asian crises, which began with the depreciation of the Thai baht, and then propagated to Malaysia, Indonesia, the Philippines, Singapore, Taiwan and South Korea; the 2001 Argentine crisis, which determined a cluster of turmoils in countries like Paraguay, Uruguay and Brazil. Finally, since the end of February 2007, the international financial system has been characterised by a situation of uncertainty, which has been mainly generated by the crisis in the US subprime mortgage market: this led to a condition of turmoil in Asian, European, and American markets.

From the previous discussion, it emerges that in more recent times financial shocks do not remain confined to the market where they generate; rather, they tend to spread to other markets, giving rise to a cluster of crises. The theoretical economic literature has identified a number of possible causes for the simultaneous occurrence of financial crises across markets; following Masson (1999) they can be classified as monsoonal effects, spillovers and pure contagion. *Monsoonal effects* are determined by the dependence of macroeconomic fundamentals upon a common source: for example, developing countries strongly depend on industrial countries, and a negative economic shock in the latter is likely to affect the former and possibly result in a cluster of financial crises. *Spillovers* are driven by the correlation between external economic linkages: for example, if two countries have strong trade linkages and one of them is hit by a crisis and has to devalue, the other one will be forced to devalue as well in order to maintain its degree of competitiveness. Finally, *pure contagion* occurs when a crisis spreads from one country to another one without any change in macroeconomic fundamentals; as such, pure contagion is modelled as a situation characterised by the presence of *multiple equilibria*, where the economic system jumps from one equilibrium to another one. Following Pesaran and Pick (2007), in what follows we will simply refer to pure contagion as *contagion*, while monsoonal effects and spillovers will be jointly referred to as *interdependence effects*.

The definition of contagion we adopt can be interpreted in two complementary ways. Following a probabilistic approach, contagion takes place if the occurrence of a crisis in one market increases the likelihood of a crisis in another market above the level implied by economic fundamentals. Alternatively, contagion is said to occur if the degree of correlation between two markets during crisis periods increases relatively to tranquil periods. These two ways of defining contagion are consistent with the restrictive and very restrictive definitions provided by the World Bank, respectively[1].

The distinction between interdependence and contagion is of interest to policy makers (either in international financial institutions, such as the IMF, or at central banks) and to profit maximisers investors. In the former case, if a random jump from a "good" to a "bad" equilibrium (i.e. contagion) occurs, then a policy intervention could be effective; conversely, in the case of interdependence a similar action is unlikely to have any significant effect. In the case of investors, the exposition to market risk can be generally reduced by portfolio diversification; however, if contagion occurs then the degree of dependence between markets increases, and portfolio diversification may not be an effective strategy to follow. Therefore, because of the different effects they have on economic agents' decision process, the *identification* of interdependence and contagion effects has to be achieved.

There now exists an extensive empirical literature that aims at identifying contagion from interdependence effects, as surveyed in Rigobon (2001), Billio and Pelizzon (2003), Pericoli and Sbracia (2003), Dungey *et al.* (2005*a*), Dungey *et al.* (2005*b*) and Massacci (2007). In particular, two sets of papers can be identified, each related to one of the definitions of contagion previously discussed.

Following the probabilistic approach, Eichengreen *et al.* (1996), Kruger *et al.* (1998), and Stone

[1] http://www1.worldbank.org/economicpolicy/managing%20volatility/contagion/definitions.html

and Weeks (2001) make use of binary choice models. In each equation, the discrete dependent variable takes a unit value if the underlying continuous variable crosses a given threshold, meaning that the associated market is in crisis. Contagion effects are then captured by the inclusion of a dummy variable as a covariate, which takes value equal to one if any other market in the sample is in crisis. Under the assumption of independent and normally distributed errors, a pooled probit model is then estimated. Esquivel and Larraín (1998) and Kumar *et al.* (2002) follow a similar approach; however, they estimate a random effect probit model and a logit model, respectively. However, this class of models suffers from two main shortcomings: first, by construction the contagion dummy is endogenous rather than exogenous; second, the assumption of independently distributed errors is unlikely to hold in practice, since their cross sectional correlation captures interdependence effects. As a consequence, the resulting Maximum Likelihood estimator delivers inconsistent estimates of the parameters[2].

Following the seminal paper by King and Wadhwani (1990), Boyer *et al.* (1999), Loretan and English (2000), Ronn *et al.* (2001) and Forbes and Rigobon (2002) assess the presence of contagion by testing whether there is a significant increase in the level of correlation between markets in crisis periods relatively to tranquil ones. This correlation-based approach is then extended in Rigobon (2000), Rigobon (2003*a*), Rigobon (2003*b*) and Dungey *et al.* (2005*a*) by developing testing procedures for contagion based on the change in the reduced form covariance matrix of the underlying structural model between tranquil and crisis periods. The problem with this methodology is that it requires *a priori* specification of crisis periods; therefore, if crisis windows are misspecified then the estimators of the parameters are likely to be inconsistent. Further, the models can only be employed to perform in-sample analysis: they become of little use when performing out-of-sample forecasting analysis, which plays a crucial role in the development of an early warning system for policy makers.

A further set of papers can be identified, made of contributions that can be related to the two groups of studies previously discussed. Favero and Giavazzi (2002) make use of a linear simultaneous equations system. In each structural equation, contagion effects are captured by the inclusion of a dummy variable as a covariate, which takes value equal to unity if the error of the corresponding reduced form equation crosses a threshold. Under the assumptions of normally distributed structural errors and of exogenous contagion dummies, the model is estimated by Full Information Maximum Likelihood. Because of the use of contagion dummies, the Favero and Giavazzi (2002) methodology can be linked to the probabilistic approach previously discussed; however, it also suffers from an analogous problem: by construction the contagion dummies are actually endogenous, and the employed estimator delivers inconsistent estimates of the parameters of the model.

A further approach is followed *inter alia* by Jeanne and Masson (2000), Fratzsher (2003) and Billio *et al.* (2005). They endogenously identify crisis periods by introducing a hidden state variable that follows a Markov Chain, so that the resulting model is a Markov switching model. In this way, crisis periods no longer need to be *a priori* specified, as it is the case in the studies where correlation or covariance-based tests of contagion are employed. However, in a multivariate framework, the hidden state variables are assumed to be orthogonal to each other, implying that contemporaneous crisis episodes are treated as independent events.

This paper aims at analysing a contagion model that overcomes the shortcomings of the models we have previously discussed. Our work focuses upon the canonical model of contagion advanced in Pesaran and Pick (2007). The model is a two-equation nonlinear simultaneous equations system with endogenous dummy variables; it is also an extension of univariate threshold autoregressive (TAR) models to a simultaneous equations framework. As it will be shown, for a range of economic fundamentals the model produces multiple (i.e. two) equilibria, and the resulting equilibrium is chosen by a selection indicator that can be modelled as a Bernoulli process. Further, the presence of multiple equilibria leads to an incoherent econometric specification.

Our analysis focuses on identification and estimation of the Pesaran and Pick's (2007) model. In particular, it is shown that the coherency issue does not affect identification of the model, which is achieved by exploiting the nonlinear nature of the system. As far as estimation is concerned, the expression for the likelihood function is derived: the analytical expression for the pdf is affected by the

[2] The use of binary choice models to identify contagion effects is justified by an interest upon extreme events taking place at the tails of the *marginal* distribution of asset returns. Bae *et al.* (2003) focus their attention on events at the tails of the *joint* distribution of returns; as such, they estimate a multinomial logit model. However, their approach suffers from the same shortcomings as that based on binary choice models.

coherency issue; however, it does not require knowledge of the Bernoulli process driving the solution choice in the multiple equilibria region. The empirical validity of the theoretical results are confirmed by the Monte Carlo analysis we conduce, which also shows that the Full Information Maximum Likelihood estimator we propose performs better than the GIVE estimators advanced in Pesaran and Pick (2007). Finally, an empirical illustration seems to provide evidence in favour of the presence of contagion effects in equity markets.

The rest of the paper is organised as follows. The contagion model is introduced in Section 2, and its solution is presented in Section 3. The coherency issue is discussed is Section 4. Identification and estimation of the model are dealt with in Section 5 and Section 6, respectively. A Monte Carlo analysis is conduced in Section 7. An empirical illustration is provided in Section 8. Finally, concluding remarks are given in Section 9. Concerning notation, $\overset{a.s.}{\rightarrow}$ and $|\cdot|$ denote almost sure convergence and the Euclidean norm, respectively.

2 Contagion Model

Consider the following two-equation model specification introduced in Pesaran and Pick (2007)

$$\begin{array}{c} y_{1t} = \boldsymbol{\delta}_1' \mathbf{z}_t + \boldsymbol{\alpha}_1' \mathbf{x}_{1t} + \beta_1 \mathbf{I}\left(y_{2t} - c_2\right) + u_{1t}, \\ y_{2t} = \boldsymbol{\delta}_2' \mathbf{z}_t + \boldsymbol{\alpha}_2' \mathbf{x}_{2t} + \beta_2 \mathbf{I}\left(y_{1t} - c_1\right) + u_{2t}, \end{array} \tag{1}$$

where the dependent variable y_{it} is a performance indicator for market $i = 1, 2,\ t = 1, \ldots T$. The regressors $\mathbf{x}_{it}$ are $k_i \times 1$ vectors of market specific regressors (which may include lagged values of y_{it}) such that $\mathbf{x}_{1t} \cap \mathbf{x}_{2t} = \varnothing$, while $\mathbf{z}_t$ is an $s \times 1$ vector of observable common factors; the elements of $\mathbf{x}_{it}$ and $\mathbf{z}_t$ constitute the information set $\digamma_t$ defined as

$$\digamma_t \equiv \left(\mathbf{z}_t', \mathbf{x}_{1t}', \mathbf{x}_{2t}'\right)'. \tag{2}$$

The indicator function $\mathbf{I}(\cdot)$ is defined as

$$\mathbf{I}(A) = \begin{cases} 1 & \text{if } A > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

The vector of residuals $\mathbf{u}_t \equiv (u_{1t}, u_{2t})'$ is a vector of real-valued martingale difference sequence with respect to a non-decreasing sigma-field $\mathcal{F}_t$ to which $\mathbf{u}_t$ and $\digamma_{t+1}$ are adapted, and it is such that

$$E\left(\mathbf{u}_t \mathbf{u}_t'\right) = \boldsymbol{\Sigma}_{\mathbf{u}} \equiv \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 \end{pmatrix}. \tag{4}$$

The system in (1) is a two-equation nonlinear simultaneous equations model with endogenous dummy variables. The system is piecewise linear; following the threshold principle first advanced by Pearson (1900), a shift in the intercept in the equation for y_{it} occurs whenever the underlying endogenous variable y_{jt} crosses the corresponding threshold c_j, for $i, j = 1, 2$ and $i \neq j$. Therefore, the system belongs to the general class of models with structural shifts, as considered in Heckman (1978). Blundell and Smith (1994) nests the specifications in Heckman (1978) with those in Gourieroux *et al.* (1980), who consider models with endogenous switching regimes. A detailed survey of this class of models is provided in Chapter 5 of Maddala (1983). Notice that in models with structural shifts as considered in the microeconometric literature, the threshold parameters are assumed to be known (and generally equal to zero); conversely, in the system in (1), the thresholds c_1 and c_2 are allowed to be unknown: in this respect, the model we are considering is also related to the time series literature on threshold models, which became popular after the work by Tong (1990), as it extends univariate self-exciting threshold autoregressive (SETAR) models to a simultaneous equations system.

At each point in time, the variable y_{it} can be in one of two possible regimes, i.e. $y_{it} \leq c_i$ or $y_{it} > c_i$, $i = 1, 2$; therefore, the system as a whole can be in one of the following four regimes:

$$\begin{array}{ll} \text{regime 1:} \quad y_{1t} \leq c_1, y_{2t} \leq c_2; & \text{regime 2:} \quad y_{1t} \leq c_1, y_{2t} > c_2; \\ \text{regime 3:} \quad y_{1t} > c_1, y_{2t} \leq c_2; & \text{regime 4:} \quad y_{1t} > c_1, y_{2t} > c_2. \end{array} \tag{5}$$

Given the model specification in (1), the following two assumptions are imposed:

Assumption 1 The elements of the vector $\mathcal{F}_t$ are stationary ergodic predetermined variables.

Assumption 2 The vector $\mathbf{u}_t$ is distributed as $\mathbf{u}_t \sim IID\left(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u}}\right)$; its joint pdf is absolutely continuous, positive everywhere on $\mathbb{R}^2$, and independent of $\mathcal{F}_t$.

These assumptions are standard from the literature on threshold models; for example, see Petruccelli and Woolford (1984), Chan and Tong (1985), and Chan (1993). In particular, they rule out the possibility of conditionally heteroskedastic residuals; as a consequence, $(u_{1t}, u_{2t})'$ have constant conditional correlation coefficient $\rho_{u_1 u_2} \equiv (\sigma_{u_1 u_2}/\sigma_{u_1}\sigma_{u_2})$.

From an economic perspective, the model in (1) allows for both interdependence and contagion effects. At each point in time, interdependence is the result of normal markets interactions, and it is captured by the non-zero value of the correlation coefficient $\rho_{u_1 u_2}$. Conversely, contagion only takes place during crisis periods. Formally, we assume that a crisis in market i is associated with an *extreme positive value* of the dependent variable y_{it}, and that it takes place whenever y_{it} becomes strictly greater than the corresponding threshold parameter c_i: therefore, from (3), crisis periods are associated to a unit value of the indicator function $\mathbf{I}(\cdot)$. As a consequence, contagion from market j to market i is said to occur if

$$\Pr\left(y_{it} > c_i \,|y_{jt} > c_j, \mathbf{z}_t, \mathbf{x}_{it}\right) > \Pr\left(y_{it} > c_i \,|y_{jt} < c_j, \mathbf{z}_t, \mathbf{x}_{it}\right), \quad i,j = 1,2, \quad i \neq j : \tag{6}$$

from (6), a necessary condition for contagion to occur is that $\beta_i > 0$; therefore, contagion from market j to market i determines an endogenous shift in the value of the intercept from δ_i to $\delta_i + \beta_i$. Finally, because of the way crisis periods and contagion are defined, without loss of generality we assume that the condition $c_i > 0$ and $\beta_i \geq 0$, $i = 1, 2$ holds[3].

The dependent variable y_{it} in (1) is an index of market performance. Several examples for y_{it} arise from the contagion literature, depending upon the nature of the underlying markets, i.e. currency, stock or bond markets. In the context of currency crises, Eichengreen *et al.* (1996) and Esquivel and Larraín (1998) use the index of exchange market pressure introduced by Girton and Roper (1977). The index is a weighted average of exchange rate devaluation, increase in short term interest rate and decrease in international reserves, where the weights are chosen so to equalise the volatility of the three components[4]. In the framework of stock market crises, King and Wadhwani (1990), Boyer *et al.* (1999), Loretan and English (2000), Forbes and Rigobon (2002) and Corsetti *et al.* (2005) make use of stock market returns. Finally, Favero and Giavazzi (2002) study the propagation of devaluation expectations amongst member countries of the European Monetary System by deploying the spreads between the 3-months interest rate for member countries, related to the 3-months German rate.

Finally, it worth linking the model in (1) with those previously employed in the contagion literature. From (6), contagion is defined in a probabilistic way; however, the underlying model allows for endogenous contagion dummies and contemporaneously correlated errors: therefore, the shortcomings arising from the application of binary choice models and the Favero and Giavazzi (2002) approach are overcome. Further, in (1) neither are crisis periods *a priori* specified, nor are contemporaneous periods of turmoil treated as independent events: therefore, the problems of correlation-based tests (or their generalisations) and Markov switching models are respectively solved.

3 Solution

In order to solve the model in (1) two cases have to be considered. First, if β_1 or β_2 (or both) are equal to zero the system has a unique equilibrium. For example, if $\beta_2 = 0$ the system simplifies to

$$\begin{aligned} y_{1t} &= \boldsymbol{\delta}_1'\mathbf{z}_t + \boldsymbol{\alpha}_1'\mathbf{x}_{1t} + \beta_1 \mathbf{I}(y_{2t} - c_2) + u_{1t}, \\ y_{2t} &= \boldsymbol{\delta}_2'\mathbf{z}_t + \boldsymbol{\alpha}_2'\mathbf{x}_{2t} + u_{2t}, \end{aligned} \tag{7}$$

[3] The assumptions $c_1, c_2 > 0$ and $\beta_1, \beta_2 \geq 0$ are imposed because of the economic interpretation of the system in (1). In principle c_1 and c_2, and β_1 and β_2 could be interior points of the compact sets $\mathbf{C}$ and $\boldsymbol{\beta}$, respectively, such that $\mathbf{C} \subset \mathbb{R}^2$ and $\boldsymbol{\beta} \subset \mathbb{R}^2$. In particular, an analysis of the model for the remaining combinations of values of β_1 and β_2 is provided in Appendix A.

[4] The theoretical motivation of the index introduced in Girton and Roper (1977) can be found in the literature on first and second generation models of currency crises. See Blackburn and Sola (1993) and Rangvid (2001) for a survey of first and second generation models, respectively. Further, Kruger *et al.* (1998) and Stone and Weeks (2001) follow an analogous approach; however, in constructing the index of exchange market pressure they exclude interest rates, on the ground that they are not market-determined in developing countries.

and the solution is given by

$$\left\{\begin{array}{l} y_{1t} = \boldsymbol{\delta}_1'\mathbf{z}_t + \boldsymbol{\alpha}_1'\mathbf{x}_{1t} + u_{1t}, \\ y_{2t} = \boldsymbol{\delta}_2'\mathbf{z}_t + \boldsymbol{\alpha}_2'\mathbf{x}_{2t} + u_{2t}, \end{array}\right. \qquad \text{if } y_{2t} \leq c_2;$$

$$\left\{\begin{array}{l} y_{1t} = \boldsymbol{\delta}_1'\mathbf{z}_t + \boldsymbol{\alpha}_1'\mathbf{x}_{1t} + \beta_1 + u_{1t}, \\ y_{2t} = \boldsymbol{\delta}_2'\mathbf{z}_t + \boldsymbol{\alpha}_2'\mathbf{x}_{2t} + u_{2t}, \end{array}\right. \qquad \text{if } y_{2t} > c_2.$$

In order to solve the system when both β_1 and β_2 are positive, define the variable w_{it} as

$$w_{it} \equiv \boldsymbol{\delta}_i'\mathbf{z}_t + \boldsymbol{\alpha}_i'\mathbf{x}_{it} + u_{it}, \quad i = 1, 2, \tag{8}$$

so that (1) can be equivalently written as

$$\begin{array}{c} y_{1t} = w_{1t} + \beta_1 \mathbf{I}\left(y_{2t} - c_2\right), \\ y_{2t} = w_{2t} + \beta_2 \mathbf{I}\left(y_{1t} - c_1\right). \end{array} \tag{9}$$

Because of the condition $\beta_1, \beta_2 > 0$, the following normalised variables can be defined

$$Y_{it} \equiv \frac{y_{it} - c_i}{\beta_i}, \quad W_{it} \equiv \frac{w_{it} - c_i}{\beta_i}, \quad i = 1, 2; \tag{10}$$

the system in (9) can then be written as

$$\begin{array}{c} Y_{1t} = W_{1t} + \mathbf{I}\left(Y_{2t}\right), \\ Y_{2t} = W_{2t} + \mathbf{I}\left(Y_{1t}\right). \end{array} \tag{11}$$

From (10), the four possible regimes in (5) are defined as

$$\begin{array}{ll} \text{regime 1:} \quad Y_{1t} \leq 0, Y_{2t} \leq 0; & \text{regime 2:} \quad Y_{1t} \leq 0, Y_{2t} > 0; \\ \text{regime 3:} \quad Y_{1t} > 0, Y_{2t} \leq 0; & \text{regime 4:} \quad Y_{1t} > 0, Y_{2t} > 0. \end{array} \tag{12}$$

In (12), the case $Y_{it} > 0$ corresponds to a crisis in market i at time t, whereas $Y_{it} \leq 0$ denotes a tranquil period. From (11), each of the regimes in (12) corresponds to a region in the (W_{1t}, W_{2t}) space:

$$\begin{array}{ll} \text{regime 1:} \quad W_{1t} \leq 0, W_{2t} \leq 0; & \text{regime 2:} \quad W_{1t} \leq -1, W_{2t} > 0. \\ \text{regime 3:} \quad W_{1t} > 0, W_{2t} \leq -1; & \text{regime 4:} \quad W_{1t} > -1, W_{2t} > -1. \end{array} \tag{13}$$

The four combinations of values of W_{1t} and W_{2t} described in (13) give rise to the following mutually exclusive stochastic solution regions in the (W_{1t}, W_{2t}) space:

$$\begin{array}{ll} \text{region A}_t\text{:} & \left\{\begin{array}{l} W_{1t} > -1 \\ W_{2t} > 0 \end{array}\right. \cup \left\{\begin{array}{l} W_{1t} > 0 \\ -1 < W_{2t} \leq 0 \end{array}\right. ; \\ \text{region B}_t\text{:} & \left\{\begin{array}{l} W_{1t} > 0 \\ W_{2t} \leq -1 \end{array}\right. ; \\ \text{region C}_t\text{:} & \left\{\begin{array}{l} W_{1t} \leq -1 \\ -1 < W_{2t} \leq 0 \end{array}\right. \cup \left\{\begin{array}{l} W_{1t} \leq 0 \\ W_{2t} \leq -1 \end{array}\right. ; \\ \text{region D}_t\text{:} & \left\{\begin{array}{l} W_{1t} \leq -1, \\ W_{2t} > 0, \end{array}\right. ; \\ \text{region E}_t\text{:} & \left\{\begin{array}{l} -1 < W_{1t} \leq 0 \\ -1 < W_{2t} \leq 0 \end{array}\right. . \end{array} \tag{14}$$

A graphical representation of the solution regions defined in (14) is given in Figure 1 below:

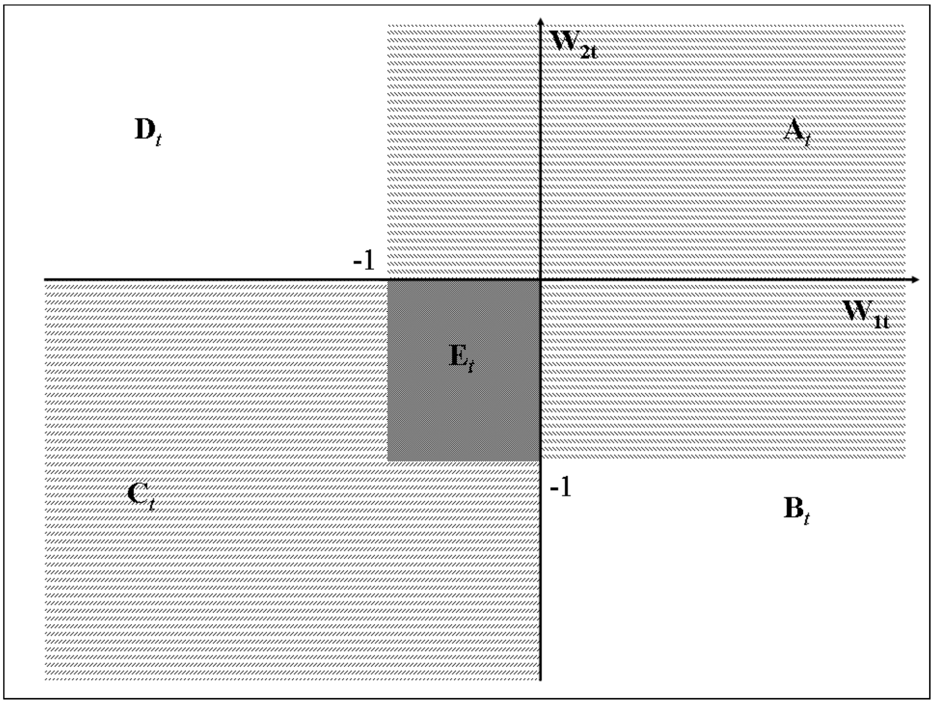

Figure 1: Solution regions.

In each of the regions defined in (14), a corresponding random event occurs. Therefore, in terms of the normalised variables Y_{it} and W_{it} defined in (10), the complete solution to the model can be written as:

$$\begin{aligned}
&\begin{cases} Y_{1t} = W_{1t} + 1 > 0 \\ Y_{2t} = W_{2t} + 1 > 0 \end{cases} ; && \text{(event A}_t\text{)} \\
&\begin{cases} Y_{1t} = W_{1t} > 0 \\ Y_{2t} = W_{2t} + 1 \leq 0 \end{cases} ; && \text{(event B}_t\text{)} \\
&\begin{cases} Y_{1t} = W_{1t} \leq 0 \\ Y_{2t} = W_{2t} \leq 0 \end{cases} ; && \text{(event C}_t\text{)} \\
&\begin{cases} Y_{1t} = W_{1t} + 1 \leq 0 \\ Y_{2t} = W_{2t} > 0 \end{cases} ; && \text{(event D}_t\text{)} \\
&\begin{cases} Y_{1t} = W_{1t} \leq 0 \\ Y_{2t} = W_{2t} \leq 0 \end{cases} \cup \begin{cases} Y_{1t} = W_{1t} + 1 > 0 \\ Y_{2t} = W_{2t} + 1 > 0 \end{cases} . && \text{(event E}_t\text{)}
\end{aligned} \tag{15}$$

From (15), the model has a unique solution in regions A_t, B_t, C_t, and D_t, but *multiple equilibria* arise in region E_t. From (12) and (13), this happens because the regions in the (W_{1t}, W_{2t}) plane where the regimes $Y_{it} \leq 0$ and $Y_{it} > 0$ occurs are not mutually exclusive, and their intersection generates region E_t. Notice that the presence of multiple equilibria in region E_t is consistent with the notion of contagion put forward in Masson (1999) and previously discussed.

Pesaran and Pick (2007) model the solutions in region E_t as the outcome of a randomisation process d_t^E defined as

$$d_t^E \sim Bernoulli\left(\pi_d^E\right), \tag{16}$$

where π_d^E is the *unknown* probability of observing $d_t^E = 1$. If π_d^E is defined as the probability of observing the favourable non-crisis equilibrium $Y_{it} = W_{it}$, that is

$$\pi_d^E \equiv \Pr\left(Y_{it} = W_{it} \,|E_t;\ \digamma_t\right), \tag{17}$$

then from (15) the solution in region E_t can be written as

$$\begin{aligned} Y_{it} &= d_t^E W_{it} + \left(1 - d_t^E\right)(1 + W_{it}) \\ &= 1 + W_{it} - d_t^E. \end{aligned} \tag{18}$$

From (17) the probability π_d^E is independent of the information set $\digamma_t$; in principle π_d^E could be allowed to depend upon the observable explanatory variables: however, this extension goes beyond the purpose of the analysis we are now undertaking. Taking into account (8), (10), (15), (16) and (18), the reduced form for the original system in (1) is given by

$$\begin{aligned}
&\begin{cases} y_{1t} = w_{1t} + \beta_1 > c_1 \\ y_{2t} = w_{2t} + \beta_2 > c_2 \end{cases} ; \qquad (\text{event } \mathrm{A}_t) \\
&\begin{cases} y_{1t} = w_{1t} > c_1 \\ y_{2t} = w_{2t} + \beta_2 \leq c_2 \end{cases} ; \qquad (\text{event } \mathrm{B}_t) \\
&\begin{cases} y_{1t} = w_{1t} \leq c_1 \\ y_{2t} = w_{2t} \leq c_2 \end{cases} ; \qquad (\text{event } \mathrm{C}_t) \\
&\begin{cases} y_{1t} = w_{1t} + \beta_1 \leq c_1 \\ y_{2t} = w_{2t} > c_2 \end{cases} ; \qquad (\text{event } \mathrm{D}_t) \\
&\begin{cases} y_{1t} = w_{1t} + \left(1 - d_t^E\right)\beta_1 \\ y_{2t} = w_{2t} + \left(1 - d_t^E\right)\beta_2 \end{cases} . \qquad (\text{event } \mathrm{E}_t)
\end{aligned} \tag{19}$$

4 Coherency

From (5), the model in (1) describes four possible regimes. The probabilities of the occurrence of *each* of these regimes are equal to

$$\begin{aligned}
\Pr\left(y_{1t} \leq c_1, y_{2t} \leq c_2 \left| \digamma_t \right.\right) &= \Pr\left(C_t \cup E_t \left| \digamma_t \right.\right), \\
\Pr\left(y_{1t} \leq c_1, y_{2t} > c_2 \left| \digamma_t \right.\right) &= \Pr\left(D_t \left| \digamma_t \right.\right), \\
\Pr\left(y_{1t} > c_1, y_{2t} \leq c_2 \left| \digamma_t \right.\right) &= \Pr\left(B_t \left| \digamma_t \right.\right), \\
\Pr\left(y_{1t} > c_1, y_{2t} > c_2 \left| \digamma_t \right.\right) &= \Pr\left(A_t \cup E_t \left| \digamma_t \right.\right),
\end{aligned} \tag{20}$$

where the information set $\digamma_t$ is defined in (2). Denote by p_t the probability of the occurrence of *any* of the regimes in (5); from (20) p_t satisfies

$$\begin{aligned}
p_t &= \Pr\left(y_{1t} \leq c_1, y_{2t} \leq c_2 \left| \digamma_t \right.\right) + \Pr\left(y_{1t} \leq c_1, y_{2t} > c_2 \left| \digamma_t \right.\right) \\
&\quad + \Pr\left(y_{1t} > c_1, y_{2t} \leq c_2 \left| \digamma_t \right.\right) + \Pr\left(y_{1t} > c_1, y_{2t} > c_2 \left| \digamma_t \right.\right) \\
&= 1 + \Pr\left(E_t \left| \digamma_t \right.\right) \\
&> 1,
\end{aligned} \tag{21}$$

where the analytical expression for $\Pr\left(E_t \left| \digamma_t \right.\right)$ is given by

$$\begin{aligned}
\Pr\left(E_t \left| \digamma_t \right.\right) &= G\left(\frac{c_1 - \boldsymbol{\delta}_1'\mathbf{z}_t - \boldsymbol{\alpha}_1'\mathbf{x}_{1t}}{\sigma_{u_1}}, \frac{c_2 - \boldsymbol{\delta}_2'\mathbf{z}_t - \boldsymbol{\alpha}_2'\mathbf{x}_{2t}}{\sigma_{u_2}}\right) \\
&\quad - G\left(\frac{c_1 - \boldsymbol{\delta}_1'\mathbf{z}_t - \boldsymbol{\alpha}_1'\mathbf{x}_{1t}}{\sigma_{u_1}}, \frac{c_2 - \boldsymbol{\delta}_2'\mathbf{z}_t - \boldsymbol{\alpha}_2'\mathbf{x}_{2t} - \beta_2}{\sigma_{u_2}}\right) \\
&\quad - G\left(\frac{c_1 - \boldsymbol{\delta}_1'\mathbf{z}_t - \boldsymbol{\alpha}_1'\mathbf{x}_{1t} - \beta_1}{\sigma_{u_1}}, \frac{c_2 - \boldsymbol{\delta}_2'\mathbf{z}_t - \boldsymbol{\alpha}_2'\mathbf{x}_{2t}}{\sigma_{u_2}}\right) \\
&\quad + G\left(\frac{c_1 - \boldsymbol{\delta}_1'\mathbf{z}_t - \boldsymbol{\alpha}_1'\mathbf{x}_{1t} - \beta_1}{\sigma_{u_1}}, \frac{c_2 - \boldsymbol{\delta}_2'\mathbf{z}_t - \boldsymbol{\alpha}_2'\mathbf{x}_{2t} - \beta_2}{\sigma_{u_2}}\right),
\end{aligned} \tag{22}$$

and $G(\cdot)$ is the joint pdf of the standardised error terms $\left(u_{1t} \backslash \sigma_{u_1}, u_{2t} \backslash \sigma_{u_2}\right)'$. From (21), we can see that *the sum of the probabilities of the four events in* (5) *is greater than unity.* This feature is a direct consequence of the presence of multiple equilibria in region E_t: for a given set of economic fundamentals more than one equilibrium exists; therefore, conditional upon those fundamentals, the probability of the occurrence of any of the equilibria is greater than one.

The system in (1) is an example of an *incoherent econometric model*. Gourieroux *et al.* (1980) define a coherent model as one with a "well defined reduced form". This is equivalent to saying there exists a one-to-one correspondence between a shock u_{it} and the related dependent variable y_{it} for given values of the explanatory variables. This is not the case for the model in (1) due to the randomisation process d_t^E defined in (16), which chooses the outcome in the multiple equilibria region E_t. This is also the technical explanation behind the result obtained in (21). Incoherent models have been widely studied in the microeconometric literature, as discussed in Chapter 5 of Maddala (1983).

For identification and estimation purposes, the well known *coherency condition* has often been imposed, as discussed in Heckman (1978), Gourieroux *et al.* (1980), Blundell and Smith (1994), and Lewbel (2007). The coherency condition is necessary and sufficient to guarantee that the probabilities of the four events in (5) add up to one and to avoid the presence of multiple equilibria; in this way, the existence of a well-defined likelihood function for the dependent variables $(y_{1t}, y_{2t})'$ is ensured. In the context of the model in (1), this means imposing the condition $\beta_1\beta_2 = 0$: for example, if $\beta_2 = 0$ the system reduces to that in (7). However, the imposition of the coherency condition *a priori* eliminates simultaneity, the key feature the model tries to capture. Therefore, identification and estimation of the system have to be achieved without imposing the coherency condition.

Some work in this direction has already been done in the literature. For example, Kooreman (1994) considers the model in (1) where both the dependent variables are discrete rather than continuous, and the threshold parameters are assumed to me known; the model of interest then becomes a simultaneous probit, and it is estimated by Maximum Likelihood by taking the probability π_d^E in the randomisation process d_t^E defined in (16) as known: however, this simplifying assumption may lead to inconsistent estimates of the parameters in case the assumed value of π_d^E turns out to be different from the true one. Bresnahan and Reiss (1990) consider a simultaneous probit model as in Kooreman (1994); however, they avoid considering the probability π_d^E as known by treating the multiple outcomes as one event: in this way the model effectively predicts three outcomes rather than four, a loss of information will be suffered, and the resulting estimator will be inefficient. Tamer (2003) solves the efficiency problem of the estimator proposed in Bresnahan and Reiss (1990) by treating the multiple equilibria outcomes as two separate events; this is achieved by estimating the model by semiparametric Maximum Likelihood, where the probability of one of the incoherent outcomes is replaced by its sample counterpart.

In this paper we extend the work by Tamer (2003) by achieving identification and estimation of the system in (1), where both the dependent variables are continuous rather than discrete, the threshold parameters are treated as unknown, and no coherency condition is imposed. As we are going to prove, identification is achieved by exploiting the set of moment restrictions arising from the multiplicity of regimes described in (5). Further, estimation is performed by parametric (rather than semiparametric) Maximum Likelihood, provided that the joint pdf of $(y_{1t}, y_{2t})'$ contains a normalisation factor which ensures that the function itself integrates to unity, so to account for the incoherent nature of the system.

5 Identification

The model in (1) is a simultaneous equations system, so it is crucial to determine under what conditions contagion effects (captured by the slope coefficients β_1 and β_2) can be identified from interdependence (captured by the correlation coefficient $\rho_{u_1u_2}$). The identification problem can be solved by exploiting the presence of the multiple regimes described in (5); this can be done so long as the variable y_{it} has *at least* one observation in each of the two regimes $y_{it} \leq c_i$ and $y_{it} > c_i$, for $i = 1, 2$. If only the regime $y_{it} \leq c_i$ is observed then no observation is available to identify β_i; conversely, if only the regime $y_{it} > c_i$ occurs then the right-hand side variables become perfectly collinear.

Formally, denote by $\hat{\pi}_{iT}(c_i)$ the *proportion* of crisis in market i associated to a general value of the threshold parameter c_i; we then require $\hat{\pi}_{iT}(c_i)$ be neither zero nor one, that is

$$0 < \hat{\pi}_{iT}(c_i) < 1, \quad \hat{\pi}_{iT}(c_i) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{I}(y_{it} - c_i), \quad i = 1, 2, \tag{23}$$

uniformly for all sample sizes T. As $T \to \infty$, condition (23) requires the *probability* of crisis in market i associated to c_i, denoted by $\pi_i(c_i)$, be neither zero nor one, for $i = 1, 2$. Formally, under the conditions imposed in Assumption 1 and Assumption 2, this means that

$$0 < \pi_i(c_i) < 1, \quad \hat{\pi}_{iT}(c_i) \overset{a.s.}{\to} \pi_i(c_i) \equiv E[\mathbf{I}(y_{it} - c_i)], \quad i = 1, 2, \tag{24}$$

where $\hat{\pi}_{iT}(c_i)$ is defined in (23). As proved in Pesaran and Pick (2007), if the error terms $(u_{1t}, u_{2t})'$ in (1) have joint pdf positive everywhere on $\mathbb{R}^2$ (as imposed in Assumption 2), then condition (24) is satisfied.

Condition (24) is sufficient to ensure that at least *two* of the regimes in (5) actually occur; without loss of generality, we consider the case where these regimes are the following:

$$\text{regime 1:} \quad y_{1t} \le c_1, y_{2t} \le c_2; \qquad \text{regime 4:} \quad y_{1t} > c_1, y_{2t} > c_2.$$

This means that the following two conditions hold:

$$0 < \Pr(y_{1t} \le c_1, y_{2t} \le c_2) < 1, \qquad 0 < \Pr(y_{1t} > c_1, y_{2t} > c_2) < 1,$$

where

$$\frac{1}{T}\sum_{t=1}^{T}[1 - \mathbf{I}(y_{1t} - c_1)][1 - \mathbf{I}(y_{2t} - c_2)] \overset{a.s.}{\to} \Pr(y_{1t} \le c_1, y_{2t} \le c_2)$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{I}(y_{1t} - c_1)\mathbf{I}(y_{2t} - c_2) \overset{a.s.}{\to} \Pr(y_{1t} > c_1, y_{2t} > c_2).$$

The identification conditions for the system in (1) are then stated in the following theorem:

Theorem 1 *Consider the model in* (1), *where the vectors of predetermined variables* $\mathbf{x}_{1t}$ *and* $\mathbf{x}_{2t}$ *are such that* $\mathbf{x}_{1t} \cap \mathbf{x}_{2t} = \varnothing$, *and where the condition stated in Assumption* 1 *and Assumption* 2 *hold. Then if* $\beta_1, \beta_2 > 0$ *the model is identified even if* $\boldsymbol{\alpha}_1 = \mathbf{0}$ *and* $\boldsymbol{\alpha}_2 = \mathbf{0}$.

Proof. Consider the system in (1), where no exclusion restrictions are imposed ($\boldsymbol{\alpha}_1 = \mathbf{0}$ and $\boldsymbol{\alpha}_2 = \mathbf{0}$), and without loss of generality suppose that $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \mathbf{0}$. The model then simplifies to

$$\begin{aligned} y_{1t} &= \beta_1 \mathbf{I}(y_{2t} - c_2) + u_{1t}, \\ y_{2t} &= \beta_2 \mathbf{I}(y_{1t} - c_1) + u_{2t}. \end{aligned} \tag{25}$$

Condition (24) is sufficient to identify the threshold parameters c_1 and c_2. This is because given $\Pr(y_{1t} \le c_1, y_{2t} \le c_2)$ and two vectors of values for $(c_1, c_2)'$, namely $(c_1^*, c_2^*)'$ and $(c_1^{**}, c_2^{**})'$ such that $(c_1^*, c_2^*)' \neq (c_1^{**}, c_2^{**})'$, then

$$\Pr(y_{1t} \le c_1^*, y_{2t} \le c_2^*) \neq \Pr(y_{1t} \le c_1^{**}, y_{2t} \le c_2^{**}), \quad a.s.$$

An analogous argument holds for $\Pr(y_{1t} > c_1, y_{2t} > c_2)$, and c_1 and c_2 are identified. The system in (25) is therefore observationally equivalent to

$$\begin{aligned} &\begin{cases} y_{1t} = u_{1t}, \\ y_{2t} = u_{2t}, \end{cases} && \text{if } y_{1t} \le c_1, y_{2t} \le c_2, \\ &\begin{cases} y_{1t} = \beta_1 + u_{1t}, \\ y_{2t} = \beta_2 + u_{2t}, \end{cases} && \text{if } y_{1t} > c_1, y_{2t} > c_2, \end{aligned} \tag{26}$$

and five parameters have to be identified, i.e. β_1, $\sigma_{u_1}^2$, β_2, $\sigma_{u_2}^2$ and $\sigma_{u_1u_2}$. From (26), the following moment restrictions are then available:

$$E(y_{1t}^2) = \sigma_{u_1}^2, \quad E(y_{2t}^2) = \sigma_{u_2}^2, \quad E(y_{1t}y_{2t}) = \sigma_{u_1u_2},$$

$$\text{if } y_{1t} \le c_1, y_{2t} \le c_2;$$

and

$$E\left[\left(y_{1t}-\beta_1\right)^2\right]=\sigma^2_{u_1},\quad E\left[\left(y_{2t}-\beta_2\right)^2\right]=\sigma^2_{u_2},\quad E\left[\left(y_{1t}-\beta_1\right)\left(y_{2t}-\beta_2\right)\right]=\sigma_{u_1u_2},$$

$$\text{if } y_{1t}>c_1, y_{2t}>c_2.$$

Therefore, six distinct moment restrictions are available to identify five parameters, and the system in (25) is overidentified. ■

Remark 2 *Theorem* 1 *assumes that* $\beta_1, \beta_2 > 0$*. Clearly, the result stated in the theorem holds also when* $\beta_1\beta_2 = 0$*, i.e. when at least one of the contagion coefficients is equal to zero.*

Theorem 1 shows that the system in (1) is identified even if equation specific explanatory variables are not included: the multiplicity of regimes provides the additional number of moment restrictions which is sufficient to identify the system. Notice that while the condition stated in (24) (and therefore the conditions imposed in Assumption 1 and Assumption 2) is sufficient to guarantee that two of the regimes in (5) occur, other regimes may actually take place; this means that an additional number of moment restrictions may be available. Further, it is important to stress that although the reduced form in (19) depends upon the randomisation process d_t^E defined in (16), identification of the probability of success π_d^E in (17) is not required to identify the parameters of the system in (1). Finally, notice that the conditions for identification required in Theorem 1 are weaker than those imposed in Pesaran and Pick (2007): under the assumption of c_1 and c_2 being known, they identify the system by including equation specific explanatory variables, i.e. by imposing the condition $\boldsymbol{\alpha}_1 \neq \mathbf{0}$ and $\boldsymbol{\alpha}_2 \neq \mathbf{0}$.

It is also interesting to compare the conditions required to identify the system in (1) with those needed in the simultaneous probit model as obtained in Tamer (2003). In particular, in the latter model the slope coefficients are identified up to a scale factor, provided that equation specific explanatory variables are included. Conversely, Theorem 1 shows that identification of the system in (1) does not require imposition of any exclusion restrictions, and that the slope coefficients are uniquely identified: this is because the continuous dependent variables $(y_{1t}, y_{2t})'$ are *observable* rather than latent. Finally, notice that a further consequence of $(y_{1t}, y_{2t})'$ being observable is that the threshold parameters c_1 and c_2 are identified; conversely, in the (simultaneous) probit model, the threshold parameters are not identified: for this reason, they are assumed to be known and equal to zero.

6 Estimation

Single equation OLS estimation of the system in (1) would deliver inconsistent estimates of the parameters due to the endogeneity of the contagion dummies. Pesaran and Pick (2007) propose to estimate the model by employing a single equation GIVE estimator. Under the assumption of the conditional distribution of the shocks $(u_{1t}, u_{2t})'$ being known, the system can be estimated by Full Information Maximum Likelihood (FIML): the first estimation method is discussed in Section 6.1, the second is introduced in Section 6.2.

6.1 Single equation GIVE estimation

Estimation of threshold models by instrumental variables is addressed in Caner and Hansen (2004). They consider a single equation model where the explanatory variables are endogenous while the threshold variable is exogenous; they then propose a two-stage least square estimator of the threshold parameter and a generalised method of moments estimator of the slope coefficients. However, this estimation procedure cannot be applied to the system in (1), as the explanatory variables are predetermined and the threshold variables are endogenous.

Pesaran and Pick (2007) assume that the threshold parameters c_1 and c_2 are known and employ a single equation GIVE estimator to estimate the remaining set of parameters. For $i, j = 1, 2$ with $i \neq j$, define the vectors

$$\boldsymbol{\phi}_i \equiv \left(\boldsymbol{\delta}_i', \boldsymbol{\alpha}_i', \beta_i\right)', \quad \mathbf{y}_i \equiv \left(y_{i1}, \ldots, y_{iT}\right)', \quad \mathbf{h}_{it} \equiv \left[\mathbf{z}_t', \mathbf{x}_{it}', \mathbf{I}\left(y_{jt}-c_j\right)\right]', \quad \mathbf{u}_i \equiv \left(u_{i1}, \ldots, u_{iT}\right)', \qquad (27)$$

and the matrices

$$\mathbf{H}_i \equiv \begin{pmatrix} \mathbf{h}'_{i1} \\ \vdots \\ \mathbf{h}'_{iT} \end{pmatrix}, \quad \mathbf{W}_i \equiv \begin{pmatrix} \mathbf{w}'_{i1} \\ \vdots \\ \mathbf{w}'_{iT} \end{pmatrix}, \quad \mathbf{P}_{\mathbf{W}_i} \equiv \mathbf{W}_i \left(\mathbf{W}'_i\mathbf{W}_i\right)^{-1} \mathbf{W}'_i,$$

$\mathbf{w}_{it}$ being the vector of instruments: the system in (1) can then be cast in matrix form as

$$\mathbf{y}_i = \mathbf{H}_i\boldsymbol{\phi}_i + \mathbf{u}_i, \quad i,j = 1,2.$$

The GIVE estimator is then obtained as

$$\hat{\boldsymbol{\phi}}_{i,GIVE} = \left(\mathbf{H}'_i\mathbf{P}_{\mathbf{W}_i}\mathbf{H}_i\right)^{-1}\left(\mathbf{H}'_i\mathbf{P}_{\mathbf{W}_i}\mathbf{y}_i\right), \tag{28}$$

with estimated covariance matrix

$$\hat{\mathbf{V}}_{i,GIVE} = \hat{\sigma}_i^2\left(\mathbf{H}'_i\mathbf{P}_{\mathbf{W}_i}\mathbf{H}_i\right)^{-1},$$

where

$$\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{H}_i\hat{\boldsymbol{\phi}}_{i,GIVE}, \qquad \hat{\sigma}_i^2 = \left(\hat{\mathbf{u}}'_i\hat{\mathbf{u}}_i\right)/T.$$

Since the model is nonlinear, an important issue is the choice of the optimal vector of instruments, denoted by $\mathbf{w}^*_{it}$. For given values of c_1 and c_2, the equations of the system in (1) are linear in the parameters, but contain regressors that are nonlinear functions of the endogenous variables; following Kelejian (1971) and Bowden and Turkington (1981), in this framework the optimal instrument for the endogenous contagion dummy $\mathbf{I}(y_{it} - c_i)$ would be[5]

$$w^*_{it} \equiv E\left[\mathbf{I}(y_{it} - c_i)\,|\mathcal{F}_t\right] = \Pr\left(y_{it} - c_i > 0\,|\mathcal{F}_t\right), \tag{29}$$

which is the conditional probability of crisis in market i, whose analytical expression is obtained in Appendix B[6].

However, the optimal instrument w^*_{it} in (29) is not feasible: although the analytical expression is available in closed form once the joint pdf of $(u_{1t}, u_{2t})'$ in (1) is known, it is a function of the vectors of unknown parameters $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. However, following the parametric approach proposed in Kelejian (1971), the endogenous dummy variable $\mathbf{I}(y_{it} - c_i)$ can be approximated by a polynomial of order m in the corresponding equation specific predetermined variables $\mathbf{x}_{it}$[7]. The vector of instruments for the system in (1) is then given by

$$\mathbf{w}_{it} = \left[\mathbf{z}'_t, \mathbf{x}'_{it}, \mathbf{x}'_{jt}, \left(\mathbf{x}^2_{jt}\right)', \ldots, \left(\mathbf{x}^m_{jt}\right)'\right]', \quad i,j = 1,2, \quad i \neq j, \quad t = 1,\ldots,T, \tag{30}$$

where $\mathbf{x}^n_{jt}$ denotes the column vector made of the $n-th$ powers of each of the elements of the vector $\mathbf{x}_{jt}$. Pesaran and Pick (2007) approximate the endogenous crisis indicators by a polynomial in the predetermined variables of order up to $m = 6$.

Estimation of the model in (1) by instrumental variables has the advantage of not being affected by the coherency issue discussed in Section 4. However, the GIVE estimator in (28) faces the following three problems:

1. it is a single equation estimator; therefore, efficiency issues arise.

2. it is likely to suffer from a weak instruments problem, as there might be a weak degree of correlation between the endogenous dummy variables and their instruments; see Stock *et al.* (2002) and Andrews and Stock (2005) for survey of the weak instruments problem. In this case, the GIVE estimator does not have an asymptotically normal distribution, and standard statistical inference provides misleading results.

[5] Notice that the optimal vector of instruments obtained in Kelejian (1971) and Bowden and Turkington (1981) is more generally valid for a system where the structural equations are linear in the parameters and contain regressors which are nonlinear functions of both the endogenous and the predetermined variables.

[6] As shown in Amemiya (1977), in a general framework the optimal vector of instruments $\mathbf{w}^*_{it}$ is given by the conditional expectation of the gradient calculated with respect to the vector of parameters and evaluated at the *true* parameters values.

[7] For nonparametric estimates of optimal instruments see Newey (1990).

3. the instruments $\mathbf{w}_{it}$ defined in (30) require inclusion of equation specific explanatory variables. However, this condition is not needed for identification purposes, as shown in Theorem 1.

The GIVE estimator in (28) assumes that the threshold parameters c_1 and c_2 are known; if this is not the case, then the threshold parameters can be estimated by grid search, and the remaining set of parameters by instrumental variables. Define the vector $\mathbf{h}_{it}(c_j)$ and the matrix $\mathbf{H}_i(c_j)$ as

$$\mathbf{h}_{it}(c_j) \equiv \left[\mathbf{z}_t', \mathbf{x}_{it}', \mathbf{I}(y_{jt} - c_j)\right]', \quad \mathbf{H}_i(c_j) \equiv \begin{bmatrix} \mathbf{h}_{i1}'(c_j) \\ \vdots \\ \mathbf{h}_{iT}'(c_j) \end{bmatrix}, \quad i, j = 1, 2, \quad i \neq j. \tag{31}$$

Further, suppose that c_j is an interior point of the compact set $\mathbf{C}_j$ (the set of admissible values of c_j) which satisfies $\mathbf{C}_j \subset \mathbb{R}$: the threshold c_j in the $i - th$ equation can be estimated by grid search as

$$\hat{c}_{j,GIVE} = \arg\min_{c_j \in \mathbf{C}_j} \left[\hat{\mathbf{u}}_i(c_j)\right]' \mathbf{P}_{\mathbf{W}_i} \left[\hat{\mathbf{u}}_i(c_j)\right], \quad i, j = 1, 2, \quad i \neq j,$$

where

$$\hat{\mathbf{u}}_i(c_j) = \mathbf{y}_i - \mathbf{H}_i(c_j)\, \hat{\boldsymbol{\phi}}_{i,GIVE}(c_j),$$

and

$$\hat{\boldsymbol{\phi}}_{i,GIVE}(c_j) = \left[\mathbf{H}_i(c_j)' \mathbf{P}_{\mathbf{W}_i} \mathbf{H}_i(c_j)\right]^{-1} \left[\mathbf{H}_i(c_j)' \mathbf{P}_{\mathbf{W}_i} \mathbf{y}_i\right].$$

The GIVE estimator $\hat{\boldsymbol{\phi}}_{i,GIVE}(c_j)$ is then obtained as

$$\hat{\boldsymbol{\phi}}_i(\hat{c}_{j,GIVE}) = \left[\mathbf{H}_i(\hat{c}_{j,GIVE})' \mathbf{P}_{\mathbf{W}_i} \mathbf{H}_i(\hat{c}_{j,GIVE})\right]^{-1} \left[\mathbf{H}_i(\hat{c}_{j,GIVE})' \mathbf{P}_{\mathbf{W}_i} \mathbf{y}_i\right], \quad i, j = 1, 2, \quad i \neq j.$$

Denote by c_j^0 and $\boldsymbol{\phi}_i^0$ the true values of the parameters c_j and $\boldsymbol{\phi}_i$, respectively. The strong consistency of $\hat{c}_{j,GIVE}$ and $\hat{\boldsymbol{\phi}}_i(\hat{c}_{j,GIVE})$ is then stated in the following theorem:

Theorem 3 *Consider the model in* (1) *under the set of conditions imposed in Assumption* 1 *and Assumption* 2. *Further, assume that* $E|\mathbf{w}_{it}| < \infty$, $E|u_{it}| < \infty$ *and* $E(\mathbf{w}_{it}\mathbf{w}_{it}')$ *is positive definite, for* $i = 1, 2$. *Then* $\hat{c}_{j,GIVE} \overset{a.s.}{\to} c_j^0$ *and* $\hat{\boldsymbol{\phi}}_i(\hat{c}_{j,GIVE}) \overset{a.s.}{\to} \boldsymbol{\phi}_i^0$, *for* $i, j = 1, 2$, $i \neq j$.

Proof. See Appendix C.1. ■

6.2 Maximum Likelihood estimation

In a full information framework, the threshold parameters c_1 and c_2 in (1) can be estimated by grid search, while the remaining set of parameters can be estimated by Full Information Maximum Likelihood (FIML). As it will be shown, the joint pdf of $(y_{1t}, y_{2t})'$ is piecewise due to the presence of multiple regimes in the system. In addition, it has to take into account the coherency issue discussed in Section 4: as such, it includes a normalisation factor p_t to ensure that it integrates to unity over $\mathbb{R}^2$.

Formally, the conditional joint pdf of $(y_{1t}, y_{2t})'$ for the model in (1) is given by

$$\begin{aligned} f(y_{1t}, y_{2t} | \mathcal{F}_t) \;=\; & \frac{1}{p_t} f(y_{1t}, y_{2t} | y_{1t} \leq c_1, y_{2t} \leq c_2; \mathcal{F}_t) \Pr(y_{1t} \leq c_1, y_{2t} \leq c_2 | \mathcal{F}_t) \\ & + \frac{1}{p_t} f(y_{1t}, y_{2t} | y_{1t} \leq c_1, y_{2t} > c_2; \mathcal{F}_t) \Pr(y_{1t} \leq c_1, y_{2t} > c_2 | \mathcal{F}_t) \\ & + \frac{1}{p_t} f(y_{1t}, y_{2t} | y_{1t} > c_1, y_{2t} \leq c_2; \mathcal{F}_t) \Pr(y_{1t} > c_1, y_{2t} \leq c_2 | \mathcal{F}_t) \\ & + \frac{1}{p_t} f(y_{1t}, y_{2t} | y_{1t} > c_1, y_{2t} > c_2; \mathcal{F}_t) \Pr(y_{1t} > c_1, y_{2t} > c_2 | \mathcal{F}_t), \end{aligned}$$

where the information set $\mathcal{F}_t$ is defined in (2). Define the following four vectors of parameters

$$\begin{aligned} \boldsymbol{\theta}_1 &\equiv \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2'\right)', & \boldsymbol{\theta}_2 &\equiv \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \beta_1, \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2'\right)', \\ \boldsymbol{\theta}_3 &\equiv \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2', \beta_2\right)', & \boldsymbol{\theta}_4 &\equiv \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \beta_1, \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2', \beta_2\right)', \end{aligned} \tag{32}$$

where $\boldsymbol{\theta}_k$ denotes the vector of slope coefficients characterising the joint pdf of $(y_{1t}, y_{2t})'$ in regime k, for $k = 1, \ldots, 4$, where each regime is defined as in (5). From the properties of the truncated distributions we have

$$f(y_{1t}, y_{2t} | y_{1t} \leq c_1, y_{2t} \leq c_2; \mathcal{F}_t) = \frac{[1 - \mathbf{I}(y_{1t} - c_1)][1 - \mathbf{I}(y_{2t} - c_2)] f(y_{1t}, y_{2t}; \boldsymbol{\theta}_1 | \mathcal{F}_t)}{\Pr(y_{1t} \leq c_1, y_{2t} \leq c_2 | \mathcal{F}_t)},$$

where $f(y_{1t}, y_{2t}; \boldsymbol{\theta}_1 | \mathcal{F}_t)$ denotes the joint pdf of $(y_{1t}, y_{2t})'$ under the regime characterised by $\boldsymbol{\theta}_1$; the other components of $f(y_{1t}, y_{2t} | \mathcal{F}_t)$ can be obtained in an analogous way. Therefore, the joint pdf $(y_{1t}, y_{2t})'$ for the model in (1) simplifies to

$$\begin{aligned} f(y_{1t}, y_{2t} | \mathcal{F}_t) &= \frac{[1 - \mathbf{I}(y_{1t} - c_1)][1 - \mathbf{I}(y_{2t} - c_2)]}{p_t} f(y_{1t}, y_{2t}; \boldsymbol{\theta}_1 | \mathcal{F}_t) \\ &\quad + \frac{[1 - \mathbf{I}(y_{1t} - c_1)] \mathbf{I}(y_{2t} - c_2)}{p_t} f(y_{1t}, y_{2t}; \boldsymbol{\theta}_2 | \mathcal{F}_t) \\ &\quad + \frac{\mathbf{I}(y_{1t} - c_1)[1 - \mathbf{I}(y_{2t} - c_2)]}{p_t} f(y_{1t}, y_{2t}; \boldsymbol{\theta}_3 | \mathcal{F}_t) \\ &\quad + \frac{\mathbf{I}(y_{1t} - c_1) \mathbf{I}(y_{2t} - c_2)}{p_t} f(y_{1t}, y_{2t}; \boldsymbol{\theta}_4 | \mathcal{F}_t). \end{aligned} \tag{33}$$

The normalisation factor p_t ensures that the condition

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_{1t}, y_{2t} | \mathcal{F}_t) \, dy_{1t} dy_{2t} = 1$$

is fulfilled. The term p_t is given by

$$\begin{aligned} p_t &= \int_{-\infty}^{c_1} \int_{-\infty}^{c_2} f(y_{1t}, y_{2t}; \boldsymbol{\theta}_1 | \mathcal{F}_t) \, dy_{2t} dy_{1t} + \int_{-\infty}^{c_1} \int_{c_2}^{\infty} f(y_{1t}, y_{2t}; \boldsymbol{\theta}_2 | \mathcal{F}_t) \, dy_{2t} dy_{1t} \\ &\quad + \int_{c_1}^{\infty} \int_{-\infty}^{c_2} f(y_{1t}, y_{2t}; \boldsymbol{\theta}_3 | \mathcal{F}_t) \, dy_{2t} dy_{1t} + \int_{c_1}^{\infty} \int_{c_2}^{\infty} f(y_{1t}, y_{2t}; \boldsymbol{\theta}_4 | \mathcal{F}_t) \, dy_{2t} dy_{1t} \\ &= \Pr(y_{1t} \leq c_1, y_{2t} \leq c_2 | \mathcal{F}_t) + \Pr(y_{1t} \leq c_1, y_{2t} > c_2 | \mathcal{F}_t) \\ &\quad + \Pr(y_{1t} > c_1, y_{2t} \leq c_2 | \mathcal{F}_t) + \Pr(y_{1t} > c_1, y_{2t} > c_2 | \mathcal{F}_t) \\ &= 1 + \Pr(E_t | \mathcal{F}_t), \end{aligned} \tag{34}$$

expression (34) being the same as (21). The log-likelihood function is then given by

$$L_T = \sum_{t=1}^{T} \log f(y_{1t}, y_{2t} | \mathcal{F}_t).$$

Define the vectors of parameters $\mathbf{c}$ and $\boldsymbol{\Delta}$ as

$$\mathbf{c} \equiv (c_1, c_2)', \qquad \boldsymbol{\Delta} \equiv \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \beta_1, \sigma_1^2, \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2', \beta_2, \sigma_2^2, \sigma_{12}\right)', \tag{35}$$

and denote by $\mathbf{c}^0$ and $\boldsymbol{\Delta}^0$ the true values of $\mathbf{c}$ and $\boldsymbol{\Delta}$, respectively. Further, denote by $\hat{\mathbf{c}}_{FIML}$ and $\hat{\boldsymbol{\Delta}}_{FIML}$ the estimators of $\mathbf{c}^0$ and $\boldsymbol{\Delta}^0$ as obtained by grid search and FIML, respectively. The strong consistency of $\hat{\mathbf{c}}$ and $\hat{\boldsymbol{\Delta}}$ is then stated in the following theorem:

Theorem 4 *Consider the model in* (1) *under the set of conditions imposed in Assumption* 1 *and Assumption* 2*. Further, assume that standard regularity conditions hold with respect to* $\boldsymbol{\Delta}$*. Then* $\hat{\mathbf{c}}_{FIML} \overset{a.s.}{\to} \mathbf{c}^0$ *and* $\hat{\boldsymbol{\Delta}}_{FIML} \overset{a.s.}{\to} \boldsymbol{\Delta}^0$*.*

Proof. See Appendix C.2. ■

The following three considerations about the FIML estimator for the model in (1) developed in this section are worth making:

1. The FIML estimator cannot be replaced by a SURE-type estimator: this is because the pdf $f(y_{1t}, y_{2t}; \boldsymbol{\theta}_k | \mathcal{F}_t)$ selected by the indicator functions $\mathbf{I}(y_{1t} - c_1)$ and $\mathbf{I}(y_{2t} - c_2)$ has to be weighted by the normalising term p_t, which depends upon the whole set of parameters of the model, and not just on the set $\boldsymbol{\theta}_k$ characterising the selected regime.

2. The reduced form in (19) depends upon the randomisation process d_t^E defined in (16); however, d_t^E does not enter the analytical expression of the joint pdf in (33). This means that knowledge of the randomisation process d_t^E is not required to estimate the system in (1).

3. Contrary to the class of GIVE estimators discussed in Section 6.1, the FIML estimator does not require any equation specific explanatory variables.

7 Monte Carlo analysis

The aim of the Monte Carlo analysis is twofold: to evaluate the absolute and relative performance of the class of single equation GIVE estimators and of the FIML estimator of the model in (1); to assess the validity of the theoretical conditions for identification stated in Theorem 1.

7.1 Experimental design

The Data Generating Process (DGP) is based on the following model

$$y_{it}^r = \delta_i + \alpha_i x_{it}^r + \beta_i \mathbf{I}\left(y_{jt}^r - c_j\right) + u_{it}^r, \quad i,j = 1,2 \quad i \neq j, \tag{36}$$

where $t = 1, \dots T$, $r = 1, \dots R$, r refers to the replication and R is the total number of replications; x_{it}^r is a simulated scalar explanatory variable; δ_i, α_i, β_i and c_i are scalar parameters, which are kept fixed throughout the replications. The estimated model is also given by (36). For the purpose of the analysis, we assume that the threshold parameters c_i are *known*. In the case of the FIML estimator, the resulting (normalised) log-likelihood function consistent with the general specification in (36) is maximised by using the BFGS algorithm, with starting values obtained from single equation OLS estimation. The whole experiment is run in Ox 3.30.

We focus upon the estimators for the contagion coefficient β_1. The performance of the estimators is assessed by computing the bias and RMSE, defined as

$$bias = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}_1^r - \beta_1\right), \qquad RMSE = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}_1^r - \beta_1\right)^2},$$

respectively, where $\hat{\beta}_1^r$ is the estimate of β_1 obtained from the $r - th$ replication. We also compute the two-sided rejection frequency, defined as the ratio between the number of times the computed test statistic lies outside the 95% confidence interval and the total number of replications R: if the test statistic is computed under the null hypothesis the rejection frequency is the size of the test; if the test statistic is computed under the alternative the rejection frequency is the power. In computing the test statistic for the FIML estimator, the Wald principle is employed, where the estimated covariance matrix is obtained as the inverse of the empirical Hessian.

In assessing the size performance of the tests, each replication can be seen as a Bernoulli trial; therefore, for a high value of R the normal approximation can be employed. As a consequence, we will not reject the null hypothesis of the size being equal to the significance level of 5% if the former lies within the interval

$$\left[0.05 \pm 1.96\sqrt{\frac{0.05 \cdot 0.95}{R}}\right].$$

The value $R = 2000$ is chosen so that the 95% confidence interval is approximately equal to $[0.04, 0.06]$.

Throughout the experiments the value of the contagion coefficient β_2 was kept fixed at $\beta_2 = 0.2$. Three classes of experiments were then considered: $\beta_1 = 0$ (so that no multiple equilibria arise) and $x_{1t}^r \neq x_{2t}^r$; $\beta_1 = 0.5$ (so that multiple equilibria do arise) and $x_{1t}^r \neq x_{2t}^r$; $\beta_1 = 0.5$ and $x_{1t}^r = x_{2t}^r = x_t^r$. The first two sets of experiments were used to assess the performance of the FIML and GIVE estimators; the third set of experiments aimed at checking the validity of the identification conditions, and, consistently with the theory developed in Section 6, only the FIML estimator was used[8].

Results of the Monte Carlo experiments for the FIML estimator and the class of GIVE estimators with $m = 1$ and $m = 6$ are reported in Table 1 (for the case $\beta_1 = 0$, $x_{1t}^r \neq x_{2t}^r$), in Table 2 (for the case $\beta_1 = 0.5$, $x_{1t}^r \neq x_{2t}^r$) and in Table 3 (for the case $\beta_1 = 0.5$, $x_{1t}^r = x_{2t}^r = x_t^r$)[9]. In generating the data, the sample size was set equal to $T = 50, 100, 200, 500, 1000$; the values for the threshold parameters were arbitrarily chosen to be equal to $c_1 = c_2 = 1.64$; the seed of the random generator was set equal to -1.

[8] For the case $x_{1t}^r \neq x_{2t}^r$ an experiment with $\beta_2 = 0.2$ was also attempted; the results are very similar to those obtained in the case $\beta_2 = 0.5$ and therefore not shown.

[9] For the case $x_{1t}^r \neq x_{2t}^r$ the OLS estimator (which delivers inconsistent estimates of the contagion coefficient β_1) and the GIVE estimator for $m = 2, 3, 4, 5$ were also analysed.

Experiment 1: $\beta_1 = 0$, $x_{1t}^r \neq x_{2t}^r$. The DGP in (36) simplifies to

$$\begin{aligned} y_{1t}^r &= \delta_1 + \alpha_1 x_{1t}^r + u_{1t}^r, \\ y_{2t}^r &= \delta_2 + \alpha_2 x_{2t}^r + \beta_2 \mathbf{I}\left(y_{1t}^r - c_1\right) + u_{2t}^r. \end{aligned}$$

The error terms are generated by adopting the following common factor structure

$$u_{it}^r = \frac{\gamma_i f_t^r + \varepsilon_{it}^r}{\sqrt{\gamma_i^2 + 1}},$$

where $\varepsilon_{it}^r \sim NID\left(0,1\right)$, $f_t^r \sim NID\left(0,1\right)$, while the coefficient $\gamma_i \sim U\left(0.8,1\right)$ is fixed in repeated samples. In this way

$$\boldsymbol{\Sigma}_{\mathbf{u}} \equiv \begin{pmatrix} 1 & \rho_{u_1 u_2} \\ \rho_{u_1 u_2} & 1 \end{pmatrix}$$

where

$$\rho_{u_1 u_2} \equiv E\left(u_{1t}^r u_{2t}^r\right) = \frac{\gamma_1 \gamma_2}{\sqrt{\gamma_1^2 + 1}\sqrt{\gamma_2^2 + 1}},$$

the average value of the correlation coefficient $\rho_{u_1 u_2}$ being equal to

$$\bar{\rho}_{u_1 u_2} \equiv E\left(\rho_{u_1 u_2}\right) = E\left[\left(\frac{\gamma_i}{\sqrt{\gamma_i^2 + 1}}\right)\right]^2 = 0.1616$$

The equation specific explanatory variables x_{it}^r, $i = 1, 2$ are generated as

$$\begin{pmatrix} x_{1t}^r \\ x_{2t}^r \end{pmatrix} \sim NID\left(\mathbf{0}; \boldsymbol{\Sigma}_{\mathbf{x}}\right).$$

The covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$ is implicitly defined by generating x_{it}^r by means of the following one-factor model

$$x_{it}^r = \frac{\phi_i h_t^r + q_{it}^r}{\sqrt{\phi_i^2 + 1}},$$

where $q_{it}^r \sim NID\left(0,1\right)$, $h_t^r \sim NID\left(0,1\right)$ while the coefficient $\phi_i \sim U\left(0.8,1\right)$ is fixed in repeated samples. In this way we have

$$\boldsymbol{\Sigma}_{\mathbf{x}} \equiv \begin{pmatrix} 1 & \rho_{x_1 x_2} \\ \rho_{x_1 x_2} & 1 \end{pmatrix},$$

where

$$\rho_{x_1 x_2} = \frac{\phi_1 \phi_2}{\sqrt{\phi_1^2 + 1}\sqrt{\phi_2^2 + 1}}.$$

In order to ensure that the regressors are independent of the errors, h_t^r and f_t^r are drawn independently of each other.

The role of the slope coefficients α_1 and α_2 is to control for the goodness of fit of the model. In the case of the equation for y_{1t}^r, since no right-hand-side variable is endogenous, the coefficient of determination can be easily obtained as

$$R_1^2 = 1 - \frac{Var\left(u_{1t}^r\right)}{Var\left(y_{1t}^r\right)},$$

where

$$Var\left(y_{1t}^r\right) = \alpha_1^2 Var\left(x_{1t}^r\right) + Var\left(u_{1t}^r\right) = \alpha_1^2 + 1$$

so that

$$R_1^2 = 1 - \frac{1}{\alpha_1^2 + 1} = \frac{\alpha_1^2}{\alpha_1^2 + 1}.$$

In the case of the equation for y_{2t}^r, because the term $\mathbf{I}\left(y_{1t}^r - c_1\right)$ is endogenous the coefficient of determination cannot be computed from the residuals u_{2t}^r. Following the approach introduced in Pesaran and Smith (1994), the prediction errors are deployed; they are defined as

$$v_{2t}^r = y_{2t}^r - E\left(y_{2t}^r \left| \mathcal{F}_t^r \right.\right) = y_{2t}^r - \delta_2 - \alpha_2 x_{2t}^r - E\left[\mathbf{I}\left(y_{1t}^r - c_1\right) \left| \mathcal{F}_t^r \right.\right]$$

where $\mathcal{F}_t^r \equiv \left(x_{1t}^r, x_{2t}^r\right)'$ and

$$E\left[\mathbf{I}\left(y_{1t}^r - c_1\right) \left| \mathcal{F}_t^r \right.\right] = \Pr\left(y_{1t}^r - c_1 > 0 \left| \mathcal{F}_t^r \right.\right) = 1 - \mathbf{\Phi}\left(c_1 - \delta_1 - \boldsymbol{\alpha}_1 x_{1t}^r\right).$$

Therefore, the coefficient of determination can be computed by simulation as

$$R_2^2 = \frac{\sum_{r=1}^R}{R}\left[1 - \frac{\sum_{t=1}^T \left(v_{2t}^r\right)^2}{\sum_{t=1}^T \left(y_{2t}^r - \bar{y}_2^r\right)^2}\right],$$

where $\bar{y}_2^r = \left(\sum_{t=1}^T y_{2t}^r\right) \Big/ T$. We set $\alpha_1 = \alpha_2 = \alpha$ and we consider two cases, $\alpha = 0.5$ and $\alpha = 1$: in the former $R_1^2 = 0.2$ and $R_2^2 \simeq 0.2$; in the latter $R_1^2 = 0.5$ and $R_2^2 \simeq 0.5$.[10]

Finally, the role of the parameters δ_1 and δ_2 is to control for the unconditional probability of crisis, so to assess the performance of the estimators as the unconditional probabilities of crisis change. In the case of y_{1t}^r the unconditional probability of crisis π_1 is obtained in closed form as

$$\begin{aligned}
\Pr\left(y_{1t} - c_1 > 0\right) &= \Pr\left(\delta_1 + \alpha_1 x_{1t}^r + u_{1t}^r - c_1 > 0\right) = \Pr\left(\frac{\alpha_1 x_{1t}^r + u_{1t}^r}{\sqrt{\alpha_1^2 + 1}} > \frac{c_1 - \delta_1}{\sqrt{\alpha_1^2 + 1}}\right) \\
&= 1 - \mathbf{\Phi}\left(\frac{c_1 - \delta_1}{\sqrt{\alpha_1^2 + 1}}\right) \equiv \pi_1,
\end{aligned}$$

so that δ_1 is given by

$$\delta_1 = c_1 - \sqrt{\alpha_1^2 + 1}\left[\mathbf{\Phi}^{-1}\left(1 - \pi_1\right)\right].$$

In the case of y_{2t}^r the probability of crisis is computed by simulation as

$$\pi_2 \equiv \Pr\left[\left(y_{2t}^r - c_2 > 0\right)\right] = \frac{\sum_{r=1}^R}{R}\left[\frac{\sum_{t=1}^T \mathbf{I}\left(y_{2t}^r - c_2\right)}{T}\right],$$

and we calibrate δ_2 so to control for π_2. We set

$$\pi_1 = \pi = 0.005, 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50 \simeq \pi_2.$$

This is an important part in the experimental design: if $\mathbf{I}\left(y_{jt}^r - c_j\right) = 0\ \forall t$, then no observations are available to estimate β_i; conversely, if $\mathbf{I}\left(y_{jt}^r - c_j\right) = 1\ \forall t$, then β_i cannot be identified from the intercept δ_i. Therefore, each replication is repeated until $\mathbf{I}\left(y_{jt}^r - c_j\right)$ is different from being a vector of all zeros or ones.

Experiment 2: $\beta_1 = 0.5$, $x_{1t}^r \neq x_{2t}^r$. The DGP is given by the reduced form obtained in (19), where the value $\pi_d^E = 0.5$ is chosen; further, the error terms u_{it}^r and the explanatory variables x_{it}^r are generated as in Experiment 1.

The goodness of fit is controlled for by means of the slope coefficients α_1 and α_2. Due to the endogeneity induced by the indicator functions $\mathbf{I}\left(y_{it}^r - c_i\right)$, we follow the approach proposed in Pesaran and Smith (1994), and employ the prediction errors v_{it}^r to compute the coefficients of determination. The prediction errors v_{it}^r are defined as

$$v_{it}^r = y_{it}^r - E\left(y_{it}^r \left| \mathcal{F}_t^r \right.\right) = y_{it}^r - \delta_i - \alpha_i x_{it}^r - \beta_i \Pr\left(y_{jt}^r - c_j > 0 \left| \mathcal{F}_t^r \right.\right), \quad i, j = 1, 2, \quad i \neq j,$$

[10] Notice that R_1^2 and R_2^2 have very similar values because the endogenous dummy $\mathbf{I}\left(y_{1t}^r - c_1\right)$ determines a shift in the intercept in the equation for y_{1t}^r. Therefore, the effect induced on explanatory power of the model is likely to be negligible.

the general expression for $\Pr\left(y_{it} - c_i > 0 \left| F_t^r \right.\right)$ being derived in Appendix B. Therefore, $R_i^2, i = 1, 2,$ may be computed by simulation as

$$R_i^2 = \frac{\sum_{r=1}^{R}}{R}\left[1 - \frac{\sum_{t=1}^{T}\left(v_{it}^r\right)^2}{\sum_{t=1}^{T}\left(y_{it}^r - \bar{y}_i^r\right)^2}\right],$$

where $\bar{y}_i^r = \left(\sum_{t=1}^{T} y_{it}^r\right) \Big/ T$. We set $\alpha_1 = \alpha_2 = \alpha$ and consider two cases, $\alpha = 0.5$ and $\alpha = 1$, which correspond to $R_i^2 \simeq 0.5$ and $R_i^2 \simeq 1$, $i = 1, 2$, respectively.

Finally, unconditional crises probabilities are controlled for by means of the parameters δ_1 and δ_2. These probabilities are computed by simulation as

$$\pi_i \equiv \Pr\left(y_{it}^r - c_i > 0\right) = \frac{\sum_{r=1}^{R}}{R}\left[\frac{\sum_{t=1}^{T} \mathbf{I}\left(y_{it}^r - c_i\right)}{T}\right],$$

and δ_i is chosen so to control for π_i. We set

$$\pi_1 \simeq \pi_2 \simeq \pi = 0.005, 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50.$$

Experiment 3: $\beta_1 = 0.5$, $x_{1t}^r = x_{2t}^r = x_t^r$. The experimental set-up is the same as in Experiment 2, with the exception of the restriction $x_{1t}^r = x_{2t}^r = x_t^r$ being imposed both upon the DGP and the estimated model given in (36). Further, for the sake of simplicity we only consider the case $\alpha_1 = \alpha_2 = \alpha = 1$.

7.2 Performance of GIVE and FIML estimators

In order to assess the absolute and relative performance of the GIVE and FIML estimators, we analyse the results reported in Table 1 and Table 2.

7.2.1 Bias and RMSE

The bias of the FIML estimator decreases with the sample size T and with the probability of crisis π (up to $\pi = 0.5$), while it does not generally show any clear correlation with α (and therefore with the goodness of fit of the model); also, the bias does not seem to depend upon the magnitude of β_1. Considering the GIVE estimators, the bias decreases with the sample size T and with the probability of crisis π (although in a more erratic way in the case $m = 1$); it also decreases with α, as the instruments become stronger. The bias of the GIVE estimators follows an unclear pattern in relation to the magnitude of β_1: for $\alpha = 0.5$ it looks like an increase in β_1 generally leads to an increase in the bias, probably because the instruments become slightly weaker; however such an effect does not seem to be present when $\alpha = 1$ and the instruments are therefore stronger. Dealing with the relative performance of the GIVE estimators for $m = 1$ and $m = 6$, the latter results in a lower value of the bias for low values of π (such as $\pi = 0.005$ and $\pi = 0.01$), while the former tends to perform better for higher values of π (such as $\pi \geq 0.05$). Finally, the bias of the FIML estimator is lower than that of the GIVE estimators for virtually any combination of T and π.

The RMSE of the FIML estimator decreases with the sample size T and with the probability of crisis π, a pattern similar to that of the bias; it also diminishes with α (this being a difference compared to the bias), while it does not show any clear pattern related to the magnitude of β_1. As far as the GIVE estimators are concerned, their bias decreases with T and π (although in a more erratic way in the case $m = 1$) as well as with α (this last feature confirming one more time the presence of the weak instruments problem), while no clear pattern seems to be related to the magnitude of β_1; also, the GIVE estimator with $m = 6$ has lower RMSE compared to that with $m = 1$, although the difference tends to disappear as both T and π increase. Finally, the FIML estimator is always more efficient than the GIVE estimators, although the efficiency loss diminishes with T and π as well as with α: for example, in the case of Experiment 2 with $\alpha = 0.5$, if $T = 50$ and $\pi = 0.005$ the efficiency loss of using the GIVE estimator with $m = 6$ rather the FIML is 230%, while it reduces to 122% when $T = 1000$ and $\pi = 0.5$; further, for $\alpha = 1$, $T = 1000$ and $\pi = 0.5$ the efficiency loss falls to 69%. The efficiency loss is due to the limited information nature of the GIVE estimators as discussed in Section 6.1.

7.2.2 Size and power

Starting from the FIML estimator, for $T = 50$ the actual size never approaches the nominal size, regardless the value of π. As the sample size T increases the actual size tends to approach the nominal size, this feature generally being true for any value of the probability of crisis π. The only exception arises when $\beta_1 = 0.5$ and $\alpha = 0.5$, where the actual size never reaches the nominal size as T increases when $\pi = 0.005, 0.01$. A possible explanation is that as the magnitude of β_1 increases a higher absolute number of observations in the crisis regime is required to provide a consistent estimator; in addition, a low value of α combined with a high value of β_1 may raise identification issues. Considering the GIVE estimators, for $m = 1$ the test is generally undersized when $\alpha = 0.5$, the nominal size being systematically reached only when $T \geq 500$ and $\pi \geq 0.10$ (apart from the case $T = 1000$ and $\pi = 0.50$); in the case $\alpha = 1$ the size performance improves and the nominal size is reached for a wider combination of T and π. In the case of the GIVE estimator with $m = 6$, the nominal value of the size is achieved for a larger combinations of T and π compared to the case $m = 1$ both for $\alpha = 0.5$ and $\alpha = 1$; further, analogously to the case $m = 1$ the size performance is better when $\alpha = 1$ than when $\alpha = 0.5$. Finally, a comparison between the FIML and the GIVE estimators shows that the former definitely achieves a better size performance.

The power of the tests is computed by testing the null $\beta_1 = 0.5$ in Experiment 1 (where the data are generated under the alternative $\beta_1 = 0$) and $\beta_1 = 1$ in Experiments 2 and 3 (where the data are generated under the alternative $\beta_1 = 0.5$); from testing theory, power comparisons are only made for combinations of T and π such that the difference between significance level and actual size is statistically insignificant. In the case of the FIML estimator, the power increases with the sample size T as well as with the probability of crisis π, as also shown in Figure 2 and Figure 3 below; the power also increases with α. Tests based upon the GIVE estimators show a similar behaviour, as the power increases with T, π and α. Further, for $\alpha = 0.5$ the GIVE estimator with $m = 1$ seems to provide a test with slightly higher power than the GIVE estimator with $m = 6$; however, for $\alpha = 1$ the GIVE estimator with $m = 6$ seems to outperform that with $m = 1$. Finally, the FIML estimator generates a test which is clearly more powerful than those obtained from the GIVE estimators.

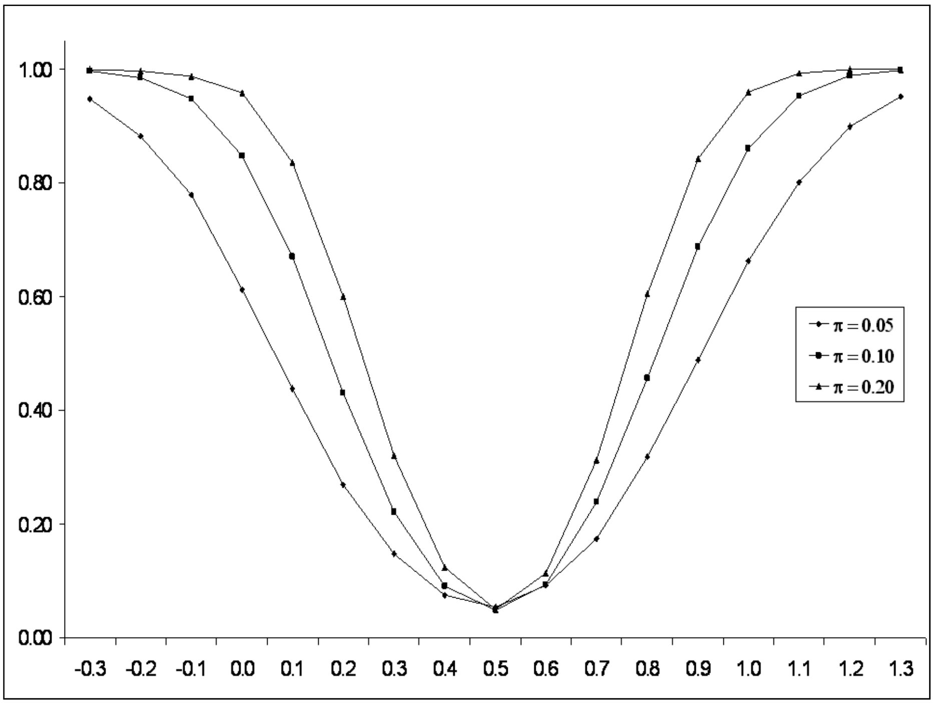

Figure 2: Power Function for FIML Estimator, Experiment 2, $T = 500$, $\alpha = 0.5$.

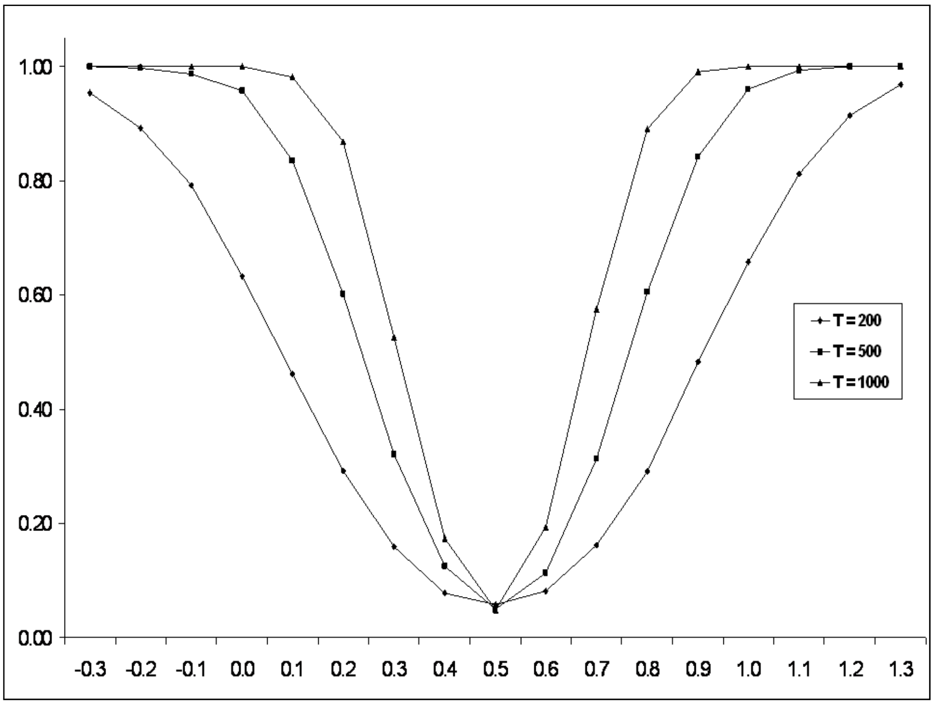

Figure 3: Power Function for FIML Estimator, Experiment 3, $\pi = 0.2$, $\alpha = 0.5$.

7.3 Identification

In order to assess the validity of Theorem 1, we consider the results reported in Table 3 and compare them with those presented in Table 2 for the case of the FIML with $\alpha = 1$.

Starting from the bias, it decreases both with T and π, with the only noticeable exception given by the case $T = 50$ and $\pi = 0.50$. Further, for low values of T and π, the bias is higher than in the case with equation specific regressors, the difference vanishing as T and π increase; also in this case the only exception arises when $T = 50$ and $\pi = 0.50$. The RMSE shows a pattern similar to the bias, in the sense that it decreases both with T and π; compared to the case with equation specific regressors, the RMSE is lower when $\pi = 0.005$ and it is generally higher in the remaining cases.

The size of the test shows a pattern similar to that for the case with equation specific regressors, the test being consistently oversized only when $T = 50$. Further, where power comparisons are applicable (that is for combinations of T and π where for *each* model the difference between significance level and actual size is statistically insignificant) the imposition of exclusion restrictions improves the power performance of the test.

In conclusion, although from Theorem 1 no exclusion restriction is required to identify the model in (1), the presence of equation specific explanatory variables does contribute towards a more precise estimation of the parameters.

8 Empirical illustration

In this section we provide an application to real data of the theoretical framework we have developed so far: in particular, we aim at analysing the interactions between equity markets. The data and the model specification are described in Section 8.1 and Section 8.2, respectively, while the results are discussed in Section 8.3.

8.1 Data

In our analysis, we follow the methodology advanced in Kaminsky and Reinhart (2002), and divide world equity markets into two broad categories, namely centre and periphery markets: the first groups consists of the biggest financial centres, i.e. New York, London and Tokyo; the latter includes all other markets. We then focus on the interactions between the New York Stock Exchange and three periphery markets, namely Frankfurt, Zurich and Paris.

We make use of daily stock market spot prices for the S&P 500 (New York), DAX 30 (Frankfurt), SMI (Zurich) and CAC 40 (Paris) recorded at 16:00 London time (pseudo-closing prices), where all the stock market indices are in US dollars; the data were obtained from Datastream for the period 3 August 1990 to 30 June 2005. All the mentioned indices describe the behaviour of the biggest firms. The S&P 500 is a capitalisation-weighted index of 500 stocks of US public companies; it approximately represents 75% of total market capitalisation. The DAX 30 includes the 30 largest German securities according to market capitalisation and turnover. The SMI is made of a maximum of 30 of the largest and most liquid stocks in the Swiss market. The CAC 40 is a weighted-average index of 40 stocks, the weights being based upon the closing price of the last traded day.

Pseudo-closing prices were chosen over actual closing prices because international stock markets have different trading hours. Indeed, New York trades from 9:30 to 16:00 Eastern standard time (which corresponds to 14:30 to 21:00 London time); Frankfurt, Paris and Zurich trade from 9:00 to 17:30 local time (which corresponds to 8:00 to 16:30 London time). Therefore, the use of daily closing prices in our analysis would have led to an underestimation of the correlation between stock markets themselves[11].

For each market i, the spot prices at time t (P_{it}) were converted into continuously compounded returns (r_{it}) as

$$r_{it} = (\log P_{it} - \log P_{i,t-1}) * 100, \quad i = 1, 2. \tag{37}$$

After removing holidays in each country, we were left with 3741 observations of common trading days for the four series[12]. Descriptive statistics for the resulting stock market returns and the correlations between them are provided in Table 4. Average daily returns are all positive, with New York providing the highest rate followed by Zurich, Frankfurt and Paris. The S&P 500 is also the least volatile index, as evidenced by the value of its sample standard deviation, followed the SMI, the CAC 40 and the DAX 30. The measure of skewness shows that the S&P 500, the DAX 30 and the SMI are negatively skewed compared to the normal distribution, while the CAC 40 is positively skewed; in addition, all returns' distributions are highly leptokurtic compared to the normal distribution. The Jarque-Bera test for normality rejects the null hypothesis at 1% level for all returns. Finally, the correlation between the S&P 500 and any other market is lower than the correlation between any other two markets: this is due to the significant time difference in *local time* between the trading hours in New York and the other markets, therefore providing evidence in favour of using pseudo rather than actual closing prices.

8.2 Model specification

In carrying out the empirical analysis, we first note that a crisis in a stock market is associated with an extreme *negative* value of r_{it}; therefore, a crisis takes place whenever $r_{it} < -c_i$, or equivalently $-r_{it} - c_i > 0$: the crisis indicator is then defined as

$$\mathbf{I}\left(-r_{it} - c_i\right), \quad i = 1, 2.$$

Further, in order to define the dependent variable y_{it}, recall that stock market returns exhibit a high degree of conditional heteroskedasticity, as extensively discussed in the literature following the work by Engle (1982) and Bollerslev (1986): therefore, consistently with the theory we developed, the returns r_{it} have to be *devolatised*. The variable y_{it} is then defined as

$$y_{it} \equiv -\frac{r_{it}}{\sigma_{i|t-1}}, \quad \sigma^2_{i|t-1} \equiv \mathrm{Var}\left(r_{it}\left|\boldsymbol{\Omega}_{i,t-1}\right.\right), \quad i = 1, 2, \tag{38}$$

where $\boldsymbol{\Omega}_{i,t-1}$ is the information set up to time $t-1$. The conditional standard deviation $\sigma_{i|t-1}$ is estimated by fitting the returns r_{it} with the GARCH$(1,1) - t$ model introduced by Bollerslev (1987):

[11] The use of pseudo-closing prices to avoid the problem of non-synchronous data was first suggested in Martens and Poon (2001); see Chapter 5 of Tsay (2005) for a discussion of the characteristics of high-frequency data. Also, notice that a similar exercise could be performed with the London Stock Exchange acting as centre market. Conversely, the same kind of analysis cannot be performed with the Japanese stock market, since it does not have any common trading time with the stock markets we chose as periphery markets.

[12] In each market a day is considered a holiday if the return on that day is exactly equal to zero.

compared to the standard GARCH model with conditionally Gaussian disturbances, this represents a more flexible approach to account for the leptokurtosis in stock market returns as evidenced in Table 4. The GARCH$(1,1)-t$ model, specified in terms of the returns r_{it}, is given by

$$
\begin{array}{l}
r_{it} = \mu_i + \sum_{k=1}^{5} \gamma_{ik} r_{i,t-k} + \varepsilon_{it}, \\
\varepsilon_{it} = z_{it} \sigma_{i|t-1}, \\
z_{it} \left| \boldsymbol{\Omega}_{i,t-1} \sim iidt_v \left(0,1\right), \right. \\
\sigma^2_{i|t-1} = \omega + \alpha \varepsilon^2_{i,t-1} + \beta \sigma^2_{i|t-2},
\end{array}
\tag{39}
$$

where v denotes the number of degrees of freedom of the t distribution[13]. The market returns r_{it} in (39) are modelled as an autoregressive process of order five so to control for serial correlation as well as weekly effects. Taking into account (38), the model specification becomes

$$
\begin{array}{l}
\hat{y}_{1t} = \delta_1 + \boldsymbol{\alpha}_1{}' \mathbf{x}_{1t} + \beta_1 \mathbf{I} \left(-r_{2t} - c_2\right) + u_{1t}, \\
\hat{y}_{2t} = \delta_2 + \boldsymbol{\alpha}_2{}' \mathbf{x}_{2t} + \beta_2 \mathbf{I} \left(-r_{1t} - c_1\right) + u_{2t},
\end{array}
\tag{40}
$$

where

$$
\hat{y}_{it} \equiv -\frac{r_{it}}{\hat{\sigma}_{i|t-1}}, \quad \mathbf{x}_{it} \equiv \left(\hat{y}_{i,t-1}, \ldots, \hat{y}_{i,t-5}\right)', \quad i = 1,2,
\tag{41}
$$

$\hat{\sigma}_{i|t-1}$ being the estimate of $\sigma_{i|t-1}$ arising from (39), and the subscript $i = 2$ always referring to the S&P 500 index. In this way we are left with 3736 observations.

The threshold parameters c_1 and c_2 are estimated by grid search, while the remaining set of parameters is estimated by FIML, under the assumption of jointly normally distributed errors $(u_{1t}, u_{2t})'$: this choice over single equation GIVE estimation is motivated by the superior performance of the FIML estimator, as it arises from the Monte Carlo analysis carried out in Section 7. Notice that the normality assumption imposed on the errors is a plausible one, since the original returns r_{it} have been devolatised, and the crisis indicators account for the presence of extreme values in the empirical distribution of $\hat{y}_{it}$. For each stock market return, the width of the grid for the corresponding threshold is chosen so to include observations between the bottom 0.5% and 20% quantiles of the empirical distribution of r_{it}. The resulting intervals for the threshold values for the S&P 500, SMI, DAX 30 and CAC 40 are $\mathbf{C}_{S\&P500} \equiv [0.65, 3.40]$, $\mathbf{C}_{SMI} \equiv [0.80, 3.60]$, $\mathbf{C}_{DAX30} \equiv [0.90, 4.70]$ and $\mathbf{C}_{CAC40} \equiv [0.90, 4.10]$, respectively, with a step equal to 0.01.

8.3 Results

Results from estimation of the model in (40) by FIML are reported in Table 5. Starting from the effect of the NYSE upon periphery markets, we can see that the latter seem to define a negative extreme event (i.e. a crisis) in the former almost in the same way: both the SMI and the CAC 40 react when the (standardised) daily returns in the S&P 500 goes below −2.69% (which corresponds to a proportion of crisis periods equal to 1.15%), whereas the analogous threshold related to the DAX 30 is equal to −2.44% (and a related proportion of crisis episodes equal to 1.85%). Consistently with this result, the effect of a crisis in the NYSE on any of the other markets is of similar magnitude: it causes a drop of 0.57% in the SMI, 0.58% in the DAX 30, and 0.64% in the CAC 40, each variation being expressed in terms of the corresponding *devolatised* stock market return[14]: therefore, the periphery markets we consider seem to have the same degree of vulnerability with respect to extreme negative events taking place in the NYSE.

Turning the attention to the effect of the periphery markets on the S&P 500, it can be seen that the CAC 40 seems to be the index that most often affects the S&P 500, being perceived to be in crisis 4.39% of the times, with an associated threshold of −2.20%; it is then followed by the DAX 30 and the SMI (2.38% and 0.70% of crisis periods, and threshold values equal to −3.06% and −3.15%, respectively). It is also interesting to note that the higher the number of times an index affects the S&P 500, the lower the magnitude of the effect (i.e. the magnitude of the corresponding contagion

[13] The GARCH$(1,1)-t$ model in (39) is estimated by means of the procedure developed by Laurent and Peters (2005).

[14] This is because of the definition of the dependent variable y_{it} given in (41), so that each return has same conditional variance equal to one.

coefficient): a crisis in the CAC 40 causes a drop of 0.34% in the S&P 500, whereas an episodes of turmoil affecting the DAX 30 and the SMI determines a fall in the S&P 500 equal to 0.50% and 0.74%, respectively.

9 Concluding remarks

This paper has dealt with the theoretical issues of identification and estimation in the incoherent model of contagion advanced in Pesaran and Pick (2007). In particular, two main results were proved. First, identification of the relevant parameters of the model (i.e. identification of the parameters capturing contagion effects from those related to interdependence) does not require identification of the process d_t, which drives the choice of the solution in the region where multiple equilibria arise. Second, FIML estimation of the model does not require estimation of the process d_t. Therefore, more generally it was proved that *in-sample* analysis of the model does not require information about the process d_t.

The work undertaken in this paper is subject to several possible extensions. In particular, we believe that three of them are worth discussing. First, the issue of statistical inference, i.e. testing for contagion, has not been addressed. The problem is that under the null hypothesis of no contagion in market i, i.e. under $H_0 : \beta_i = 0$, the threshold parameter c_j is not identified, for $i, j = 1, 2$ and $i \neq j$. The problem of performing hypothesis testing when a nuisance parameter is identified only under the alternative is known in the statistical literature as the Davies problem, after the work by Davies (1977, 1987). Given the model in (1), a possible solution would be to follow the procedure advanced in Massacci (2007).

Second, the FIML estimator we propose is likely to be sensitive to misspecifications in the underlying distribution of the error terms $(u_{1t}, u_{2t})'$: therefore, alternative estimation methods, such as GMM, are worth considering.

Finally, the analysis we have carried out focuses on *in-sample* inference, without dealing with *out-of-sample* forecasting. For example, this aspect could be relevant when forecasting the direction of the market. Therefore, a test of market timing of the kind proposed in Pesaran and Timmermann (1992) could represent a useful starting point.

These possible developments go beyond the purpose of this paper, and will be the subject of separate studies.

Appendix

A Further analysis of the model

In this appendix we provide a brief analysis of the model in (1) when $\beta_1, \beta_2 < 0$, and when $\beta_1 < 0$ and $\beta_2 > 0$.[15] Consider first the case $\beta_1, \beta_2 < 0$. Taking into account (8) and (10), the system in (1) can be written as

$$\begin{aligned} Y_{1t} &= W_{1t} + \mathbf{I}(-Y_{2t}), \\ Y_{2t} &= W_{2t} + \mathbf{I}(-Y_{1t}). \end{aligned}$$

The five mutually exclusive solution regions are then given by

$$\text{region A}_t\text{:} \quad \begin{cases} W_{1t} \geq 0 \\ W_{2t} \geq 0 \end{cases};$$

$$\text{region B}_t\text{:} \quad \begin{cases} W_{1t} \geq 0 \\ -1 \leq W_{2t} < 0 \end{cases} \cup \begin{cases} W_{1t} \geq -1 \\ W_{2t} < -1 \end{cases};$$

$$\text{region C}_t\text{:} \quad \begin{cases} W_{1t} < -1 \\ W_{2t} < -1 \end{cases};$$

$$\text{region D}_t\text{:} \quad \begin{cases} W_{1t} < -1 \\ -1 \leq W_{2t} < 0 \end{cases} \cup \begin{cases} W_{1t} < 0 \\ W_{2t} \geq 0 \end{cases};$$

$$\text{region E}_t\text{:} \quad \begin{cases} -1 \leq W_{1t} < 0 \\ -1 \leq W_{2t} < 0 \end{cases}.$$

The reduced form is then given by

$$\begin{cases} y_{1t} = w_{1t} \leq c_1 \\ y_{2t} = w_{2t} \leq c_2 \end{cases}, \qquad \text{(event A}_t\text{)}$$

$$\begin{cases} y_{1t} = w_{1t} + \beta_1 \leq c_1 \\ y_{2t} = w_{2t} > c_2 \end{cases}, \qquad \text{(event B}_t\text{)}$$

$$\begin{cases} y_{1t} = w_{1t} + \beta_1 > c_1 \\ y_{2t} = w_{2t} + \beta_2 > c_2 \end{cases}, \qquad \text{(event C}_t\text{)}$$

$$\begin{cases} y_{1t} = w_{1t} > c_1 \\ y_{2t} = w_{2t} + \beta_2 \leq c_2, \end{cases}, \qquad \text{(event D}_t\text{)}$$

$$\begin{cases} y_{1t} = w_{1t} + d_t^E \beta_1 \\ y_{2t} = w_{2t} + \left(1 - d_t^E\right) \beta_2 \end{cases}, \qquad \text{(event E}_t\text{)}$$

the process d_t^E being defined in (16). The normalising term p_t is then equal to

$$p_t = 1 + \Pr\left(E_t \left| \mathcal{F}_t \right.\right),$$

the analytical expression for $\Pr\left(E_t \left| \mathcal{F}_t \right.\right)$ being the same as that obtained in (22). Therefore, the expression for the joint pdf given in (33) remains valid for the case $\beta_1, \beta_2 < 0$.

Consider now the case $\beta_1 < 0$ and $\beta_2 > 0$. The model can be written as

$$\begin{aligned} Y_{1t} &= W_{1t} + \mathbf{I}(Y_{2t}), \\ Y_{2t} &= W_{2t} + \mathbf{I}(-Y_{1t}). \end{aligned}$$

[15] The case where $\beta_1 > 0$ and $\beta_2 < 0$ is analogous to that where $\beta_1 < 0$ and $\beta_2 > 0$, and it is therefore omitted.

In this case only *four* mutually exclusive solution regions arise

$$\text{region A}_t\text{:} \quad \begin{cases} W_{1t} \geq -1 \\ W_{2t} > 0 \end{cases},$$

$$\text{region B}_t\text{:} \quad \begin{cases} W_{1t} \geq 0 \\ W_{2t} \leq 0 \end{cases},$$

$$\text{region C}_t\text{:} \quad \begin{cases} W_{1t} < 0 \\ W_{2t} \leq -1 \end{cases},$$

$$\text{region D}_t\text{:} \quad \begin{cases} W_{1t} < -1 \\ W_{2t} > -1 \end{cases},$$

while in region E_t defined as

$$\text{region E}_t\text{:} \quad \begin{cases} -1 \leq W_{1t} < 0 \\ -1 < W_{2t} \leq 0 \end{cases}$$

no solution arises. The reduced form is then given by

$$\begin{cases} y_{1t} = w_{1t} + \beta_1 \leq c_1 \\ y_{2t} = w_{2t} > c_2 \end{cases}, \quad (\text{event A}_t)$$

$$\begin{cases} y_{1t} = w_{1t} \leq c_1 \\ y_{2t} = w_{2t} \leq c_2 \end{cases}, \quad (\text{event B}_t)$$

$$\begin{cases} y_{1t} = w_{1t} > c_1 \\ y_{2t} = w_{2t} + \beta_2 \leq c_2 \end{cases}, \quad (\text{event C}_t)$$

$$\begin{cases} y_{1t} = w_{1t} + \beta_1 > c_1 \\ y_{2t} = w_{2t} + \beta_2 > c_2 \end{cases}, \quad (\text{event D}_t)$$

and no multiple equilibria arise. In this case the normalising term p_t is given by

$$p_t = 1 - \Pr(E_t \,|\mathcal{F}_t),$$

where the analytical expression for $\Pr(E_t \,|\mathcal{F}_t)$ is of the same magnitude but opposite sign compared to the expression provided in (22). Therefore, the expression for the joint pdf given in (33) remains valid for the case $\beta_1 < 0$ and $\beta_2 > 0$.

B Conditional probability of crisis

In this appendix we provide the analytical derivation of the general expression for $\Pr(y_{it} - c_i > 0 \,|\mathcal{F}_t)$, $i = 1, 2$. From the reduced form equation (19) it follows that

$$\begin{aligned} \Pr(y_{1t} - c_1 > 0 \,|\mathcal{F}_t) \;=\; & \Pr(A_t \,|\mathcal{F}_t) \Pr(y_{1t} - c_1 > 0 \,|A_t; \mathcal{F}_t) \\ & + \Pr(B_t \,|\mathcal{F}_t) \Pr(y_{1t} - c_1 > 0 \,|B_t; \mathcal{F}_t) \\ & + \Pr(C_t \,|\mathcal{F}_t) \Pr(y_{1t} - c_1 > 0 \,|C_t; \mathcal{F}_t) \\ & + \Pr(D_t \,|\mathcal{F}_t) \Pr(y_{1t} - c_1 > 0 \,|D_t; \mathcal{F}_t) \\ & + \Pr(E_t \,|\mathcal{F}_t) \Pr(y_{1t} - c_1 > 0 \,|E_t; \mathcal{F}_t) \end{aligned}$$

where

$$\begin{aligned} \Pr(y_{1t} - c_1 > 0 \,|A_t; \mathcal{F}_t) &= \Pr(y_{1t} - c_1 > 0 \,|B_t; \mathcal{F}_t) = 1, \\ \Pr(y_{1t} - c_1 > 0 \,|C_t; \mathcal{F}_t) &= \Pr(y_{1t} - c_1 > 0 \,|D_t; \mathcal{F}_t) = 0, \\ \Pr(y_{1t} - c_1 > 0 \,|E_t; \mathcal{F}_t) &= \left(1 - \pi_d^E\right), \end{aligned}$$

so that

$$\Pr(y_{1t} - c_1 > 0 \,|\mathcal{F}_t) \;=\; \Pr(A_t \,|\mathcal{F}_t) + \Pr(B_t \,|\mathcal{F}_t) + \left(1 - \pi_d^E\right) \Pr(E_t \,|\mathcal{F}_t);$$

therefore, taking into account (14), we have

$$\begin{aligned}\Pr\left(y_{1t}-c_1>0\left|\digamma_t\right.\right) &= \Pr\left(W_{1t}>0\left|\digamma_t\right.\right)+\Pr\left(-1<W_{1t}\leq 0,W_{2t}>0\left|\digamma_t\right.\right)\\ &\quad +\left(1-\pi_d^E\right)\Pr\left(-1<W_{1t}\leq 0,-1<W_{2t}\leq 0\left|\digamma_t\right.\right).\end{aligned}$$

In the same way, we obtain

$$\begin{aligned}\Pr\left(y_{2t}-c_2>0\left|\digamma_t\right.\right) &= \Pr\left(W_{2t}>0\left|\digamma_t\right.\right)+\Pr\left(W_{1t}>0,-1<W_{2t}\leq 0\left|\digamma_t\right.\right)\\ &\quad +\left(1-\pi_d^E\right)\Pr\left(-1<W_{1t}\leq 0,-1<W_{2t}\leq 0\left|\digamma_t\right.\right)\end{aligned}$$

and in general we have

$$\begin{aligned}\Pr\left(y_{it}-c_i>0\left|\digamma_t\right.\right) &= \Pr\left(W_{it}>0\left|\digamma_t\right.\right)+\Pr\left(-1<W_{it}\leq 0,W_{jt}>0\left|\digamma_t\right.\right)\\ &\quad +\left(1-\pi_d^E\right)\Pr\left(-1<W_{it}\leq 0,-1<W_{jt}\leq 0\left|\digamma_t\right.\right)\end{aligned},$$

for $i,j=1,2$, $i\neq j$.

C Proofs

In this appendix we provide the proofs of Theorem 3 and Theorem 4. In what follows, $A\left(\mathbf{A}^0,a\right)$ denotes a generic open ball of radius a centred around $\mathbf{A}^0$.

C.1 Strong consistency of the single equation GIVE estimator

Given the definition of $\boldsymbol{\phi}_i$ in (27), define the vector of parameters $\boldsymbol{\theta}_{ij}$ as

$$\boldsymbol{\theta}_{ij}\equiv\left(\boldsymbol{\phi}_i,c_j\right),$$

with true value $\boldsymbol{\theta}_{ij}^0$ defined as

$$\boldsymbol{\theta}_{ij}^0\equiv\left(\boldsymbol{\phi}_i^0,c_j^0\right).$$

Denote

$$u_{it}\left(\boldsymbol{\theta}_{ij}\right)\equiv u_{it}=y_{it}-\mathbf{h}_{it}'\left(c_j\right)\boldsymbol{\phi}_i,$$

where $\mathbf{h}_{it}\left(c_j\right)$ is defined in (31). Consider the function

$$\begin{aligned}Q_i\left(\boldsymbol{\theta}_{ij}\right) &= E\left[\mathbf{w}_{it}'u_{it}\left(\boldsymbol{\theta}_{ij}\right)\right]E\left(\mathbf{w}_{it}\mathbf{w}_{it}'\right)E\left[\mathbf{w}_{it}u_{it}\left(\boldsymbol{\theta}_{ij}\right)\right]\\ &= E\left\{\left[1-\mathbf{I}\left(y_{jt}-c_j\right)\right]\mathbf{w}_{it}'u_{1it}\left(\boldsymbol{\theta}_{ij}\right)\right\}E\left(\mathbf{w}_{it}\mathbf{w}_{it}'\right)E\left\{\left[1-\mathbf{I}\left(y_{jt}-c_j\right)\right]\mathbf{w}_{it}u_{1it}\left(\boldsymbol{\theta}_{ij}\right)\right\}\\ &\quad +E\left[\mathbf{I}\left(y_{jt}-c_j\right)\mathbf{w}_{it}'u_{2it}\left(\boldsymbol{\theta}_{ij}\right)\right]E\left(\mathbf{w}_{it}\mathbf{w}_{it}'\right)E\left[\mathbf{I}\left(y_{jt}-c_j\right)\mathbf{w}_{it}u_{2it}\left(\boldsymbol{\theta}_{ij}\right)\right],\end{aligned}$$

where

$$u_{1it}\left(\boldsymbol{\theta}_{ij}\right)=y_{it}-\mathbf{h}_{1it}'\left(c_j\right)\boldsymbol{\phi}_{1i},\quad \mathbf{h}_{1it}\left(c_j\right)\equiv\left(\mathbf{z}_t',\mathbf{x}_{it}'\right)',\quad \boldsymbol{\phi}_{1i}\equiv\left(\boldsymbol{\delta}_i',\boldsymbol{\alpha}_i'\right)',$$

$$u_{2it}\left(\boldsymbol{\theta}_{ij}\right)=y_{it}-\mathbf{h}_{2it}'\left(c_j\right)\boldsymbol{\phi}_{2i},\quad \mathbf{h}_{2it}\left(c_j\right)\equiv\left(\mathbf{z}_t',\mathbf{x}_{it}',1\right)',\quad \boldsymbol{\phi}_{2i}=\boldsymbol{\phi}_i.$$

The function $Q_i\left(\boldsymbol{\theta}_{ij}\right)$ can be equivalently written as

$$Q_i\left(\boldsymbol{\theta}_{ij}\right) = E\left[g_{1it}\left(\boldsymbol{\theta}_{ij}\right)'\right]E\left(\mathbf{w}_{it}\mathbf{w}_{it}'\right)E\left[g_{1it}\left(\boldsymbol{\theta}_{ij}\right)\right]+E\left[g_{2it}\left(\boldsymbol{\theta}_{ij}\right)'\right]E\left(\mathbf{w}_{it}\mathbf{w}_{it}'\right)E\left[g_{2it}\left(\boldsymbol{\theta}_{ij}\right)\right],$$

where

$$g_{1it}\left(\boldsymbol{\theta}_{ij}\right)\equiv\left[1-\mathbf{I}\left(y_{jt}-c_j\right)\right]\mathbf{w}_{it}u_{1it}\left(\boldsymbol{\theta}_{ij}\right),\qquad g_{2it}\left(\boldsymbol{\theta}_{ij}\right)\equiv\mathbf{I}\left(y_{jt}-c_j\right)\mathbf{w}_{it}u_{2it}\left(\boldsymbol{\theta}_{ij}\right).$$

In order to prove Theorem 3 it is sufficient to prove the following lemma:

Lemma 5 *Consider the model in* (1) *under the conditions imposed in Theorem* 3*; then*

$$\lim_{b_{ij}\to 0}E\left[\sup_{\boldsymbol{\theta}_{ij}\in B\left(\boldsymbol{\theta}_{ij}^0,b_{ij}\right)}\left|g_{kit}\left(\boldsymbol{\theta}_{ij}\right)-g_{kit}\left(\boldsymbol{\theta}_{ij}^0\right)\right|\right]=0,\quad k=1,2.$$

Proof. Consider the case $k=1$; we then have

$$\begin{aligned}\left|g_{1it}\left(\boldsymbol{\theta}_{ij}\right)-g_{1it}\left(\boldsymbol{\theta}_{ij}^{0}\right)\right| &= \left|\left[1-\mathbf{I}\left(y_{jt}-c_j\right)\right]\mathbf{w}_{it}u_{1it}\left(\boldsymbol{\theta}_{ij}\right)-\left[1-\mathbf{I}\left(y_{jt}-c_j^0\right)\right]\mathbf{w}_{it}u_{1it}\left(\boldsymbol{\theta}_{ij}^{0}\right)\right| \\ &\leq \left|\left[1-\mathbf{I}\left(y_{jt}-c_j\right)\right]\mathbf{w}_{it}\left[u_{1it}\left(\boldsymbol{\theta}_{ij}\right)-u_{1it}\left(\boldsymbol{\theta}_{ij}^{0}\right)\right]\right| \\ &\quad+\left|\left\{\left[1-\mathbf{I}\left(y_{jt}-c_j\right)\right]-\left[1-\mathbf{I}\left(y_{jt}-c_j^0\right)\right]\right\}\mathbf{w}_{it}u_{1it}\left(\boldsymbol{\theta}_{ij}^{0}\right)\right| \\ &\leq \left[1-\mathbf{I}\left(y_{jt}-c_j\right)\right]\left|\mathbf{w}_{it}\right|\left|u_{1it}\left(\boldsymbol{\theta}_{ij}\right)-u_{1it}\left(\boldsymbol{\theta}_{ij}^{0}\right)\right| \\ &\quad+\mathbf{I}\left(\left|c_1-c_1^0\right|-\left|y_{jt}-c_j^0\right|\right)\left|\mathbf{w}_{it}\right|\left|u_{1it}\left(\boldsymbol{\theta}_{ij}^{0}\right)\right|.\end{aligned}$$

The case $k=2$ can be proved in a similar way, which completes the proof of Lemma 5. ■

Remark 6 *The sufficiency of the above lemma to prove Theorem* 3 *follows from Chan* (1993)*, whose Lemma* 1 *is analogous to Lemma* 5 *here.*

C.2 Strong consistency of the FIML estimator

Consider the vectors of parameters $\mathbf{c}$ and $\boldsymbol{\Delta}$ defined in (35), define the vector $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta}\equiv\left(\mathbf{c}',\boldsymbol{\Delta}'\right)',$$

and denote by $\mathbf{a}^0$ the true value of the generic parameters vector $\mathbf{a}$. Further, denote the normalising term p_t defined in (34) by

$$p_t\equiv p_t\left(\mathbf{c},\boldsymbol{\Delta}\right)\equiv p_t\left(\boldsymbol{\theta}\right).$$

Finally, denote by g_{kt} the contribution to the pdf defined in (33) under regime k, $k=1,2,3,4$. Therefore, $g_{1t}\left(\boldsymbol{\theta}\right)$ is defined as

$$g_{1t}\left(\boldsymbol{\theta}\right)\equiv\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1\left|F_t\right.\right)/p_t\left(\boldsymbol{\theta}\right)\right],$$

$\boldsymbol{\theta}_1$ being given in (32); in the same way $g_{kt}\left(\boldsymbol{\theta}\right)$ can be defined for $k=2,3,4$. Theorem 4 is proved by proving the following lemma:

Lemma 7 *Consider the model in* (1) *under the conditions imposed in Theorem* 4*; then*

$$E\left[g_{kt}\left(\boldsymbol{\theta}\right)\right]<E\left[g_{kt}\left(\boldsymbol{\theta}^0\right)\right],\quad\forall\boldsymbol{\theta}\neq\boldsymbol{\theta}^0,\quad k=1,2,3,4, \tag{C.1}$$

and

$$\lim_{b\to0}E\left[\sup_{\boldsymbol{\theta}\in B\left(\boldsymbol{\theta}^0,b\right)}\left|g_{kt}\left(\boldsymbol{\theta}\right)-g_{kt}\left(\boldsymbol{\theta}^0\right)\right|\right]=0,\quad k=1,2,3,4. \tag{C.2}$$

Proof. The result in (C.1) can be proved in two steps, namely

$$E\left[g_{kt}\left(\mathbf{c},\boldsymbol{\Delta}\right)\right]<E\left[g_{kt}\left(\mathbf{c},\boldsymbol{\Delta}^0\right)\right],\quad\forall\boldsymbol{\Delta}\neq\boldsymbol{\Delta}^0, \tag{C.1.1}$$

and

$$E\left[g_{kt}\left(\mathbf{c},\boldsymbol{\Delta}^0\right)\right]<E\left[g_{kt}\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right],\quad\forall\mathbf{c}\neq\mathbf{c}^0, \tag{C.1.2}$$

for $k=1,2,3,4$. Consider the case $k=1$; then (C.1.1) is equivalent to

$$\begin{aligned}&E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}\right)\right]\right\} \\ <\;&E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)\right]\right\},\end{aligned} \tag{C.1.1a}$$

since $\boldsymbol{\theta}_1^0\subset\boldsymbol{\Delta}^0$. From Jensen's inequality we have

$$\begin{aligned}&E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[\frac{f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}\right)}{f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)}\right]\right\} \\ <\;&\log E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\left[\frac{f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}\right)}{f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)}\right]\right\} \\ =\;&\log\int_{-\infty}^{c_1}\int_{-\infty}^{c_2}\frac{f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1\left|F_t\right.\right)}{p_t\left(\mathbf{c},\boldsymbol{\Delta}\right)}dy_{2t}dy_{1t} \\ <\;&0,\end{aligned}$$

so that (C.1.1a) holds. By applying the same procedure for $k = 2, 3, 4$, the result in (C.1.1) is shown to hold. Similarly, to prove (C.1.2) consider the case $k = 1$; therefore

$$\begin{aligned} & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)\right]\right\} \\ < \quad & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1^0\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2^0\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]\right\}, \end{aligned}$$

which is equivalent to

$$\begin{aligned} & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)\right]\right\} \\ - \quad & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]\right\} \\ < \quad & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1^0\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2^0\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]\right\} \\ - \quad & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]\right\}. \end{aligned} \tag{C.1.2a}$$

We then have

$$\begin{aligned} & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)\right]\right\} \\ - \quad & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]\right\} \\ = \quad & E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left\{\frac{\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)\right]}{\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]}\right\}\right\} \\ < \quad & \log E\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\frac{\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)\right]}{\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]}\right\} \\ = \quad & \log\int_{-\infty}^{c_1}\int_{-\infty}^{c_2}\frac{f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)}{p_t\left(\mathbf{c},\boldsymbol{\Delta}^0\right)}dy_{2t}dy_{1t} \\ < \quad & 0; \end{aligned}$$

further,

$$\begin{aligned} & \mathrm{E}\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]\right\} \\ - \quad & \mathrm{E}\left\{\left[1-\mathbf{I}\left(y_{1t}-c_1^0\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2^0\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)/p_t\left(\mathbf{c}^0,\boldsymbol{\Delta}^0\right)\right]\right\} \\ < \quad & 0 \end{aligned}$$

since $f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0\left|F_t\right.\right)$ is positive if and only if $y_{1t} \leq c_1^0$ and $y_{2t} \leq c_2^0$. Therefore, (C.1.2a) is fulfilled. Applying the same procedure for $k = 2, 3, 4$ the result in (C.1.2) is shown to hold: therefore, (C.1) also holds. The result in (C.2) is proved by noting that for $k = 1$

$$\begin{aligned}
\left|g_{1t}(\boldsymbol{\theta}) - g_{1t}\left(\boldsymbol{\theta}^0\right)\right| &= |[1-\mathbf{I}(y_{1t}-c_1)][1-\mathbf{I}(y_{2t}-c_2)]\log\left[f(y_{1t},y_{2t};\boldsymbol{\theta}_1|F_t)/p_t(\boldsymbol{\theta})\right] \\
&\quad - \left[1-\mathbf{I}\left(y_{1t}-c_1^0\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2^0\right)\right]\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0|F_t\right)/p_t\left(\boldsymbol{\theta}^0\right)\right]| \\
&\leq \left|[1-\mathbf{I}(y_{1t}-c_1)][1-\mathbf{I}(y_{2t}-c_2)]\log\left[\frac{f(y_{1t},y_{2t};\boldsymbol{\theta}_1|F_t)/p_t(\boldsymbol{\theta})}{f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0|F_t\right)/p_t\left(\boldsymbol{\theta}^0\right)}\right]\right| \\
&\quad + \left[\begin{array}{l} \left|\mathbf{I}(y_{1t}-c_1)-\mathbf{I}\left(y_{1t}-c_1^0\right)\right| + \left|\mathbf{I}(y_{2t}-c_2)-\mathbf{I}\left(y_{2t}-c_2^0\right)\right| \\ +\left|\mathbf{I}(y_{1t}-c_1)\mathbf{I}(y_{2t}-c_2)-\mathbf{I}\left(y_{1t}-c_1^0\right)\mathbf{I}\left(y_{2t}-c_2^0\right)\right| \end{array}\right] \\
&\quad \times \left|\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0|F_t\right)/p_t\left(\boldsymbol{\theta}^0\right)\right]\right| \\
&\leq |[1-\mathbf{I}(y_{1t}-c_1)][1-\mathbf{I}(y_{2t}-c_2)]|\left|\log\left[\frac{f(y_{1t},y_{2t};\boldsymbol{\theta}_1|F_t)/p_t(\boldsymbol{\theta})}{f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0|F_t\right)/p_t\left(\boldsymbol{\theta}^0\right)}\right]\right| \\
&\quad + \left[\begin{array}{l} \mathbf{I}\left(\left|c_1-c_1^0\right|-\left|y_{1t}-c_1^0\right|\right)+\mathbf{I}\left(\left|c_2-c_2^0\right|-\left|y_{2t}-c_2^0\right|\right) \\ +\mathbf{I}\left(\left[\left|c_1-c_1^0\right|-\left|y_{1t}-c_1^0\right|\right]\left[\left|c_2-c_2^0\right|-\left|y_{2t}-c_2^0\right|\right]\right) \end{array}\right] \\
&\quad \times \left|\log\left[f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1^0|F_t\right)/p_t\left(\boldsymbol{\theta}^0\right)\right]\right|.
\end{aligned}$$

A similar result holds for $k = 2, 3, 4$. Therefore, (C.2) is also verified. ■

Remark 8 *The result stated in* $(C.1)$ *is a standard condition in Maximum Likelihood estimation, and it is sufficient to ensure that the pdf is maximised at the true parameters values* $\boldsymbol{\theta}^0$. *The result in* $(C.2)$ *is equivalent to the one stated in Lemma* 1 *in Chan* (1993).

References

[1] Amemiya, T. (1977), "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model", *Econometrica*, 45(4), $955-968$.

[2] Andrews, D. W. K. and J. H. Stock (2005), "Inference with Weak Instruments", Cowles Foundation Discussion Paper No. 1530.

[3] Bae, K. H., G. A. Karolyi and R. M. Stulz (2003), "A New Approach to Measuring Financial Contagion", *The Review of Financial Studies*, 16(3), $717-763$.

[4] Billio, M., M. Lo Duca and L. Pelizzon (2005), "Contagion Detection with Switching Regimes Models: a Short and Long Run Analysis", GRETA, Working Paper 0501.

[5] Billio, M. and L. Pelizzon (2003), "Contagion and Interdependence in Stock Markets: Have They Been Misdiagnosed?", *Journal of Economics and Business*, $55(5-6)$, $405-426$.

[6] Blackburn, K. and M. Sola (1993), "Speculative Currency Attacks and Balance of Payments Crises", *Journal of Economic Surveys*, 7(2), $119-144$.

[7] Blundell R. and R. J. Smith (1994), "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables", *Journal of Econometrics*, $64(1-2)$, $355-373$.

[8] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31(3), $307-327$.

[9] Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return", *The Review of Economics and Statistics*, 69(3), $542-547$.

[10] Boyer, B. H., M. S. Gibson and M. Loretan (1999), "Pitfalls in Tests for Changes in Correlations", Federal Reserve Board International Finance Division, Working Paper 597R.

[11] Bowden, R. J. and D. A. Turkington (1981), "A Comparative Study of Instrumental Variables Estimators for Nonlinear Simultaneous Models", *Journal of the American Statistical Association*, 76(376), $988-995$.

[12] Bresnahan, T. F. and P. C. Reiss (1990), "Entry in Monopoly Markets", *Review of Economic Studies*, 57(4), $531-553$.

[13] Caner, M. and B. E. Hansen (2004), "Instrumental Variable Estimation of a Threshold Model", *Econometric Theory*, 20(5), $813-843$.

[14] Chan, K. S. (1993), "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model", *The Annals of Statistics*, 21(1), $520-533$.

[15] Chan, K. S. and H. Tong (1985), "On the Use of the Deterministic Lyapunov Function for the Ergodicity of Stochastic Difference Equations", *Advances in Applied Probability*, 17 (3), $666-678$.

[16] Corsetti, G., M. Pericoli and M. Sbracia (2005), "Some Contagion, Some Interdependence': More Pitfalls in Tests of Financial Contagion" *Journal of International Money and Finance*, 24(8), $1177-1199$.

[17] Davies, R. B. (1977), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative", *Biometrika*, 64(2), $247-254$.

[18] Davies, R. B. (1987), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative", *Biometrika*, 74(1), $33-43$.

[19] Dungey, M., R. Fry, B. González-Hermosillo and V. L. Martin (2005*a*), "Empirical Modelling of Contagion: a Review of Methodologies", *Quantitative Finance*, 5 (1), $9-24$.

[20] Dungey, M., R. Fry, B. González-Hermosillo and V. L. Martin (2005*b*), "A Comparison of Alternative Tests of Contagion with Applications", in Dungey, M. and D. Tambakis (eds), *Identifying International Financial Contagion: Progress and Challenges*, Oxford University Press, Oxford.

[21] Eichengreen, B., A. K. Rose and C. Wyplosz (1996), "Contagious Currency Crises: First Tests", *Scandinavian Journal of Economics*, 98(4), 463 – 484.

[22] Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity of the Variance of United Kingdom Inflation", *Econometrica*, 50(4), 987 – 1008.

[23] Esquivel, G. and F. Larraín B. (1998), "Explaining Currency Crises", mimeo, Harvard University.

[24] Favero, C. A. and F. Giavazzi (2002), "Is the International Propagation of Financial Shocks Non-Linear? Evidence from the ERM", *Journal of International Economics*, 57 (1), 231 – 246.

[25] Forbes, K. and R. Rigobon (2002), "No Contagion, Only Interdependence: Measuring Stock Market Comovements", *Journal of Finance*, 57(5), 2223 – 2261.

[26] Fratzscher, M. (2003), "On Currency Crises and Contagion", *International Journal of Finance and Economics*, 8 (2), 109 – 129.

[27] Girton, L. and D. Roper (1977), "A Monetary Model of Exchange Market Pressure Applied to the Postwar Canadian Experience", *The American Economic Review*, 67(4), 537 – 548.

[28] Gourieroux, C., J. J. Laffont and A. Monfort (1980), "Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes", *Econometrica*, 48(3), 675 – 696.

[29] Heckman, J. J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, 46(4), 931 – 959.

[30] Jeanne, O. and P. Masson (2000), "Currency Crises, Sunspots and Markov-Switching Regimes", *Journal of International Economics*, 50(2), 327 – 350.

[31] Kaminsky, G. L. and C. Reinhart (2002), "The Center and the Periphery: The Globalization of Financial Turmoil", NBER Working Paper 9479.

[32] Kelejian, H. H. (1971), "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables", *Journal of the American Statistical Association*, 66(334), 373 – 374.

[33] King, M. A. and S. Wadhwani (1990), "Transmission of Volatility Between Stock Markets", *The Review of Financial Studies*, 3(1), 5 – 33.

[34] Kooreman, P. (1994), "Estimation of Econometric Models of Some Discrete Games", *Journal of Applied Econometrics*, 9(3), 255 – 268.

[35] Kruger, M., P. N. Osakwe and J. Page (1998), "Fundamentals, Contagion and Currency Crises: An Empirical Analysis", Bank of Canada Working Paper 98 – 10.

[36] Kumar, M., U. Moorthy and W. Perraudin (2002), "Predicting Emerging Market Currency Crashes", IMF Working Paper 02/7.

[37] Laurent, S. and J. P. Peters (2005), "G@RCH 4.0, Estimation and Forecasting ARCH Models", *Timberlake Consultants*

[38] Lewbel, A. (2007), "Coherence and Completeness of Structural Models Containing a Dummy Endogenous Variable", *International Economic Review*, forthcoming.

[39] Loretan, M. and W. English (2000), "Evaluating "Correlation Breakdowns" During Periods of Market Volatility", Board of Governors of the Federal Reserve System, International Finance Discussion Paper No. 658.

[40] Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

[41] Martens, M. and S. Poon, (2001), "Returns Synchronization and Daily Correlation Dynamics Between International Stock Markets ", *Journal of Banking and Finance*, 25(10), 1805 − 1827.

[42] Massacci, D. (2007), "Econometric Analysis of Financial Contagion", PhD dissertation, University of Cambridge.

[43] Masson, P. (1999), "Contagion: Monsoonal Effects, Spillovers, and Jumps Between Multiple Equilibria", in Agenor, P. R., M. Miller, D. Vines and D. Weber (eds), *The Asian Financial Crisis: Causes, Contagion and Consequences,* Cambridge University Press, Cambridge.

[44] Newey, W. K. (1990), "Efficient Instrumental Variables Estimation of Nonlinear Models", *Econometrica*, 58(4), 809 − 837.

[45] Pearson, K. (1900), "On the Correlation of Characters not Quantitatively Measurable", *Royal Society Philosophical Transactions*, Series A, 195, 1 − 47.

[46] Pericoli, M. and M. Sbracia (2003), "A Primer on Financial Contagion", *Journal of Economic Surveys*, 17(4), 571 − 608.

[47] Pesaran, M. H. and A. Pick (2007), "Econometric Issues in the Analysis of Contagion", *Journal of Economic Dynamics and Control*, 31 (4), 1245 − 1277.

[48] Pesaran, M. H. and R. J. Smith (1994), "A Generalised R^2 Criterion for Regression Models Estimated by the Instrumental Variables Method", *Econometrica*, 62(3), 705 − 710.

[49] Pesaran, M. H. and A. Timmermann (1992), "A Simple Nonparametric Test of Predictive Performance", *Journal of Business and Economic Statistics*, 10(4), 461 − 465.

[50] Petruccelli, J. P. and S. W. Woolford (1984), "A Threshold AR(1) Model", *Journal of Applied Probability*, 21 (2), 270 − 286.

[51] Rangvid, J. (2001), "Second Generation Models of Currency Crises", *Journal of Economic Surveys*, 15 (5), 613 − 646.

[52] Rigobon, R. (2000), "A Simple Test for the Stability of Linear Models Under Heteroskedasticity, Omitted Variables, and Endogenous Variables Problems", mimeo, MIT.

[53] Rigobon, R. (2001), "Contagion: How To Measure It?", NBER Working Paper 8118.

[54] Rigobon, R. (2003*a*), "Identification Through Heteroskedasticity", *Review of Economics and Statistics*, 85 (4), 777 − 792.

[55] Rigobon, R. (2003*b*), "On the Measurement of the International Propagation of Shocks: Is the Transmission Stable?", *Journal of International Economics*, 61(2), 261 − 283.

[56] Ronn E., A. Sayrak and S. Tompaidis (2001), The Impact of Large Changes in Asset Prices on Intra Market Correlations in the Domestic and International Markets, mimeo, University of Texas, Austin.

[57] Stock, J. H., J. H. Wright and M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments", *Journal of Business and Economic Statistics*, 20(4), 518 − 529.

[58] Stone M. R. and M. Weeks (2001), "Systemic Financial Crises, Balance Sheets, and Model Uncertainty", IMF Working Paper 01/162.

[59] Tamer, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria", *Review of Economic Studies*, 70(1), 147 − 165.

[60] Tsay, R. S. (2005), *Analysis of Financial Time Series*, 2nd Edition, John Wiley & Sons, Inc., Hoboken, New Jersey.

[61] Tong, H. (1990), *Non-Linear Time Series: A Dynamical Systems Approach*, Oxford University Press, Oxford.

Table 1: Bias, RMSE, Size and Power in the Case of Experiment 1

	Bias					RMSE				
$\alpha_1 = \alpha_2 = 0.5$										
(π, T)	**50**	**100**	**200**	**500**	**1000**	**50**	**100**	**200**	**500**	**1000**
	FIML									
0.005	0.3365	0.0978	0.0407	0.0352	0.0090	1.1403	0.9474	0.8232	0.6806	0.5075
0.01	0.1906	0.0530	0.0309	0.0109	-0.0008	1.0172	0.8481	0.7190	0.5022	0.3258
0.05	0.0098	0.0084	-0.0002	0.0040	-0.0009	0.7458	0.5370	0.3500	0.2032	0.1422
0.10	-0.0188	-0.0001	0.0051	0.0016	-0.0021	0.5556	0.3837	0.2607	0.1588	0.1117
0.20	-0.0110	-0.0047	-0.0015	0.0021	0.0010	0.4322	0.3035	0.2093	0.1333	0.0939
0.30	-0.0211	-0.0063	-0.0071	-0.0028	0.0019	0.4114	0.2756	0.1927	0.1254	0.0871
0.40	-0.0189	-0.0112	-0.0106	-0.0045	0.0023	0.4027	0.2775	0.1913	0.1230	0.0858
0.50	-0.0233	-0.0040	-0.0036	-0.0022	-0.0001	0.4071	0.2778	0.1943	0.1215	0.0849
	GIVE, $m = 1$									
0.005	23.3750	5.0538	-1.3774	-5.4219	0.6535	860.7200	249.6600	87.9290	166.4500	508.6500
0.01	7.5300	-2.6424	0.7927	-1.3885	-0.1173	159.0100	115.6700	54.1780	54.5050	4.9310
0.05	27.0600	0.2628	-0.3140	-0.0448	-0.0219	1009.0000	32.6530	5.6176	1.3038	0.7818
0.10	-0.4459	-1.6106	-0.1570	-0.0227	-0.0115	31.7830	53.9930	2.8133	0.7032	0.4491
0.20	0.3352	0.0203	-0.0789	-0.0093	0.0070	65.8750	10.4750	0.7326	0.4214	0.2793
0.30	0.2083	-0.0919	-0.0604	-0.0062	-0.0054	15.3230	1.4431	0.5897	0.3420	0.2272
0.40	-0.2902	-0.0932	-0.0519	-0.0057	-0.0059	6.3756	1.2605	0.5162	0.3132	0.2079
0.50	0.0403	-0.0834	-0.0486	-0.0060	-0.0053	5.8288	1.5639	0.5091	0.3094	0.2059
	GIVE, $m = 6$									
0.005	1.5123	1.2912	1.1796	0.6736	0.5867	3.6240	4.8315	5.9986	7.3796	6.0088
0.01	1.2101	1.0030	0.9293	0.5643	0.3472	3.3271	4.3309	5.2081	4.4033	2.7120
0.05	0.8519	0.5646	0.3313	0.1807	0.1080	2.5011	2.0547	1.4033	0.9200	0.6378
0.10	0.6285	0.3964	0.2228	0.1187	0.0626	1.7170	1.2117	0.8830	0.5589	0.3986
0.20	0.4708	0.2808	0.1377	0.0731	0.0345	1.0363	0.8172	0.5768	0.3777	0.2653
0.30	0.4213	0.2410	0.1088	0.0536	0.0261	0.8877	0.6832	0.4859	0.3208	0.2202
0.40	0.3957	0.2296	0.0980	0.0497	0.0207	0.8446	0.6461	0.4524	0.2970	0.2026
0.50	0.3918	0.2182	0.0973	0.0477	0.0200	0.8429	0.6359	0.4471	0.2972	0.2002
$\alpha_1 = \alpha_2 = 1$										
	FIML									
0.005	0.3195	0.1408	0.0272	0.0187	0.0030	1.1148	0.9466	0.8193	0.6585	0.4613
0.01	0.2005	0.0498	0.0183	0.0321	-0.0090	1.0159	0.8181	0.7357	0.4733	0.2981
0.05	0.0065	-0.0087	-0.0030	0.0045	-0.0010	0.7250	0.4975	0.3186	0.1944	0.1370
0.10	0.0072	-0.0065	-0.0015	0.0029	-0.0012	0.5276	0.3578	0.2434	0.1488	0.1054
0.20	-0.0159	-0.0066	-0.0038	0.0024	-0.0003	0.4006	0.2745	0.1913	0.1202	0.0846
0.30	-0.0052	0.0017	0.0016	0.0009	0.0021	0.3658	0.2503	0.1815	0.1139	0.0768
0.40	-0.0107	-0.0034	-0.0042	0.0001	0.0014	0.3569	0.2374	0.1704	0.1084	0.0745
0.50	-0.0153	-0.0007	-0.0085	-0.0021	-0.0006	0.3602	0.2415	0.1709	0.1079	0.0748
	GIVE, $m = 1$									
0.005	21.4380	-0.7559	0.4001	0.4183	-0.2641	859.5400	65.2020	57.1470	30.3990	6.1158
0.01	1.3279	-0.9211	-1.0822	-0.3590	-0.0797	46.0970	83.6690	17.7820	6.7448	2.2480
0.05	20.9690	-0.1499	-0.1164	-0.0076	-0.0080	934.8900	7.7591	1.3300	0.7427	0.4837
0.10	-0.1055	-0.0948	-0.0499	-0.0006	-0.0027	5.7757	1.2589	0.6867	0.4168	0.2783
0.20	-0.0964	-0.0411	-0.0231	0.0024	-0.0009	1.0901	0.6372	0.4114	0.2578	0.1742
0.30	-0.0549	-0.0270	-0.0176	0.0023	-0.0007	0.7812	0.4951	0.3274	0.2088	0.1416
0.40	-0.0387	-0.0217	-0.0152	0.0025	-0.0003	0.6700	0.4460	0.2987	0.1903	0.1289
0.50	-0.0398	-0.0218	-0.0140	0.0024	-0.0003	0.6624	0.4397	0.2921	0.1872	0.1266
	GIVE, $m = 6$									
0.005	1.1439	0.9439	0.7949	0.3483	0.1124	2.4759	3.0206	3.6763	4.2947	2.7626
0.01	0.9248	0.7151	0.3537	0.2110	0.0902	2.2931	2.8683	3.2232	2.2001	1.1328
0.05	0.4628	0.2318	0.1040	0.0394	0.0244	1.8185	1.3823	0.8378	0.4971	0.3399
0.10	0.2766	0.1199	0.0492	0.0221	0.0135	1.1910	0.7574	0.5127	0.3226	0.2247
0.20	0.1665	0.0727	0.0226	0.0148	0.0074	0.7422	0.5100	0.3573	0.2254	0.1581
0.30	0.1285	0.0453	0.0153	0.0119	0.0048	0.6303	0.4259	0.3040	0.1933	0.1343
0.40	0.1176	0.0374	0.0115	0.0110	0.0040	0.5799	0.4030	0.2825	0.1812	0.1245
0.50	0.1194	0.0350	0.0099	0.0109	0.0032	0.5746	0.4028	0.2771	0.1794	0.1222

(Table 1 continued)

	Size (**5% level**, $H_0: \beta_1 = 0.00$)					**Power** (**5% level**, $H_0: \beta_1 = 0.50$)				
$\alpha_1 = \alpha_2 = 0.5$										
(π, T)	**50**	**100**	**200**	**500**	**1000**	**50**	**100**	**200**	**500**	**1000**
	FIML									
0.005	0.1385	0.0715	0.0505	0.0600	0.0555	0.1030	0.0985	0.1090	0.1775	0.2875
0.01	0.1020	0.0650	0.0600	0.0645	0.0565	0.1025	0.1105	0.1430	0.2815	0.4310
0.05	0.0730	0.0720	0.0630	0.0465	0.0525	0.1605	0.2235	0.3535	0.6830	0.9295
0.10	0.0645	0.0605	0.0565	0.0500	0.0520	0.1940	0.3060	0.5115	0.8775	0.9945
0.20	0.0575	0.0530	0.0540	0.0525	0.0515	0.2410	0.4115	0.6760	0.9665	0.9995
0.30	0.0665	0.0495	0.0495	0.0555	0.0435	0.2795	0.4410	0.7495	0.9835	1.0000
0.40	0.0660	0.0525	0.0510	0.0570	0.0530	0.2660	0.4780	0.7670	0.9865	1.0000
0.50	0.0655	0.0605	0.0590	0.0515	0.0465	0.2805	0.4655	0.7610	0.9880	1.0000
	GIVE, $m = 1$									
0.005	0.0055	0.0040	0.0030	0.0045	0.0080	0.0050	0.0025	0.0015	0.0030	0.0055
0.01	0.0065	0.0040	0.0035	0.0065	0.0155	0.0035	0.0015	0.0010	0.0065	0.0120
0.05	0.0075	0.0135	0.0175	0.0280	0.0330	0.0035	0.0065	0.0075	0.0295	0.0710
0.10	0.0150	0.0230	0.0280	0.0365	0.0360	0.0040	0.0105	0.0225	0.0765	0.1725
0.20	0.0255	0.0315	0.0350	0.0415	0.0385	0.0100	0.0165	0.0535	0.2020	0.4295
0.30	0.0315	0.0370	0.0350	0.0425	0.0385	0.0105	0.0225	0.0860	0.3020	0.5950
0.40	0.0340	0.0380	0.0365	0.0440	0.0390	0.0115	0.0310	0.1125	0.3700	0.6880
0.50	0.0360	0.0400	0.0370	0.0440	0.0405	0.0145	0.0300	0.1205	0.3775	0.6910
	GIVE, $m = 6$									
0.005	0.1095	0.0585	0.0330	0.0325	0.0355	0.0625	0.0335	0.0180	0.0190	0.0215
0.01	0.0830	0.0480	0.0295	0.0410	0.0395	0.0390	0.0195	0.0195	0.0200	0.0230
0.05	0.0670	0.0520	0.0490	0.0505	0.0450	0.0290	0.0200	0.0220	0.0385	0.0730
0.10	0.0800	0.0700	0.0580	0.0550	0.0515	0.0305	0.0240	0.0290	0.0685	0.1810
0.20	0.1075	0.0815	0.0520	0.0590	0.0485	0.0320	0.0285	0.0490	0.1735	0.4080
0.30	0.1090	0.0845	0.0560	0.0595	0.0460	0.0305	0.0310	0.0700	0.2685	0.5600
0.40	0.1070	0.0840	0.0615	0.0580	0.0435	0.0310	0.0345	0.0915	0.3235	0.6565
0.50	0.1105	0.0890	0.0625	0.0590	0.0410	0.0310	0.0345	0.0895	0.3360	0.6555
$\alpha_1 = \alpha_2 = 1$										
	FIML									
0.005	0.1440	0.0835	0.0665	0.0530	0.0480	0.1105	0.1015	0.1250	0.1530	0.2425
0.01	0.1140	0.0595	0.0655	0.0615	0.0480	0.1105	0.1020	0.1470	0.2320	0.4330
0.05	0.0765	0.0575	0.0585	0.0460	0.0475	0.1520	0.2305	0.3585	0.7195	0.9500
0.10	0.0610	0.0630	0.0520	0.0575	0.0490	0.2065	0.3245	0.5650	0.9195	0.9980
0.20	0.0615	0.0605	0.0470	0.0475	0.0420	0.2755	0.4595	0.7550	0.9885	1.0000
0.30	0.0640	0.0615	0.0560	0.0565	0.0465	0.2990	0.5220	0.8050	0.9945	1.0000
0.40	0.0640	0.0490	0.0530	0.0525	0.0495	0.3310	0.5675	0.8535	0.9975	1.0000
0.50	0.0660	0.0470	0.0520	0.0580	0.0440	0.3410	0.5545	0.8495	0.9985	1.0000
	GIVE, $m = 1$									
0.005	0.0090	0.0080	0.0060	0.0110	0.0200	0.0065	0.0055	0.0045	0.0105	0.0195
0.01	0.0110	0.0090	0.0070	0.0200	0.0305	0.0065	0.0040	0.0080	0.0185	0.0335
0.05	0.0165	0.0255	0.0285	0.0395	0.0410	0.0115	0.0215	0.0425	0.0930	0.1605
0.10	0.0240	0.0345	0.0380	0.0440	0.0410	0.0225	0.0405	0.0945	0.2165	0.4305
0.20	0.0320	0.0425	0.0410	0.0470	0.0425	0.0410	0.1035	0.2225	0.5070	0.8180
0.30	0.0415	0.0475	0.0445	0.0470	0.0420	0.0700	0.1585	0.3315	0.6780	0.9495
0.40	0.0430	0.0480	0.0425	0.0475	0.0425	0.0835	0.2005	0.3980	0.7710	0.9770
0.50	0.0465	0.0485	0.0440	0.0480	0.0425	0.0910	0.2070	0.4160	0.7805	0.9830
	GIVE, $m = 6$									
0.005	0.1310	0.0770	0.0495	0.0485	0.0415	0.0720	0.0390	0.0330	0.0355	0.0375
0.01	0.0975	0.0550	0.0450	0.0490	0.0475	0.0520	0.0275	0.0330	0.0385	0.0515
0.05	0.0645	0.0500	0.0470	0.0480	0.0455	0.0355	0.0370	0.0595	0.1365	0.2630
0.10	0.0695	0.0510	0.0455	0.0505	0.0480	0.0470	0.0595	0.1165	0.3060	0.5890
0.20	0.0640	0.0580	0.0450	0.0505	0.0495	0.0540	0.1125	0.2560	0.5895	0.8940
0.30	0.0645	0.0520	0.0495	0.0515	0.0490	0.0775	0.1590	0.3570	0.7285	0.9630
0.40	0.0665	0.0500	0.0465	0.0480	0.0455	0.0845	0.1845	0.4035	0.7950	0.9815
0.50	0.0705	0.0535	0.0460	0.0485	0.0385	0.0925	0.1890	0.4160	0.7980	0.9895

Notes: The DGP is $y^r_{it} = \delta_i + \alpha_i x^r_{it} + \beta_i \mathbf{I}\left(y^r_{jt} - c_j\right) + u^r_{it}$, $i, j = 1, 2$, $i \neq j$, with $\beta_1 = 0$, $\beta_2 = 0.2$ and $c_1 = c_2 = 1.64$; $\delta_1 = c_1 - \sqrt{\alpha_1^2 + 1}\left[\Phi^{-1}(1-\pi)\right]$, δ_2 is calibrated so to control for the unconditional probability of crisis π. $u^r_{it} = \left(\gamma_i f^r_t + \varepsilon^r_{it}\right)\left(\gamma_i^2 + 1\right)^{-1/2}$, where $\varepsilon^r_{it} \sim NID(0,1)$, $f^r_t \sim NID(0,1)$ and $\gamma_i \sim U(0.8, 1)$, γ_i fixed in repeated samples. Regressors generated as $x^r_{it} = \left(\phi_i h^r_t + q^r_{it}\right)\left(\phi_i^2 + 1\right)^{-1/2}$, where $q^r_{it} \sim NID(0,1)$, $h^r_t \sim NID(0,1)$ and $\phi_i \sim U(0.8, 1)$, ϕ_i fixed in repeated samples.

Table 2: Bias, RMSE, Size and Power in the Case of Experiment 2

	Bias					**RMSE**				
$\alpha_1 = \alpha_2 = 0.5$										
(π, T)	**50**	**100**	**200**	**500**	**1000**	**50**	**100**	**200**	**500**	**1000**
	FIML									
0.005	0.4310	0.2209	0.0928	0.0322	0.0208	1.1169	0.9104	0.8173	0.6992	0.5338
0.01	0.2294	0.0949	0.0072	0.0088	-0.0064	0.9769	0.8235	0.7510	0.5277	0.3444
0.05	0.0018	-0.0098	-0.0078	-0.0138	-0.0030	0.7671	0.5607	0.3624	0.2232	0.1533
0.10	-0.0067	0.0030	-0.0070	-0.0046	-0.0027	0.6022	0.3775	0.2597	0.1647	0.1198
0.20	0.0024	-0.0152	-0.0049	-0.0017	-0.0057	0.4739	0.3099	0.2163	0.1366	0.0960
0.30	-0.0135	0.0052	-0.0004	-0.0017	0.0005	0.4244	0.2904	0.1982	0.1290	0.0913
0.40	-0.0074	0.0011	-0.0075	-0.0055	0.0029	0.4167	0.2887	0.1907	0.1251	0.0877
0.50	-0.0088	0.0041	-0.0056	0.0013	0.0001	0.4089	0.2802	0.2020	0.1218	0.0870
	GIVE, $m = 1$									
0.005	9.6778	-4.8751	2.2369	-9.5134	-0.0638	352.3700	683.6100	111.7300	618.6900	29.2580
0.01	2.8382	1.7954	-4.4530	2.4026	-0.4521	105.3900	102.4400	223.7300	158.2600	5.5134
0.05	-0.6524	-0.5033	-2.1864	-0.5075	-0.0276	76.6690	54.9430	88.1270	17.0750	0.8826
0.10	-5.8597	3.9361	-0.1695	-0.0424	0.0087	244.8900	172.2500	1.6451	0.7411	0.4946
0.20	-0.2877	-0.0976	-0.0613	-0.0161	-0.0162	11.7670	2.1990	0.7608	0.4296	0.2939
0.30	-0.7818	-0.1109	-0.0537	-0.0147	-0.0088	21.2290	1.1750	0.6216	0.3445	0.2381
0.40	-0.3326	-2.8504	-0.0568	-0.0309	-0.0063	11.3580	123.2400	0.5294	0.3247	0.2184
0.50	-0.0290	-0.1392	-0.0697	-0.0139	-0.0082	6.8367	1.6120	0.5322	0.2979	0.1997
	GIVE, $m = 6$									
0.005	1.6760	1.5688	1.4124	1.0284	0.4905	3.6957	4.7504	6.1861	7.9464	6.4147
0.01	1.3449	1.2179	1.0248	0.6203	0.1280	3.3504	4.3551	5.4093	4.6099	2.9890
0.05	0.8261	0.7258	0.3617	0.1521	0.0952	2.5407	2.4675	1.7684	0.9891	0.6941
0.10	0.6668	0.4999	0.2600	0.1211	0.0705	1.7984	1.3279	0.9580	0.5995	0.4323
0.20	0.5311	0.3135	0.1488	0.0682	0.0251	1.1181	0.8355	0.6072	0.3855	0.2697
0.30	0.4276	0.2609	0.1291	0.0471	0.0201	0.9313	0.7118	0.5336	0.3231	0.2274
0.40	0.4340	0.2481	0.1056	0.0260	0.0188	0.8638	0.6572	0.4640	0.3024	0.2120
0.50	0.4012	0.2379	0.0915	0.0391	0.0162	0.8221	0.6383	0.4359	0.2833	0.1937
$\alpha_1 = \alpha_2 = 1$										
	FIML									
0.005	0.4534	0.1901	0.0931	-0.0083	-0.0066	1.1080	0.9208	0.8427	0.6561	0.4738
0.01	0.2641	0.0866	0.0144	0.0062	-0.0195	0.9666	0.8369	0.7356	0.4779	0.3140
0.05	0.0330	0.0103	-0.0065	-0.0016	-0.0090	0.7117	0.5286	0.3317	0.2061	0.1424
0.10	-0.0025	-0.0111	-0.0098	-0.0066	-0.0084	0.5389	0.3539	0.2437	0.1537	0.1068
0.20	-0.0146	0.0060	-0.0058	-0.0022	0.0008	0.4146	0.2881	0.1941	0.1237	0.0854
0.30	-0.0091	-0.0059	-0.0065	0.0033	-0.0026	0.3836	0.2532	0.1815	0.1147	0.0805
0.40	-0.0074	0.0006	-0.0025	-0.0020	-0.0005	0.3589	0.2543	0.1707	0.1050	0.0776
0.50	-0.0029	-0.0003	-0.0014	0.0012	-0.0018	0.3592	0.2475	0.1727	0.1095	0.0747
	GIVE, $m = 1$									
0.005	3.6306	0.6777	2.4918	-0.5485	-0.2554	90.0590	26.6220	108.7800	18.3800	8.3601
0.01	-3.7421	1.3135	0.1553	-0.2521	-0.1108	142.8100	45.0780	24.8280	6.9161	2.4211
0.05	-0.2076	-0.2253	-0.1438	-0.0233	-0.0338	29.9500	5.3762	2.1257	0.7764	0.5271
0.10	-0.2836	-0.0762	-0.0209	-0.0215	-0.0189	3.3814	1.2088	0.7330	0.4286	0.2945
0.20	-0.0469	-0.0354	-0.0278	-0.0070	0.0009	1.3700	0.6527	0.4324	0.2602	0.1781
0.30	-0.0384	-0.0195	0.0000	0.0020	-0.0051	1.0424	0.5014	0.3418	0.2099	0.1479
0.40	-0.2217	-0.0091	-0.0044	-0.0078	0.0008	10.2830	0.4544	0.3017	0.1916	0.1362
0.50	-0.0329	-0.0118	-0.0042	0.0004	-0.0002	0.6776	0.4174	0.2869	0.1807	0.1292
	GIVE, $m = 6$									
0.005	1.2721	1.1871	0.8028	0.4517	0.1461	2.4921	2.9440	3.7086	4.1390	3.2506
0.01	1.1110	0.7616	0.5858	0.1017	0.0410	2.5329	2.9238	3.5634	2.1502	1.1860
0.05	0.5461	0.2649	0.1070	0.0655	0.0161	1.8499	1.5439	0.8950	0.5232	0.3579
0.10	0.2438	0.1537	0.0510	0.0147	-0.0035	1.1606	0.7521	0.5385	0.3389	0.2348
0.20	0.2038	0.0605	0.0136	0.0099	0.0065	0.7782	0.5407	0.3562	0.2327	0.1600
0.30	0.1515	0.0559	0.0258	0.0120	0.0007	0.6280	0.4365	0.3098	0.1982	0.1376
0.40	0.1310	0.0490	0.0235	-0.0004	0.0028	0.5653	0.4200	0.2844	0.1828	0.1297
0.50	0.1279	0.0478	0.0214	0.0092	0.0035	0.5670	0.3886	0.2734	0.1758	0.1261

(Table 2 continued)

	Size (5% level, $H_0 : \beta_1 = 0.50$)					**Power** (5% level, $H_0 : \beta_1 = 1.00$)				
$\alpha_1 = \alpha_2 = 0.5$										
(π, T)	**50**	**100**	**200**	**500**	**1000**	**50**	**100**	**200**	**500**	**1000**
	FIML									
0.005	0.1060	0.0675	0.0615	0.0765	0.0870	0.0850	0.0775	0.1130	0.1905	0.2735
0.01	0.0770	0.0625	0.0710	0.0785	0.0720	0.0915	0.1145	0.1835	0.2725	0.4145
0.05	0.0780	0.0750	0.0635	0.0545	0.0520	0.1650	0.2300	0.3270	0.6630	0.9045
0.10	0.0765	0.0505	0.0425	0.0490	0.0580	0.2035	0.2745	0.5005	0.8610	0.9900
0.20	0.0790	0.0640	0.0580	0.0500	0.0490	0.2380	0.3900	0.6585	0.9600	1.0000
0.30	0.0660	0.0590	0.0445	0.0590	0.0500	0.2610	0.4245	0.7095	0.9775	1.0000
0.40	0.0670	0.0645	0.0410	0.0575	0.0585	0.2540	0.4355	0.7590	0.9805	1.0000
0.50	0.0710	0.0575	0.0565	0.0495	0.0525	0.2600	0.4450	0.7390	0.9845	1.0000
	GIVE, $m = 1$									
0.005	0.0075	0.0055	0.0055	0.0030	0.0050	0.0040	0.0025	0.0055	0.0040	0.0045
0.01	0.0030	0.0035	0.0040	0.0100	0.0130	0.0025	0.0010	0.0005	0.0090	0.0110
0.05	0.0100	0.0110	0.0150	0.0370	0.0355	0.0035	0.0035	0.0075	0.0290	0.0670
0.10	0.0150	0.0230	0.0280	0.0415	0.0495	0.0070	0.0050	0.0205	0.0735	0.1435
0.20	0.0210	0.0270	0.0385	0.0430	0.0405	0.0090	0.0150	0.0450	0.1870	0.4035
0.30	0.0320	0.0385	0.0385	0.0405	0.0440	0.0080	0.0185	0.0815	0.2820	0.5660
0.40	0.0430	0.0415	0.0360	0.0485	0.0495	0.0095	0.0275	0.1055	0.3820	0.6745
0.50	0.0355	0.0445	0.0335	0.0425	0.0395	0.0125	0.0245	0.1180	0.3845	0.7145
	GIVE, $m = 6$									
0.005	0.1170	0.0730	0.0355	0.0310	0.0375	0.0605	0.0405	0.0175	0.0200	0.0225
0.01	0.0850	0.0395	0.0415	0.0335	0.0400	0.0350	0.0160	0.0170	0.0190	0.0275
0.05	0.0630	0.0550	0.0390	0.0515	0.0610	0.0220	0.0210	0.0135	0.0365	0.0695
0.10	0.0780	0.0645	0.0635	0.0585	0.0520	0.0315	0.0140	0.0260	0.0635	0.1515
0.20	0.1045	0.0715	0.0665	0.0545	0.0465	0.0305	0.0250	0.0450	0.1665	0.4065
0.30	0.1065	0.0950	0.0705	0.0565	0.0470	0.0190	0.0280	0.0675	0.2490	0.5490
0.40	0.1190	0.0955	0.0690	0.0540	0.0525	0.0275	0.0330	0.0910	0.3375	0.6500
0.50	0.1165	0.0940	0.0590	0.0520	0.0405	0.0250	0.0275	0.0945	0.3415	0.6915
$\alpha_1 = \alpha_2 = 1$										
	FIML									
0.005	0.1195	0.0785	0.0690	0.0575	0.0500	0.0840	0.0815	0.1200	0.1845	0.2550
0.01	0.0880	0.0680	0.0690	0.0540	0.0550	0.0825	0.1080	0.1420	0.2430	0.4385
0.05	0.0695	0.0605	0.0520	0.0510	0.0510	0.1425	0.2220	0.3480	0.6830	0.9445
0.10	0.0655	0.0580	0.0575	0.0520	0.0495	0.1945	0.3280	0.5550	0.9150	0.9985
0.20	0.0610	0.0680	0.0615	0.0535	0.0515	0.2745	0.4390	0.7440	0.9840	1.0000
0.30	0.0735	0.0460	0.0520	0.0615	0.0485	0.3035	0.5280	0.8085	0.9915	1.0000
0.40	0.0715	0.0650	0.0510	0.0410	0.0530	0.3090	0.5505	0.8360	0.9965	1.0000
0.50	0.0725	0.0625	0.0530	0.0590	0.0485	0.3220	0.5635	0.8385	0.9965	1.0000
	GIVE, $m = 1$									
0.005	0.0120	0.0070	0.0060	0.0120	0.0300	0.0070	0.0060	0.0035	0.0075	0.0245
0.01	0.0120	0.0105	0.0105	0.0220	0.0300	0.0090	0.0045	0.0105	0.0205	0.0340
0.05	0.0200	0.0230	0.0245	0.0445	0.0470	0.0110	0.0160	0.0275	0.0960	0.1680
0.10	0.0270	0.0260	0.0475	0.0455	0.0450	0.0190	0.0300	0.0895	0.2140	0.4235
0.20	0.0385	0.0520	0.0420	0.0505	0.0430	0.0470	0.0955	0.2055	0.4995	0.7970
0.30	0.0415	0.0390	0.0470	0.0505	0.0485	0.0540	0.1375	0.2960	0.6700	0.9425
0.40	0.0400	0.0535	0.0420	0.0505	0.0590	0.0795	0.1935	0.3675	0.7705	0.9675
0.50	0.0445	0.0425	0.0385	0.0405	0.0495	0.0835	0.1890	0.4110	0.7895	0.9760
	GIVE, $m = 6$									
0.005	0.1485	0.0830	0.0540	0.0465	0.0510	0.0640	0.0455	0.0340	0.0265	0.0370
0.01	0.1070	0.0615	0.0410	0.0430	0.0550	0.0560	0.0355	0.0280	0.0400	0.0535
0.05	0.0715	0.0590	0.0420	0.0510	0.0455	0.0315	0.0370	0.0425	0.1250	0.2535
0.10	0.0570	0.0475	0.0530	0.0570	0.0475	0.0395	0.0385	0.1195	0.3020	0.5825
0.20	0.0785	0.0675	0.0455	0.0635	0.0460	0.0605	0.1065	0.2375	0.5735	0.8845
0.30	0.0695	0.0460	0.0520	0.0485	0.0425	0.0630	0.1420	0.3160	0.7105	0.9595
0.40	0.0610	0.0610	0.0460	0.0455	0.0555	0.0715	0.1770	0.3755	0.8000	0.9790
0.50	0.0635	0.0540	0.0480	0.0495	0.0510	0.0820	0.1780	0.4035	0.8120	0.9830

Notes: The DGP is based upon $y^r_{it} = \delta_i + \alpha_i x^r_{it} + \beta_i \mathbf{I}\left(y^r_{jt} - c_j\right) + u^r_{it}$, $i, j = 1, 2$, $i \neq j$, and is given by the reduced form in (19) with $\pi^E_d = 0.5$, $\beta_1 = 0.5$, $\beta_2 = 0.2$ and $c_1 = c_2 = 1.64$; δ_1 and δ_2 are is calibrated so to control for the unconditional probability of crisis π. $u^r_{it} = \left(\gamma_i f^r_t + \varepsilon^r_{it}\right)\left(\gamma^2_i + 1\right)^{-1/2}$, where $\varepsilon^r_{it} \sim NID(0,1)$, $f^r_t \sim NID(0,1)$ and $\gamma_i \sim U(0.8,1)$, γ_i fixed in repeated samples. Regressors generated as $x^r_{it} = \left(\phi_i h^r_t + q^r_{it}\right)\left(\phi^2_i + 1\right)^{-1/2}$, where $q^r_{it} \sim NID(0,1)$, $h^r_t \sim NID(0,1)$ and $\phi_i \sim U(0.8,1)$, ϕ_i fixed in repeated samples.

Table 3: Bias, RMSE, Size and Power in the Case of Experiment 3

panel 1)

	Bias					**RMSE**				
$\alpha_1 = \alpha_2 = 1$										
(π, T)	**50**	**100**	**200**	**500**	**1000**	**50**	**100**	**200**	**500**	**1000**
	FIML									
0.005	0.5698	0.3420	0.1488	0.0281	0.0071	1.0650	0.8936	0.7844	0.6398	0.5120
0.01	0.3769	0.1618	0.0898	0.0060	0.0009	0.9296	0.7975	0.7057	0.4978	0.3335
0.05	0.0512	0.0247	0.0003	-0.0022	-0.0028	0.7311	0.5405	0.3620	0.2152	0.1527
0.10	0.0168	-0.0095	-0.0084	-0.0034	-0.0015	0.5968	0.4004	0.2712	0.1737	0.1199
0.20	-0.0111	0.0059	0.0030	-0.0060	-0.0057	0.4823	0.3332	0.2374	0.1424	0.1029
0.30	-0.0092	-0.0197	0.0045	-0.0050	-0.0013	0.4568	0.3264	0.2284	0.1368	0.0975
0.40	-0.0054	-0.0062	0.0009	0.0018	-0.0011	0.4673	0.3143	0.2157	0.1366	0.0982
0.50	-0.0293	-0.0034	0.0027	0.0061	0.0015	0.4732	0.3187	0.2286	0.1423	0.1045

panel 2)

	Size (5% level, $H_0 : \beta_1 = 0.50$)					**Power (5% level, $H_0 : \beta_1 = 1.00$)**				
$\alpha_1 = \alpha_2 = 1$										
(π, T)	**50**	**100**	**200**	**500**	**1000**	**50**	**100**	**200**	**500**	**1000**
	FIML									
0.005	0.0975	0.0600	0.0550	0.0515	0.0870	0.0565	0.0540	0.0935	0.1840	0.2800
0.01	0.0690	0.0495	0.0655	0.0730	0.0680	0.0535	0.0840	0.1350	0.2620	0.3825
0.05	0.0800	0.0695	0.0580	0.0535	0.0500	0.1390	0.1945	0.3275	0.6395	0.9095
0.10	0.0740	0.0560	0.0545	0.0500	0.0490	0.1755	0.2915	0.4690	0.8345	0.9890
0.20	0.0670	0.0625	0.0635	0.0465	0.0485	0.2265	0.3440	0.6010	0.9430	0.9970
0.30	0.0680	0.0625	0.0630	0.0475	0.0440	0.2310	0.3940	0.6225	0.9540	0.9980
0.40	0.0760	0.0595	0.0485	0.0465	0.0505	0.2365	0.3750	0.6295	0.9555	1.0000
0.50	0.0775	0.0600	0.0660	0.0565	0.0655	0.2525	0.3725	0.6245	0.9425	0.9985

Notes: The DGP is based upon $y_{it}^r = \delta_i + \alpha_i x_t^r + \beta_i \mathbf{I}\left(y_{jt}^r - c_j\right) + u_{it}^r$, $i, j = 1, 2$, $i \neq j$, and is given by the reduced form in (19) with $\pi_d^E = 0.5$, $\beta_1 = 0.5$, $\beta_2 = 0.2$ and $c_1 = c_2 = 1.64$; δ_1 and δ_2 are is calibrated so to control for the unconditional probability of crisis π. $u_{it}^r = \left(\gamma_i f_t^r + \varepsilon_{it}^r\right)\left(\gamma_i^2 + 1\right)^{-1/2}$, where $\varepsilon_{it}^r \sim NID(0,1)$, $f_t^r \sim NID(0,1)$ and $\gamma_i \sim U(0.8, 1)$, γ_i fixed in repeated samples. Regressors generated as $x_t^r = \left(\phi h_t^r + q_t^r\right)\left(\phi^2 + 1\right)^{-1/2}$, where $q_t^r \sim NID(0,1)$, $h_t^r \sim NID(0,1)$ and $\phi \sim U(0.8, 1)$, ϕ_i fixed in repeated samples.

Table 4: Daily stock market returns. Period: 06/08/1990 to 30/06/2005.

panel a)

Descriptive statistics

Statistics	S&P 500	DAX 30	SMI	CAC 40
Mean	0.0323	0.0211	0.0311	0.0183
Median	0.0631	0.0742	0.0488	0.0498
Maximum	5.7708	7.1683	7.0489	10.3560
Minimum	-5.5327	-13.0580	-9.1340	-10.2510
Std. Dev.	1.0251	1.4215	1.1597	1.3371
Skewness	-0.0292	-0.3517	-0.1293	0.0130
Kurtosis	6.0354	7.4717	7.0317	7.4367
Jarque-Bera	1436.7**	3194**	2544.1**	3068.4**
	[0.0000]	[0.0000]	[0.0000]	[0.0000]

panel b)

Correlation matrix

	S&P 500	DAX 30	SMI	CAC 40
S&P 500	1.0000	0.5567	0.4720	0.5766
DAX 30		1.0000	0.7214	0.7588
SMI			1.0000	0.6890
CAC 40				1.0000

Notes: ** denotes significance at 1% level.

Table 5: Empirical application, FIML Estimation.
Period: 13/08/1990 to 30/06/2005.

	Market 1		Market 2
	DAX 30	**vs**	**S&P 500**
c_i	3.06		2.44
n_i	89		69
π_i	0.0238		0.0185
β_i	0.5823		0.5035
$\sigma^2_{u_i}$	0.9954		0.9754
$\rho_{u_1 u_2}$		0.4485	
$\log L$		-10137.40	
	SMI	**vs**	**S&P 500**
c_i	3.15		2.69
n_i	26		43
π_i	0.0070		0.0115
β_i	0.5731		0.7403
$\sigma^2_{u_i}$	0.9906		0.9877
$\rho_{u_1 u_2}$		0.3923	
$\log L$		-10250.70	
	CAC 40	**vs**	**S&P 500**
c_i	2.20		2.69
n_i	164		43
π_i	0.0439		0.0115
β_i	0.6438		0.3412
$\sigma^2_{u_i}$	0.9893		0.9628
$\rho_{u_1 u_2}$		0.5130	
$\log L$		-9955.72	

Notes: The table reports the results of the estimation of the model in (40) by FIML. c_i denotes the threshold parameter; n_i the number of crises periods; π_i the proportion of crisis periods; β_i the contagion coefficient; $\sigma^2_{u_i}$ the variance of u_i and $\rho_{u_1 u_2}$ the correlation between u_1 and u_2; $\log L$ the value of the log-likelihood function