Modelling Contagion in a Multiple Equilibria Setting: an Application to Stock Market Returns*

Daniele Massacci†
Faculty of Economics
University of Cambridge

Preliminary and incomplete draft: not for circulation or quotation.

August 10, 2006

Abstract

This paper deals with estimation and inference in the incoherent econometric model of contagion of Pesaran and Pick (2006). The Likelihood function is obtained. By Monte Carlo simulation the resulting Maximum Likelihood estimator is shown to perform better than the GIVE estimators proposed by Pesaran and Pick (2006). Testing for contagion turns out to be nonstandard, as a nuisance parameter is not identified under the null of no contagion. An empirical application to stock market returns shows that the New York Stock Exchange seems to be unaffected by a crisis taking place in any of the major European stock markets.

JEL classification: C10, C12, C13, C32, G15.

Keywords: Contagion, Estimation, Hypothesis Testing, Coherent Models, Threshold Models.

*I would like to thank Hashem Pesaran for his extremely helpful comments. This paper also benefits from conversations with Alessio Sancetta and Takashi Yamagata. Constructive comments from participants at the Econometrics seminars, Faculty of Economics, University of Cambridge have been most helpful. Errors and omissions are my own responsability.

†dm355@cam.ac.uk

1 Introduction

The financial system has always been characterised by a high degree of instability. Episodes of distress appear systematically and seem to have become more frequent over the last few years. Famous recent examples are the crises that hit the US stock market in 1987, the European Monetary System in 1992-1993, Mexico in December 1994, Asia in 1997, Russia in August 1998 and Argentina in December 2001.

One of the most important features of recent episodes of financial turmoil is that they tend to appear in clusters. Economists have paid attention to this feature only since the Asian crises, when they started adopting the term *contagion*: intuitively, the idea of contagion suggests a pathological situation where the degree of dependence amongst markets increases when a financial crisis occurs in any of them.

The issue of contagion in financial markets is of particular interest for investors. From standard portfolio theory we know that the exposition to market risk can be reduced by diversifying the investment portfolio, for instance by investing in more than one market. However, if contagion does occur in periods of financial turmoil then portfolio diversification may not be as effective as it would be in the absence of contagion.

There now exists a large body of empirical literature trying to assess the presence of contagion in financial markets, as surveyed by Dungey *et al.* (2005). However, many of these studies make use of unsuitable econometric models, and the results they obtain are likely to be misleading. An exception is the work by Pesaran and Pick (2006), where a canonical model of contagion is proposed; however, the issue of estimation and inference is left incomplete. The aim of this paper is therefore to fill this gap, so to provide a complete theoretical framework for the analysis of contagion.

The paper is organised as follows. Section 2 provides a review of the relevant literature. Section 3 describes the canonical model of contagion proposed in Pesaran and Pick (2006), while the solution is provided in Section 4. Section 5 discusses estimation of the model, obtained by GIVE and Maximum Likelihood estimation, while Monte Carlo evidence is presented in Section 6. Section 7 deals with the issue of testing for contagion. An empirical application to stock market returns is provided in Section 8. Finally Section 9 presents the conclusions. **Add directions for future work**

2 Review of the literature

2.1 Theoretical models of financial crises

The theoretical literature on financial crises starts with the first generation models of currency crises[1]. In this stream of literature, a speculative attack to a country's currency occurs because the government is no longer able to

[1] The seminal paper of first generation models is Krugman (1979).

defend a fixed exchange rate due to an inconsistent economic policy: therefore, in this case a currency crisis is seen as unavoidable. First generation models therefore provide a plausible explanation for the failure of stabilization policies in Latin America during the 1970s and 1980s; however, they are not able to fully explain the collapse of the European Monetary System in 1992-93: indeed, while countries like Italy and Spain were running inconsistent monetary and fiscal policies, this was not the case of France and Britain.

While in first generation models a currency crisis is unavoidable, in second generation models speculative attacks are self-fulfilling and a crisis results from the influence of expectations upon macroeconomic policy decisions[2]. In other words, a speculative attack occurs because economic agents expect them to arise, even if the maintenance of the peg would be consistent with the state of economic fundamentals. A key feature of second generation models is that they exhibit multiple equilibria: for a given state of economic fundamentals, a speculative attack may or may not occur depending on agents' expectations. Further, a shift in expectations may move the system from a tranquil state to a crisis scenario. Second generation models therefore offer a plausible explanation of why France and Britain had to leave the European Monetary System. However, they focus on a single economy, so they are unable to explain why financial crises tend to cluster together.

According to Masson (1998) the simultaneous occurrence of financial crises can be caused by 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 determine a cluster of crises. *Spillovers* are driven by the correlation between external economic linkages; for example, if two countries have trade linkages and one of them is hit by a crisis and has to devalue, then the other one will be forced to devalue in order to keep its level of competitiveness. Finally, *pure contagion* occurs when a crisis spreads from one country to another without any change in macroeconomic fundamentals; it can therefore be modelled as a situation characterised by the presence of multiple equilibria, where the economic system shifts from one equilibrium to another one. Following Pesaran and Pick (2006), we will simply refer to pure contagion as *contagion*, while monsoonal effects and spillovers will be jointly referred to as *inter-dependence effects*. In this set up contagion can be associated to a system affected by financial fragility, where the banking system faces the possibility of a short term liquidity shortage, which may lead investors to generate bank runs[3].

2.2 Empirical models of financial contagion

In the empirical literature two approaches have become particularly popular in assessing the presence of contagion: the first one is based upon measuring the

[2] The seminal paper of second generation models is Obstfeld (1986).

[3] The seminal paper of models of financial fragility is Diamond and Dybvig (1983).

probability of crisis in one market, conditional upon a crisis in another market; the second one is based upon the estimation of first and/or second moments.

According to the probability approach, contagion occurs if a crisis in one market increases the likelihood of a crisis in another market above the level that would be implied by the interdependence between the two markets. This approach has been mainly used in the context of currency crises, and has become popular after the work by Eichengreen *et al.* (1996). The empirical studies make use of a pooled probit model, where each equation corresponds to a market at a given time period, and a unit value of the dependent variable corresponds to a situation of crisis. In order to capture possible contagion effects, a dummy variable is included as a covariate, which takes value equal to unity if any other market in the sample is in crisis. By treating the contagion dummy as exogenous, standard Maximum Likelihood estimation is applied. However, the contagion dummy is actually endogenous; the resulting the Maximum Likelihood estimator is then inconsistent and statistical inference based upon it provides misleading results.

The moments-based approach has been mainly used in the context of stock markets crises, and defines contagion as an increase in the correlation between markets during crisis periods compared to tranquil periods. It has become popular after the work by Forbes and Rigobon (2002). The empirical studies are generally based upon a bivariate VAR model with no equation specific explanatory variables, so that estimation of the structural parameters and identification of contagion effects from interdependence is not possible. The identification issue is then circumvented by *a priori* identifying the market where the crisis starts; the correlation coefficients of the VAR residuals for the whole sample and the crisis periods are then compared. Although the endogeneity issue is circumvented, this methodology suffers from endogeneity bias, since it requires *a priori* identification of the crisis periods.

3 Model

Consider the following model introduced in Pesaran and Pick (2006) for a two-market set up

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

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 common explanatory variables; both the $\mathbf{x}_{it}$ and $\mathbf{z}_t$ are assumed to be predetermined and independent upon the shocks u_{1t} and u_{2t}. The indicator function $\mathbf{I}\left(A\right)$ takes value equal to unity if $A > 0$ and zero otherwise; c_1 and c_2 are non-negative thresholds. The error terms u_{1t} and u_{2t} are assumed to be serially uncorrelated and such that

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} |\mathbf{z}_t, \mathbf{x}_{1t}, \mathbf{x}_{2t} \sim IID\left[\mathbf{0}, \mathbf{\Sigma}_u\right],$$

where the symmetric and positive definite covariance matrix $\mathbf{\Sigma}_u$ is defined as

$$\mathbf{\Sigma}_u \equiv \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 \end{bmatrix}$$

so that the constant correlation coefficient between u_{1t} and u_{2t} is defined as $\rho_{u_1 u_2} = (\sigma_{u_1 u_2} / \sigma_{u_1} \sigma_{u_2})$. Finally, the slope coefficients β_1 and β_2 are assumed to be non-negative[4].

In the set up given by (1) and (2) a crisis in market i occurs whenever the dependent variable y_{it} is strictly greater than the corresponding threshold c_i: in other words, a crisis is associated to an *extreme positive value* of y_{it}. The contagion literature provides several examples for the performance indicator y_{it}. In analysing currency crises, Eichengreen *et al.* (1996) use the index of exchange market pressure first introduced by Girton and Roper (1977): this is obtained as 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[5]. Favero and Giavazzi (2002) deploy the spreads between 3-months German rate and 3-months interest rate in other Exchange Rate Mechanism countries. Finally, in analysing 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.

In the model given by (1) and (2), interdependence between markets 1 and 2 is captured by the non-zero value of the correlation coefficient $\rho_{u_1 u_2}$; conversely, contagion from market j to market i is said to occur if

$$\Pr\left[y_{it} > c_i \left| y_{jt} > c_j ; \mathbf{z}_t, \mathbf{x}_{it}\right.\right] > \Pr\left[y_{it} > c_i \left| y_{jt} \leq c_j ; \mathbf{z}_t, \mathbf{x}_{it}\right.\right], \quad i, j = 1, 2, \quad i \neq j,$$

meaning that contagion is reflected in a positive value of β_i: therefore, in this paper a probability-based definition of crisis is adopted. It is important to note that interdependence is the result of normal market interaction, whereas contagion only takes place in time of crisis. Formally, testing for contagion from market j to market i results in testing the null hypothesis $\beta_i = 0$ against the one-sided alternative $\beta_i > 0$.

The model in (1) and (2) represents a two-equation nonlinear simultaneous equations model, which has been extensively studied in the econometric literature, as summarised in Amemiya (1985). More precisely, it is an example of a

[4]From a theoretical point of view, β_1 and β_2 can take any real value. However, given the economic application of the model, they are assumed to be non-negative. An analysis of the model for cases where β_1 and/or β_2 can be negative is provided in Appendix A.

[5]Esquivel and Larrain (1998) define the market pressure index in the same way as Eichengreen et al. (1996). Coversely, Kruger et al. (1998) and Stone ane Weeks (2002) exclude interest rates since they are not market-determined in developing countries.

simultaneous equations model with endogenous switching regimes: the model is piecewise linear, where the shifts in the intercepts are driven by an endogenous dummy variable equal to unity whenever the associated endogenous continuous variable crosses the threshold. In this sense the model relates to the work by Heckman (1978), Maddala (1983) and more recently Tamer (2003), which all study the issue of endogenous switching in the context of simultaneous equations models[6]. Those models however differ from that in (1) and (2) in taking the threshold parameters as known (and generally equal to zero). The issue of estimation of threshold parameters is discussed in the time series literature following the work by Tong (1990), which introduces the threshold principle to nonlinear time series models: however, in this context the shift in the regime is determined by exogenous variables; for a survey about inference in TAR models see also Hansen (1997).

4 Solving the model

Define

$$w_{it} = \boldsymbol{\delta}_i' \mathbf{z}_t + \boldsymbol{\alpha}_i' \mathbf{x}_{it} + u_{it}, \tag{3}$$

so that (1) and (2) become

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

Two possible cases then arise. If β_1 or β_2 (or both) is equal to zero the system has a unique solution. For example, if $\beta_1 = 0$ the system becomes triangular and the solution is simply given by

$$\begin{aligned} y_{1t} &= w_{1t} + \beta_1 \mathbf{I}\left(y_{2t} - c_2\right), \\ y_{2t} &= w_{2t}. \end{aligned}$$

If β_1 and β_2 are both strictly positive, the system can be more easily solved by defining the following normalised variables

$$Y_{it} = \frac{y_{it} - c_i}{\beta_i}, \quad W_{it} = \frac{w_{it} - c_i}{\beta_i}, \quad i = 1, 2; \tag{4}$$

in this way the system can be equivalently written as

$$Y_{1t} = W_{1t} + \mathbf{I}\left(Y_{2t}\right), \tag{5}$$
$$Y_{2t} = W_{2t} + \mathbf{I}\left(Y_{1t}\right). \tag{6}$$

From (5) and (6), depending on the sign of the dependent variables Y_{1t} and Y_{2t} four possible cases arise, each of them corresponding to a region in the

[6] These models also assume that *at least* one continuous endogenous variable is latent, while in the model in (1) and (2) *both* endogenous variables are observable.

(W_{1t}, W_{2t}) plane:

$$\begin{array}{ll} \text{Case 1: } Y_{1t} \leq 0, Y_{2t} \leq 0, & \text{Case 2: } Y_{1t} \leq 0, Y_{2t} > 0, \\ \text{Case 3: } Y_{1t} > 0, Y_{2t} \leq 0, & \text{Case 4: } Y_{1t} > 0, Y_{2t} > 0, \end{array} \tag{7}$$

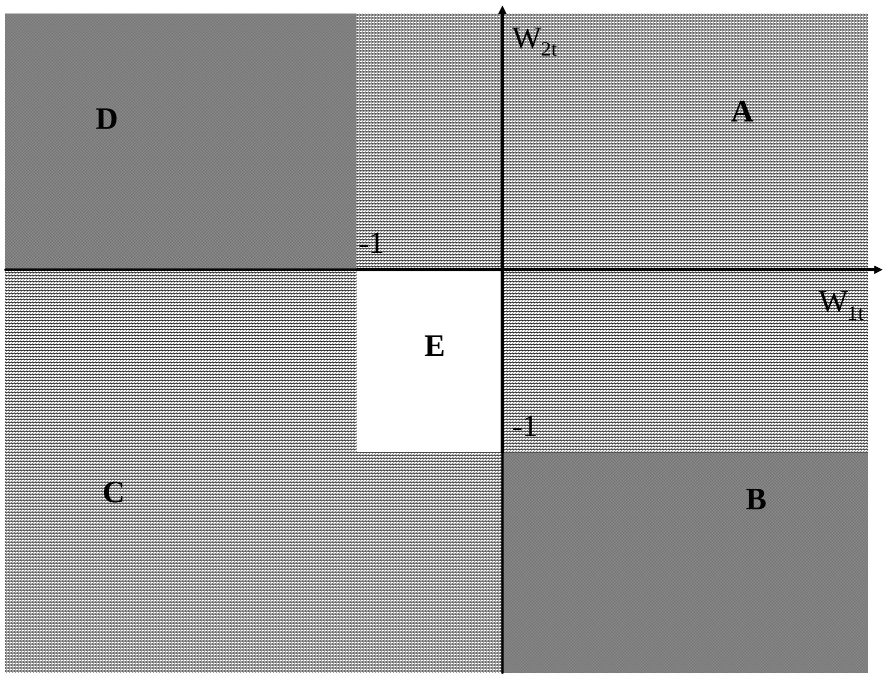
where $Y_{it} > 0$ corresponds to a crisis in market i at time t, whereas $Y_{it} \leq 0$ denotes a tranquil period. The cases described in (7) correspond to the following values for W_{1t} and W_{2t}

$$\begin{array}{ll} \text{Case 1: } W_{1t} \leq 0, W_{2t} \leq 0, & \text{Case 2: } W_{1t} \leq -1, W_{2t} > 0, \\ \text{Case 3: } W_{1t} > 0, W_{2t} \leq -1, & \text{Case 4: } W_{1t} > -1, W_{2t} > -1. \end{array} \tag{8}$$

The four combinations of values for W_{1t} and W_{2t} described in (8) give rise to the following mutually exclusive solution regions in the (W_{1t}, W_{2t}) space:

$$\begin{array}{ll} \text{Region A: } \left\{ \begin{array}{l} W_{1t} > -1 \\ W_{2t} > 0 \end{array} \right. \cup \left\{ \begin{array}{l} W_{1t} > 0 \\ -1 < W_{2t} \leq 0 \end{array} \right. \\ \text{Region B: } \left\{ \begin{array}{l} W_{1t} > 0 \\ W_{2t} \leq -1 \end{array} \right. \\ \text{Region C: } \left\{ \begin{array}{l} W_{1t} \leq -1 \\ -1 < W_{2t} \leq 0 \end{array} \right. \cup \left\{ \begin{array}{l} W_{1t} \leq 0 \\ W_{2t} \leq -1 \end{array} \right. \\ \text{Region D: } \left\{ \begin{array}{l} W_{1t} \leq -1 \\ W_{2t} > 0 \end{array} \right. \\ \text{Region E: } \left\{ \begin{array}{l} -1 < W_{1t} \leq 0 \\ -1 < W_{2t} \leq 0 \end{array} \right. \end{array} \tag{9}$$

The solution regions A to E in (9) are summarised in the following figure:



In terms of the normalised variables Y_{it} and W_{it} defined in (4) the complete solution to the model can be written as:

$$\begin{cases} Y_{1t} = W_{1t} + 1 > 0 \\ Y_{2t} = W_{2t} + 1 > 0 \end{cases} \qquad \text{(Region A)}$$

$$\begin{cases} Y_{1t} = W_{1t} > 0 \\ Y_{2t} = W_{2t} + 1 \leq 0 \end{cases} \qquad \text{(Region B)}$$

$$\begin{cases} Y_{1t} = W_{1t} \leq 0 \\ Y_{2t} = W_{2t} \leq 0 \end{cases} \qquad \text{(Region C)} \qquad (10)$$

$$\begin{cases} Y_{1t} = W_{1t} + 1 \leq 0 \\ Y_{2t} = W_{2t} > 0 \end{cases} \qquad \text{(Region D)}$$

$$\begin{cases} Y_{1t} = W_{1t} \leq 0 \\ Y_{2t} = W_{2t} \leq 0 \end{cases} \cup \begin{cases} Y_{1t} = W_{1t} + 1 > 0 \\ Y_{2t} = W_{2t} + 1 > 0 \end{cases} \qquad \text{(Region E)}$$

From (10), the model has a unique solution in regions A, B, C, and D, whereas multiple solutions arise in region E: this is because region E is generated by the intersection between the region where Y_{1t} and Y_{2t} are both less than or equal to zero and the region where Y_{1t} and Y_{2t} are both positive (Case 1 and

4 respectively in (7) and (8)). Pesaran and Pick (2006) model the solutions in region E as the outcome of a randomisation process d_t defined as

$$d_t \sim Bernoulli\left(\pi_d\right), \tag{11}$$

where π_d is the *unknown* probability of observing $d_t = 1$: therefore, assuming that π_d represents the probability of observing the favourable non-crisis equilibrium $Y_{it} = W_{it}$, from (10) the solution in region E can be written as

$$Y_{it} = d_t W_{it} + (1 - d_t)(1 + W_{it}) = 1 + W_{it} - d_t. \tag{12}$$

Taking into account (3), (4), (10) and (12), the reduced form for the original model in (1) and (2) is given by

$$\begin{aligned}
&\begin{cases} y_{1t} = w_{1t} + \beta_1 > c_1 \\ y_{2t} = w_{2t} + \beta_2 > c_2 \end{cases} && \text{(Region A)} \\
&\begin{cases} y_{1t} = w_{1t} > c_1 \\ y_{2t} = w_{2t} + \beta_2 \leq c_2 \end{cases} && \text{(Region B)} \\
&\begin{cases} y_{1t} = w_{1t} \leq c_1 \\ y_{2t} = w_{2t} \leq c_2 \end{cases} && \text{(Region C)} \\
&\begin{cases} y_{1t} = w_{1t} + \beta_1 \leq c_1 \\ y_{2t} = w_{2t} > c_2 \end{cases} && \text{(Region D)} \\
&\begin{cases} y_{1t} = w_{1t} + (1 - d_t)\beta_1 \\ y_{2t} = w_{2t} + (1 - d_t)\beta_2 \end{cases} && \text{(Region E)}
\end{aligned} \tag{13}$$

Defining the information set $\digamma_t$ as

$$\digamma_t = \left(\mathbf{z}_t^{'}, \mathbf{x}_{1t}^{'}, \mathbf{x}_{2t}^{'}\right)^{'},$$

the probabilities of *each* of the four events in (7) are given by

$$\begin{aligned}
\Pr\left[y_{1t} \leq c_1, y_{2t} \leq c_2 \left| \digamma_t \right.\right] &= \Pr\left[C \left| \digamma_t \right.\right] + \Pr\left[E \left| \digamma_t \right.\right], \\
\Pr\left[y_{1t} \leq c_1, y_{2t} > c_2 \left| \digamma_t \right.\right] &= \Pr\left[D \left| \digamma_t \right.\right], \\
\Pr\left[y_{1t} > c_1, y_{2t} \leq c_2 \left| \digamma_t \right.\right] &= \Pr\left[B \left| \digamma_t \right.\right], \\
\Pr\left[y_{1t} > c_1, y_{2t} > c_2 \left| \digamma_t \right.\right] &= \Pr\left[A \left| \digamma_t \right.\right] + \Pr\left[E \left| \digamma_t \right.\right].
\end{aligned} \tag{14}$$

As a consequence, from (14) the probability q_t of *any* of the events in (7) is

$$\begin{aligned}
q_t &= \Pr\left[y_{1t} \leq c_1, y_{2t} \leq c_2 \left| \digamma_t \right.\right] + \Pr\left[y_{1t} \leq c_1, y_{2t} > c_2 \left| \digamma_t \right.\right] \\
&\quad + \Pr\left[y_{1t} > c_1, y_{2t} \leq c_2 \left| \digamma_t \right.\right] + \Pr\left[y_{1t} > c_1, y_{2t} > c_2 \left| \digamma_t \right.\right] \\
&= 1 + \Pr\left[E \left| \digamma_t \right.\right] \\
&> 1,
\end{aligned} \tag{15}$$

meaning that the sum of the probabilities of the four events is greater than unity. The conditional probability of being in region E is

$$\begin{aligned}\Pr\left[E\,|\mathcal{F}_t\right] &= F_{12}\left(\frac{c_1-\boldsymbol{\delta}_1'\mathbf{z}_t-\boldsymbol{\alpha}_1'\mathbf{x}_{1t}}{\sigma_1},\frac{c_2-\boldsymbol{\delta}_2'\mathbf{z}_t-\boldsymbol{\alpha}_2'\mathbf{x}_{2t}}{\sigma_2}\right)\\ &\quad -F_{12}\left(\frac{c_1-\boldsymbol{\delta}_1'\mathbf{z}_t-\boldsymbol{\alpha}_1'\mathbf{x}_{1t}}{\sigma_1},\frac{c_2-\boldsymbol{\delta}_2'\mathbf{z}_t-\boldsymbol{\alpha}_2'\mathbf{x}_{2t}-\beta_2}{\sigma_2}\right)\\ &\quad -F_{12}\left(\frac{c_1-\boldsymbol{\delta}_1'\mathbf{z}_t-\boldsymbol{\alpha}_1'\mathbf{x}_{1t}-\beta_1}{\sigma_1},\frac{c_2-\boldsymbol{\delta}_2'\mathbf{z}_t-\boldsymbol{\alpha}_2'\mathbf{x}_{2t}}{\sigma_2}\right)\\ &\quad +F_{12}\left(\frac{c_1-\boldsymbol{\delta}_1'\mathbf{z}_t-\boldsymbol{\alpha}_1'\mathbf{x}_{1t}-\beta_1}{\sigma_1},\frac{c_2-\boldsymbol{\delta}_2'\mathbf{z}_t-\boldsymbol{\alpha}_2'\mathbf{x}_{2t}-\beta_2}{\sigma_2}\right),\end{aligned} \tag{16}$$

F_{12} being the conditional joint probability distribution function of the error terms u_{1t} and u_{2t}.

The model in (1) and (2) 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 the shock u_{it} and the related dependent variable y_{it} for given values of the exogenous variables $\mathbf{z}_t$ and $\mathbf{x}_{it}$. This is clearly not the case for the model in (1) and (2), due to the randomisation process d_t arising in region E, which leads to the existence of multiple equilibria[7].

Incoherent models have been widely studied in the econometric literature, as summarised in Chapter 5 of Maddala (1983). However, for identification and estimation purposes, most of the literature has been concerned with imposing the well-known "coherency" conditions; see for example Heckman (1978) and Gourieroux et al. (1980). In the context of the model in (1) and (2) this means imposing the condition $\beta_1\beta_2 = 0$.

Some attempts have been made to avoid imposing the coherency condition. For example, Kooreman (1994) considers the model in (1) and (2) where the dependent variables y_{it} are latent rather than observable, and the model itself becomes a multivariate probit. The model is then estimated by Maximum Likelihood under the assumption of the probability π_d in the randomisation process (11) being *known*: however, this restrictive assumption may lead to an inconsistent estimator. A further improvement is presented in Tamer (2003), which treats 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 the sample counterpart. The resulting estimator is shown to be consistent and $\sqrt{n}$ normal, and more efficient than the Maximum Likelihood estimator obtained from considering the multiple equilibria outcomes as a single event.

This paper also contributes to the literature of incoherent models by developing a fully parametric Maximum Likelihood estimator for the model in (1) and (2) for the case where both the dependent variables y_{it} are observable and without imposition any restriction upon the randomisation process (11): as it

[7] Note that the coherency issue of the model is still present even if β_1 and/or β_2 are negative, as it arises from the analysis in Appendix A

will be shown in Section 5.3, the likelihood function does not depend upon the randomisation process d_t.

5 Estimation

5.1 Identification

The model in (1) and (2) is a simultaneous equations system, so that it is crucial to determine under what conditions contagion effects can be identified from interdependence: therefore, the issue of identification needs to be addressed first. Because of the non-linearity induced by the endogenous indicator functions, the reduced form in (13) cannot be obtained from (1) and (2) by a simple non-singular linear transformation: therefore, the exclusion restrictions cannot be derived as in the case of linear models[8]. However, notice that for each combination of the threshold parameters c_1 and c_2, the model is linear in the remaining set of parameters: the identification problem can then be solved by employing elementary linear algebra.

Theorem 1 *Consider the model in* (1) *and* (2) *where* $\mathbf{x}_{1t} \cap \mathbf{x}_{2t} = \varnothing$*; then the model is identified if* $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \neq \mathbf{0}$*.*

Proof. Consider (1) and (2) and, without loss of generality, assume that $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \mathbf{0}$. Define

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{\alpha}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}_2' \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0 & \beta_1 \\ \beta_2 & 0 \end{bmatrix}$$

so that (1) and (2) can be written in matrix form as

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{bmatrix} + \mathbf{B} \begin{bmatrix} \mathbf{I}(y_{1t} - c_1) \\ \mathbf{I}(y_{2t} - c_2) \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{17}$$

Pre-multiply (17) by the a 2×2 non-singular matrix $\mathbf{F}$, so to obtain

$$\mathbf{F} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \mathbf{FA} \begin{bmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{bmatrix} + \mathbf{FB} \begin{bmatrix} \mathbf{I}(y_{1t} - c_1) \\ \mathbf{I}(y_{2t} - c_2) \end{bmatrix} + \mathbf{F} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{18}$$

Therefore, in order for (17) to be identified, (17) itself and (18) must lead to different reduced forms. Alternatively, if (17) is identified the only admissible matrix $\mathbf{F}$ is $\mathbf{F} = \mathbf{I}_2$, meaning that $\mathbf{F} = \mathbf{I}_2$ is the only matrix satisfying the restrictions imposed upon the matrices of parameters $\mathbf{A}$ and $\mathbf{B}$. Therefore, identification of (17) requires $\mathbf{F}$ to be unique and equal to $\mathbf{I}_2$. In order for $\mathbf{F}$ to be unique, from (17) and (18) the system

$$[\mathbf{F} \otimes \mathbf{I}_3 - \mathbf{I}_6] \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B} \end{bmatrix} = 0 \tag{19}$$

[8] See Hausman (1983) for a discussion of identification issues in linear simultaneous equation models.

has to have a unique solution with respect to the elements of $\mathbf{F}$. However, (19) is a system with 4 unknowns (the elements of $\mathbf{F}$) and $(k_1 + k_2 + 4)$ equations. Therefore, solving (19) is equivalent to solving

$$[\mathbf{F} \otimes \mathbf{I}_2 - \mathbf{I}_4] \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} = 0$$

with respect to $\mathbf{F}$. In order for $\mathbf{F}$ to be unique, the condition

$$rank\left(\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \right) = 4 \tag{20}$$

must hold, which can be interpreted as the sufficient rank condition for identification of (17). In order for (20) to hold the condition $k_1 + k_2 \geq 2$ needs to be fulfilled. However, if $\boldsymbol{\alpha}_1 = \mathbf{0}$ or $\boldsymbol{\alpha}_2 = \mathbf{0}$ then (20) does not hold, implying that (17) is not identified. ■

From Theorem 1, the presence of equation specific explanatory variables is sufficient to identify the model in (1) and (2). This is in contrast to the standard linear case, where equation specific explanatory variables are necessary to guarantee identification of the model, but where a further sufficient rank condition is required. In addition, whereas in the linear case the exogenous variables represent the optimal set of instruments, this is no longer the case in the model in (1) and (2) and more generally in nonlinear simultaneous equation models; this point is further discussed in Section 5.2.

Having obtained the condition for identification in Theorem 1 we can now turn our attention to the issue of estimation. Single equation OLS estimation would deliver inconsistent estimates due to the endogeneity of the contagion dummies. Therefore, in Section 5.2 the GIVE estimator of the model proposed in Pesaran and Pick (2006) is discussed, while the Conditional Full Information Maximum Likelihood (CFIML) estimator is derived in Section 5.3.

5.2 Single equation GIVE estimation

Under the simplifying assumption of the threshold parameters c_1 and c_2 being known and equal to c_1^0 and c_2^0 respectively, Pesaran and Pick (2006) propose to estimate the model in (1) and (2) by single equation GIVE estimation.

Define the vectors

$$\boldsymbol{\phi}_i = \left(\boldsymbol{\delta}_i', \boldsymbol{\alpha}_i', \beta_i\right)', \quad \mathbf{y}_i = \left(y_{i1}, \ldots, y_{iT}\right)', \quad \mathbf{h}_{it} = \left(\mathbf{z}_t', \mathbf{x}_{it}', \mathbf{I}\left(y_{jt} - c_j^0\right)\right)'$$

and the matrices

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

for $i, j = 1, 2$, $i \neq j$ and $t = 1, \ldots, T$, $\mathbf{w}_{it}$ being the vector of instruments. The GIVE estimator is then given by

$$\hat{\boldsymbol{\phi}}_i = \left(\mathbf{H}_i'\mathbf{P}_{\mathbf{W}_i}\mathbf{H}_i\right)^{-1} \left(\mathbf{H}_i'\mathbf{P}_{\mathbf{W}_i}\mathbf{y}_i\right), \quad i = 1, 2. \tag{21}$$

Given the nonlinear nature of the model, an important issue is the choice of the optimal vector of instruments $\mathbf{w}_{it}^*$. For given values of c_1 and c_2, the structural equations of the model in (1) and (2) 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 $\mathbf{I}\left(y_{it}-c_i^0\right)$ would be[9]

$$w_{it}^* = \mathrm{E}\left[\mathbf{I}\left(y_{it}-c_i^0\right)|\digamma_t\right] = \Pr\left[y_{it}-c_i^0>0\,|\digamma_t\right], \tag{22}$$

expression (22) being the conditional probability of crisis in market i.[10]

The optimal instrument in (22) is not feasible: although the analytical expression is available in closed form once the joint density of the shocks u_{1t} and u_{2t} in (1) and (2) is known, it also depends upon the unknown values of the parameters [11]. However, following the parametric approach proposed in Kelejian (1971), the endogenous regressor $\mathbf{I}\left(y_{it}-c_i^0\right)$ can be approximated by means of a polynomial of order m in the corresponding equation specific predetermined variables $\mathbf{x}_{it}$.[12] The vector of instruments for the model in (1) and (2) is then given by

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

where $\mathbf{x}_{jt}^n$ denotes the column vector made of the $n-th$ powers of each of the elements of the vector $\mathbf{x}_{jt}$. Pesaran and Pick (2006) approximate the endogenous crisis indicators by a polynomial in the predetermined variables of order up to $m=6$.

The instrumental variables approach to estimation of the model in (1) and (2) however faces several problems. First, (21) is a single equation estimator, so that efficiency issues arise. Second, it is likely to suffer from a weak instruments problem, meaning that the instruments are only weakly correlated with the endogenous explanatory variables; if this is the case, the resulting GIVE estimators do not have asymptotically normal distribution and standard statistical inference provides misleading results[13]. Both the efficiency issues and the weak instruments problem are investigated in Monte Carlo experiment carried out in Section 6.

The GIVE estimator in (21) takes the threshold parameters as known; due to the discontinuity induced by the contagion dummies, the threshold parameters can be estimated by grid search, and the remaining set of parameters by

[9] 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.

[10] 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.

[11] The analytical expression of $\Pr\left[y_{it}-c_i^0>0\,|\digamma_t\right]$ is obtained in Appendix B.

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

[13] For a survey of the weak instruments problem see Stock et al. (2002).

instrumental variables. Define

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

where the notation is chosen so to stress that the elements of $\mathbf{h}_{it}(c_j)$ and $\mathbf{H}_i(c_j)$ depend upon the *unknown* threshold c_j. For the $i-th$ equation the parameter c_j can then be estimated by grid search, the estimator of c_j being

$$\hat{c}_j = \arg\min_{c_j} \left[\mathbf{W}_i'\left(\mathbf{y}_i - \mathbf{H}_i(c_j)\hat{\boldsymbol{\phi}}_i(c_j)\right)\right]' (\mathbf{W}_i'\mathbf{W}_i)^{-1} \left[\mathbf{W}_i'\left(\mathbf{y}_i - \mathbf{H}_i(c_j)\hat{\boldsymbol{\phi}}_i(c_j)\right)\right],$$

for $i, j = 1, 2$, $i \neq j$, where $\hat{\boldsymbol{\phi}}_i(c_j)$ is defined as in (21) for each value of c_j in the grid; the resulting estimator for $\boldsymbol{\phi}_i$ will be given by $\hat{\boldsymbol{\phi}}_i(\hat{c}_j)$.

5.3 Maximum Likelihood Estimation

In Section 5.2 it was pointed out that the limited information GIVE estimator proposed in Pesaran and Pick (2006) may raise efficiency issues as well as suffer from weak instruments problems. Under the additional assumption of the joint distribution of the shocks being known, the threshold parameters c_1 and c_2 in the model in (1) and (2) can be estimated by grid search, while the remaining set of parameters can be estimated by Conditional Full Information Maximum Likelihood (CFIML).

As discussed in Amemiya (1985), in the presence of a one-to-one correspondence between the dependent variables and the shocks the likelihood function for a nonlinear simultaneous equations model can be written as the product between the joint density of the shocks and the Jacobian. However, as the model in (1) and (2) is incoherent, this one-to-one correspondence is no longer present. Therefore, the construction of the likelihood function has to take into account the coherency issue of the model. Intuitively this can be done by considering the following two aspects. First, recall from (15) that the sum of the probabilities of the four possible regimes differs from unity: therefore the joint density of y_{1t} and y_{2t} has to include a normalisation term q_t that ensures the density itself integrates to unity. Second, the joint density of y_{1t} and y_{2t} can be obtained from that of u_{1t} and u_{2t} once the appropriate regime is selected: this means the resulting joint density function will be piecewise.

Formally, define the following four vectors of parameters

$$\boldsymbol{\theta}_1 = \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2'\right)', \qquad \boldsymbol{\theta}_2 = \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \beta_1, \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2'\right)',$$

$$\boldsymbol{\theta}_3 = \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2', \beta_2\right)', \qquad \boldsymbol{\theta}_4 = \left(\boldsymbol{\delta}_1', \boldsymbol{\alpha}_1', \beta_1, \boldsymbol{\delta}_2', \boldsymbol{\alpha}_2', \beta_2\right)',$$

where $\boldsymbol{\theta}_i$ describes the joint density function in regime i, for $i = 1, \ldots, 4$. Given the model in (1) and (2), the joint pdf is then given by

$$\begin{aligned} f\left(y_{1t},y_{2t}\left|F_t\right.\right) &= \frac{1}{q_t} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1\left|y_{1t}\leq c_1, y_{2t}\leq c_2; F_t\right.\right)\Pr\left[y_{1t}\leq c_1, y_{2t}\leq c_2\left|F_t\right.\right] \\ &+\frac{1}{q_t} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_2\left|y_{1t}\leq c_1, y_{2t}> c_2; F_t\right.\right)\Pr\left[y_{1t}\leq c_1, y_{2t}> c_2\left|F_t\right.\right] \\ &+\frac{1}{q_t} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_3\left|y_{1t}> c_1, y_{2t}\leq c_2; F_t\right.\right)\Pr\left[y_{1t}> c_1, y_{2t}\leq c_2\left|F_t\right.\right] \\ &+\frac{1}{q_t} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_4\left|y_{1t}> c_1, y_{2t}> c_2; F_t\right.\right)\Pr\left[y_{1t}> c_1, y_{2t}> c_2\left|F_t\right.\right], \end{aligned}$$

which simplifies to

$$\begin{aligned} f\left(y_{1t},y_{2t}\left|F_t\right.\right) &= \frac{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]}{q_t} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1\left|F_t\right.\right) \\ &+\frac{\left[1-\mathbf{I}\left(y_{1t}-c_1\right)\right]\mathbf{I}\left(y_{2t}-c_2\right)}{q_t} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_2\left|F_t\right.\right) \\ &+\frac{\mathbf{I}\left(y_{1t}-c_1\right)\left[1-\mathbf{I}\left(y_{2t}-c_2\right)\right]}{q_t} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_3\left|F_t\right.\right) \\ &+\frac{\mathbf{I}\left(y_{1t}-c_1\right)\mathbf{I}\left(y_{2t}-c_2\right)}{q_t} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_4\left|F_t\right.\right) \end{aligned} \tag{23}$$

once the properties of the truncated distributions are taken into account[14]. The normalising variable q_t, which ensures that the condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left(y_{1t},y_{2t}\left|F_t\right.\right) dy_{1t}dy_{2t} = 1$$

is fulfilled, is given by

$$\begin{aligned} q_t &= \int_{-\infty}^{c_1}\int_{-\infty}^{c_2} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_1\left|F_t\right.\right) dy_{2t}dy_{1t} + \int_{-\infty}^{c_1}\int_{c_2}^{\infty} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_2\left|F_t\right.\right) dy_{2t}dy_{1t} \\ &+\int_{c_1}^{\infty}\int_{-\infty}^{c_2} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_3\left|F_t\right.\right) dy_{2t}dy_{1t} + \int_{c_1}^{\infty}\int_{c_2}^{\infty} f\left(y_{1t},y_{2t};\boldsymbol{\theta}_4\left|F_t\right.\right) dy_{2t}dy_{1t} \\ &= \Pr\left[y_{1t}\leq c_1, y_{2t}\leq c_2\left|F_t\right.\right] + \Pr\left[y_{1t}\leq c_1, y_{2t}> c_2\left|F_t\right.\right] \\ &+\Pr\left[y_{1t}> c_1, y_{2t}\leq c_2\left|F_t\right.\right] + \Pr\left[y_{1t}> c_1, y_{2t}> c_2\left|F_t\right.\right] \\ &= 1+\Pr\left[E\left|F_t\right.\right], \end{aligned} \tag{24}$$

expression (24) being the same as (15). The log-likelihood function is then given by

$$L_n = \sum_{i=1}^{T} \log f\left(y_{1t},y_{2t}\left|F_t\right.\right).$$

Note that the CFIML estimator cannot be replaced by a SURE-type estimator: this is because the density selected by the indicator functions $\mathbf{I}\left(y_{1t}-c_1\right)$ and $\mathbf{I}\left(y_{2t}-c_2\right)$ has to be weighted by the normalising term q_t, which depends upon the whole set of parameters of the model, and not just on that characterising the selected regime.

[14] From the poperties of the truncated distributions we have

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

The other three components of the density function are obtained in a similar way.

6 Monte Carlo Analysis

6.1 Experimental design

The Data Generating Processes (DGPs) for the Monte Carlo experiments are based on the following model

$$y_{it}^r = \delta_i + \alpha_i x_{it}^r + \beta_i \mathbf{I}\left(y_{jt}^r - c_j\right) + u_{it}^r \quad i, j = 1, 2 \quad i \neq j \tag{25}$$

where $t = 1, 2, \ldots T$, $r = 1, 2, \ldots R$, 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 through the replications. The estimated model is also given by (25). Without loss of generality, we assume that the threshold parameters c_i are *known* and equal to c_i^0. In the case of the CFIML estimator, the resulting (normalised) log-likelihood function consistent with (25) 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 on the parameter β_1. The performance of the estimators we consider in Section 5 is assessed by computing the bias and RMSE, respectively defined as

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

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 the rejection frequency is the size of the test; conversely, if the test statistic is computed under the alternative the rejection frequency is the power. In computing the test statistic for the CFIML 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 actual size being equal to the nominal size 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 $[0.04, 0.06]$.

In generating the data, the sample size was set equal to $T = 50, 100, 200, 500, 1000$, the values $c_1^0 = c_2^0 = 1.64$ were arbitrarily chosen and the seed was set equal to -1. Finally, we set $\beta_2 = 0.2$ and we considered two different classes of experiments, namely $\beta_1 = 0$ (so that no multiple equilibria arise) and $\beta_1 = 0.5$ (so that multiple equilibria do arise)[15].

[15] 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.

Results of the Monte Carlo experiments for the CFIML estimator and the GIVE estimator with $m = 1$ and $m = 6$ are reported in Table 1 (for the case $\beta_1 = 0$) and in Table 2 (for the case $\beta_1 = 0.5$)[16].

Experiment 1: $\beta_1 = 0$**.** In this case the DGP in (25) simplifies to

$$\begin{aligned} y_{1t}^r &= \delta_1 + \alpha_1 x_{1t}^r + u_{1t}^r, \\ y_{2t}^r &= \delta_2 + \alpha_2 x_{2t}^r + \beta_2 \mathbf{I}\left(y_{1t}^r - c_1^0\right) + u_{2t}^r. \end{aligned}$$

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

$$\mathbf{\Sigma}_u = \begin{bmatrix} 1 & \rho_{u_1 u_2} \\ \rho_{u_1 u_2} & 1 \end{bmatrix}$$

where

$$\rho_{u_1 u_2} = \text{Cov}\left[u_{1t}^r u_{2t}^r\right] = \frac{\gamma_1 \gamma_2}{\sqrt{\gamma_1^2 + 1}\sqrt{\gamma_2^2 + 1}},$$

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

$$q_{u_1 u_2} = \text{E}\left[\rho_{u_1 u_2}\right] = \left(\text{E}\left[\frac{\gamma_i}{\sqrt{\gamma_i^2 + 1}}\right]\right)^2 = 0.1616$$

The equation specific explanatory variable x_{it}^r is generated as

$$\begin{pmatrix} x_{1t}^r \\ x_{2t}^r \end{pmatrix} \sim N\left[\mathbf{0}; \mathbf{\Sigma}_x\right].$$

The covariance matrix $\mathbf{\Sigma}_x$ 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 $q_{it}^r \sim NID(0,1)$, $h_t^r \sim NID(0,1)$ while the coefficient ϕ_i is fixed in repeated samples and $\phi_i \sim U(0.8, 1)$. In this way we have

$$\mathbf{\Sigma}_x = \begin{bmatrix} 1 & \rho_{x1x2} \\ \rho_{x1x2} & 1 \end{bmatrix},$$

[16] 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.

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 regressor is endogenous, the coefficient of determination can be easily obtained as

$$R_1^2 = 1 - \frac{\mathrm{Var}\left[u_{1t}^r\right]}{\mathrm{Var}\left[y_{1t}^r\right]}$$

where

$$\mathrm{Var}\left[y_{1t}^r\right] = \alpha_1^2 \mathrm{Var}\left[x_{1t}^r\right] + \mathrm{Var}\left[u_{1t}^r\right] = \alpha_1^2 + 1$$

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 $\mathbf{I}\left(y_{1t}^r - c_1^0\right)$ 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

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

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

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

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 $R_2^2 \simeq 0.5$.[17]

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 when very few observations for the crisis regime are available. This is an important issue in multiple regimes models. For example, in the context of univariate TAR models, Hansen (1997) suggests to select the grid range for the threshold

[17] The coefficients of determination R_1^2 and R_2^2 have very similar results because the endogenous dummy $\mathbf{I}\left(y_{1t}^r - c_1^0\right)$ determines a shift in the intercept in the equation for y_{2t}^r. Therefore, the effect upon the explanatory power of the model is negligible, unless β_2 assumes very high values.

parameter so to trim the top and bottom 15% quantiles of the distribution of the dependent variable: in this way each regime has enough observations to estimate the relevant parameter. In the case of y_{1t}^r the unconditional probability of crisis π_1 is obtained in closed form as

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

so that δ_1 is given by

$$\delta_1 = c_1^0 - \sqrt{\alpha_1^2 + 1}\left[\mathbf{\Phi}^{-1}\left(1 - \pi_1\right)\right].$$

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

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

We then choose 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, since if $\mathbf{I}\left(y_{2t}^r - c_2^0\right) = 0, \forall t$ then no observations are available to estimate β_1; conversely, if $\mathbf{I}\left(y_{2t}^r - c_2^0\right) = 1, \forall t$ then β_1 cannot be identified from the intercept δ_1: therefore, each replication is repeated until $\mathbf{I}\left(y_{2t}^r - c_2^0\right)$ is different from being a vector of all zeros or ones.

Experiment 2: $\beta_1 = 0.5$. In this case the DGP is given by the reduced form obtained in (13), where the value $\pi_d = 0.5$ 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 $\mathbf{I}\left(y_{it}^r - c_i^0\right)$, 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

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

the general expression for $\Pr\left[y_{it} - c_i > 0 \left| \mathcal{F}_t^r\right.\right]$ 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 respectively correspond to $R_i^2 \simeq 0.5$ and $R_i^2 \simeq 1$, $i = 1, 2$.

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

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

and δ_i is chosen so to control for π_i. We set $\pi_1 \simeq \pi_2 \simeq \pi = 0.005, 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50$.

6.2 Bias and RMSE

The bias of the CFIML 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 $m = 1$) as well as with α, this last feature being driven by a strengthening in the instruments. 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 performance of the 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 $\pi = 0.01$). Finally, the bias of the CFIML estimator is lower than that of the GIVE estimators for virtually any combination of T and π.

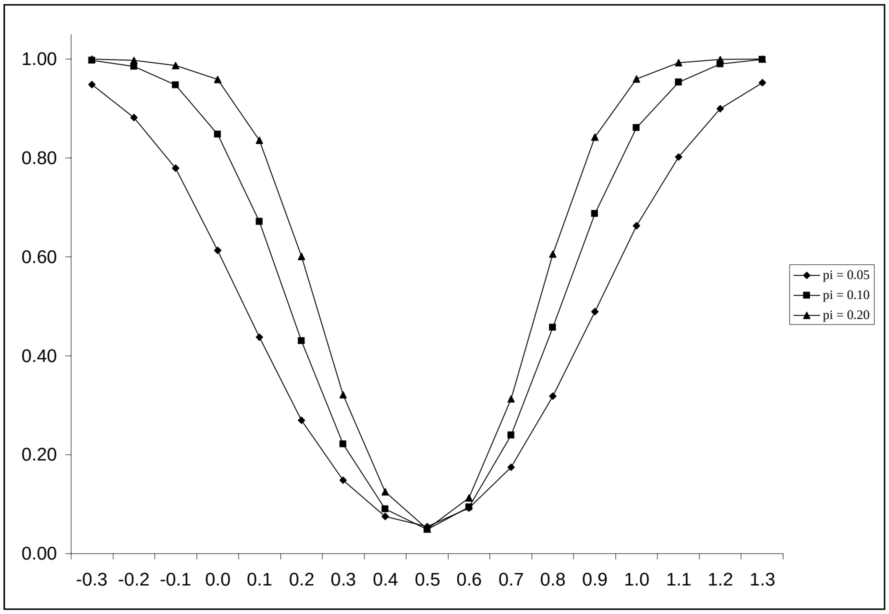
The RMSE of the CFIML estimator decreases with the sample size T and with the probability of crisis π, a pattern similar to that of the bias; it also diminishes with α, while it does not show any clear pattern related to the magnitude of β_1. As far as the GIVE estimators are concerned, their bias decreases 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 CFIML 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 CFIML 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 5.2.

6.3 Size and power

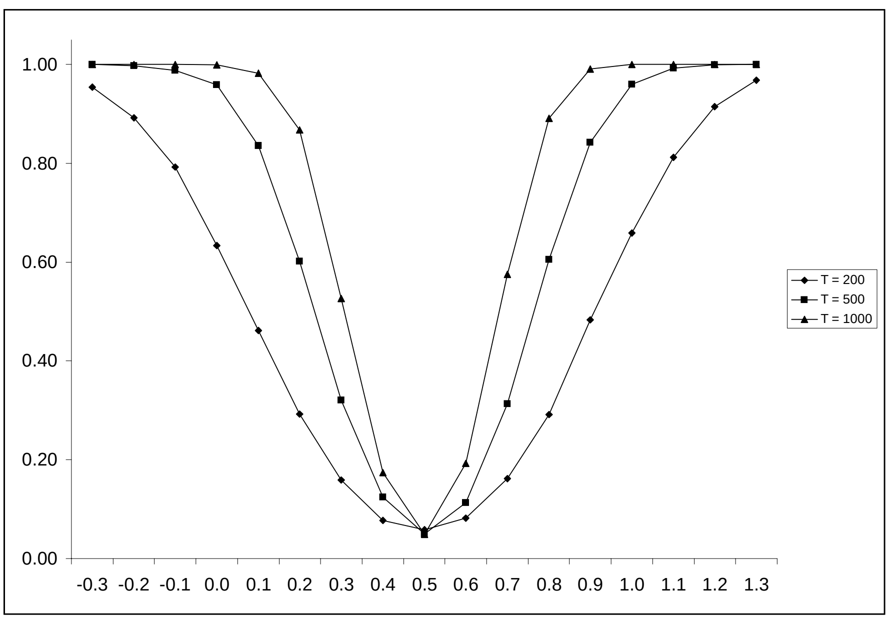
Starting from the CFIML estimator, for $T = 50$ the actual size never approaches the nominal size, regardless of the value of π. As the sample size T increases the actual size tends to approach the nominal size, this feature generally being true for any value of the probability of crisis π. The only 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 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, this last feature being in line with Theorem 1 in Section 5. Considering the GIVE estimators, for $m = 1$ the test is generally undersized when $\alpha = 0.5$, the nominal size being sistematically 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 CFIML and the GIVE estimators shows that the former achieves a better size performance.

The power of the test is computed by testing the null $\beta_1 = 0.5$ in Experiment 1 and $\beta_1 = 1$ in Experiment 2. In the case of the CFIML estimator, the power increases with the sample size T as well as with the probability of crisis π, as also shown in Figures 2 and 3 below[18]; it also increases with α. Tests based upon the GIVE estimators show a similar behaviour. Further, for $\alpha = 0.5$ the GIVE estimator with $m = 1$ seems to have slightly better power than that with $m = 6$, while for $\alpha = 1$ the estimator with $m = 6$ seems to provide a better performance. Finally, the CFIML estimator is clearly more powerful than the GIVE estimators.

[18] Comparisons based on the power of the tests are made for combinations of T and π such that the actual size is equal to the nominal size.



Power Function for CFIML Estimator, Experiment 2, $T = 500$, $\alpha = 0.5$.



Power Function for CFIML Estimator, Experiment 2, $\pi = 0.2$, $\alpha = 0.5$.

7 Testing for contagion

Given the model in (1) and (2), testing for contagion from market j to market i is equivalent to testing the null hypothesis $\beta_i = 0$ against the one-sided alternative $\beta_i > 0$, for $i, j = 1, 2$ and $i \neq j$: this is equivalent to testing for linearity in the equation for y_i. The implementation of tests for linearity is an important issue in the econometrics literature; for the case of SETAR models (which are a particular case of the model in (1) and (2)) a comprehensive survey can be found in Hansen (1999). The main problem arises because under the null hypothesis $\beta_i = 0$ the threshold parameter c_j is not identified. This is an example of the Davies problem, named after the work by Davies (1977, 1987); the problem describes the situation where a nuisance parameter is not identified under the null hypothesis, so that standard testing procedures cannot be applied.

It is therefore important to understand what the Davies problem actually involves and what kind of solutions have been proposed so far in the econometric literature. As a starting point consider the univariate model

$$y_{2t} = \boldsymbol{\alpha}_2' \mathbf{x}_{2t} + \boldsymbol{\beta}_2' \mathbf{h}\left(\mathbf{x}_t, \mathbf{c}_1\right) + u_{2t}, \tag{26}$$

where $\mathbf{x}_{2t}$ is a $k_2 \times 1$ subvector of the $k \times 1$ vector of predetermined variables $\mathbf{x}_t$, $\boldsymbol{\beta}_2$ and $\mathbf{c}_1$ are $p \times 1$ a $q \times 1$ vectors of parameters respectively, and $\mathbf{h}(\cdot, \cdot)$ is a $m \times 1$ vector of functions. Further, suppose the interest lies in testing the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$. If $\mathbf{c}_1$ was known the model in (26) could be estimated by OLS, and the relevant test statistic for the null hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}$ would be distributed as χ^2_p. However, since $\mathbf{c}_1$ is generally unknown, then it can be estimated by grid search while estimates of $\boldsymbol{\alpha}_2$ and $\boldsymbol{\beta}_2$ are obtained by OLS; further, under H_0 the vector of parameters parameter $\mathbf{c}_1$ is not identified and the null hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}$ cannot be tested in the usual manner. In order to solve the problem, Davies (1977) proposes to apply Roy's type I principle and take the supremum of the LM, Wald or LR statistic over the admissible values of $\mathbf{c}_1$. The resulting test statistic can therefore be seen as an empirical process, meaning that the underlying stochastic process depends both on time and on the nuisance parameters. However, the distribution of the test statistic will in general be non-standard. Indeed, for given values of $\mathbf{c}_1$ the distribution of the test statistic under the null hypothesis would still be χ^2_p; however, the *joint* distribution of the test statistic cannot be obtained from that of the individual test statistics, as the latter are correlated across different values of the $\mathbf{c}_1$: as we will see, the joint distribution can only be tabulated for special cases.

As discussed in Pesaran (1981), a special case arises when $\mathbf{h}\left(\mathbf{x}_t, \mathbf{c}_1\right) = \mathbf{c}_1' \mathbf{x}_t$, so that (26) becomes

$$y_{2t} = \boldsymbol{\alpha}_2' \mathbf{x}_{2t} + \beta_2 \left(\mathbf{c}_1' \mathbf{x}_t\right) + u_{2t},$$

β_2 being a scalar[19]. In this case, application of Roy's type I principle for testing $\beta_2 = 0$ provides a test statistic distributed as χ^2_q, which is equivalent to that

[19] A formal proof of the intuition presented in Pesaran (1981) is derived in McAleer and Pesaran (1986).

used for testing $(\beta_2 \mathbf{c}_1) = 0$. This means that the lack of knowledge of the true value of $\mathbf{c}_1$ leads to a loss in power that can be quantified[20].

Hansen (1996*b*) extends the work in Pesaran (1981) by considering the general set up given by (26) without imposing any restriction upon the parametric form of $\mathbf{h}(\mathbf{x}_t, \mathbf{c}_1)$. Since for given values of $\mathbf{c}_1$ the model can be estimated by OLS, the vector of scores is obtained in closed form, and the OLS estimator of the relevant parameters (including $\boldsymbol{\beta}_2$) is itself available in closed form as a function of the scores. Hansen (1996*b*) proposes to test the null hypothesis $\boldsymbol{\beta}_2 = 0$ by constructing a Wald statistic: for given values of $\mathbf{c}_1$, under the null hypothesis the test statistic is distributed as χ^2_p; however, the critical values of the joint distribution arising from taking the supremum of the Wald statistic over $\mathbf{c}_1$ can be obtained by simulation since the OLS estimator has closed form in terms of the scores[21]. Notice that in this general case the loss in power induced by the Davies problem cannot be quantified, due to the nonlinear nature of the model in (26).

The model in (1) and (2) can be seen as an extension of (26) to a system of simultaneous equations; therefore, the testing procedure developed in Hansen (1996*b*) cannot be directly applied. However an extension to it can be easily obtained if the model is estimated by instrumental variables, as discussed in Section 5.2. However, as shown in the Monte Carlo experiment conducted in Section 6, the GIVE estimator is likely to suffer from a weak instruments problem as well as from efficiency issues: therefore, statistical inference based upon it is likely to provide misleading results.

In order to test for contagion in the model in (1) and (2) we therefore need a testing procedure based upon the CFIML estimator developed in Section 5.3. However, since in this case the scores are not available in closed form, a direct extension of the Wald statistic proposed in Hansen (1996*b*) is not available. We therefore deploy the general testing procedure proposed in Hansen (1992, 1996*a*) to deal with the presence of the Davies problem. The idea is to work directly with the likelihood ratio function, so that no assumption about the scores is needed. More precisely, the likelihood ratio can be seen as a function of the nuisance parameter, the likelihood ratio statistic being the supremum over the admissible values of the nuisance parameters; by using empirical process theory, an upper bound to the asymptotic distribution of a suitably standardised likelihood ratio statistic can be found. However, since the asymptotic distribution is a bound, the resulting test may be conservative (or undersized), meaning that a loss in power may be suffered. The procedure used to obtain the critical values is discussed in Appendix C.

[20] This is because of the inverse relationship between power and number of degrees of freedom in a noncentral χ^2 test with a given noncentrality parameter.

[21] Following Andrews and Ploberger (1994), Hansen (1996*b*) also considers tests statistic resulting from taking the average and exponential average of the χ^2 processes, since they are supposed to deliver superior local power under the null hypothesis.

8 Empirical application

8.1 Data and model specification

In this section we apply the model given in (1) and (2) to analyse the interaction between the New York Stock Exchange (NYSE) and the other major European stock markets, namely London, Frankfurt, Zurich and Paris. In order to carry out the analysis we make use of daily stock market prices for the S&P 500 (New York), FTSE 100 (London), 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 above 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 FTSE 100 is the principal UK index; it consists of the UK's biggest companies by 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 London trades from 8:00 to 16:30 London time; New York 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[22].

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

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

After removing holidays in each countries, we were left with 3741 observations of common trading days for the five series[23]. Descriptive statistics for the resulting stock market returns and the correlations between them are provided in Table 3. Average daily returns are all positive, with New York providing the highest rate followed by Zurich, Frankfurt, Paris and London. The S&P 500 is also the least volatile index, as evidenced by the value of its sample standard deviation, followed by the FTSE 100, 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 FTSE 100

[22] The use of pseudo-closing prices to avoid the problem of non-synchronous data was first suggested in Martens and Poon (2001). Also, the Japanese stock market had to be excluded since it does not have any common trading time with any of the other stock stock markets we consider.

[23] For each market we defined as holidays any day where the return is exactly equal to zero.

and the CAC 40 are 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 market: this may be due to the significant time difference in *local time* between New York and the other markets.

In carrying out the empirical analysis, we first note that a crisis is associated with an extreme *negative* value of r_{it}, meaning that 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, the returns r_{it} have to be devolatised. The variable y_{it} is then defined as

$$y_{it} = -\frac{r_{it}}{\sigma_{i|t-1}}, \quad i = 1, 2,$$

where

$$\sigma^2_{i|t-1} = \mathrm{Var}\left[r_{it}\left|\boldsymbol{\Omega}_{i,t-1}\right.\right],$$

$\boldsymbol{\Omega}_{i,t-1}$ being 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): 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 evidenced in Table 3. The GARCH$(1,1)-t$ model, specified in terms of the returns r_{it}, is given by

$$\begin{aligned}
&r_{it} = \mu_i + \sum_{k=1}^{5} \gamma_{ik} r_{i,t-k} + \varepsilon_{it}, \\
&\varepsilon_{it} = z_{it}\sigma_{i|t-1}, \\
&z_{it}\left|\boldsymbol{\Omega}_{i,t-1}\right. \sim iidt_v\left(0,1\right), \\
&\sigma^2_{i|t-1} = \omega + \alpha\varepsilon^2_{i,t-1} + \beta\sigma^2_{i|t-2},
\end{aligned} \tag{27}$$

where v denotes the number of degrees of freedom of the t distribution[24]. The market returns r_{it} are modelled as an autoregressive process of order five so to control for serial correlation as well as weekly effects. The model specification then becomes

$$\hat{y}_{1t} = \delta_1 + \boldsymbol{\alpha}_1{}'\mathbf{x}_{1t} + \beta_1\mathbf{I}\left(-r_{2t} - c_2\right) + u_{1t}, \tag{28}$$

$$\hat{y}_{2t} = \delta_2 + \boldsymbol{\alpha}_2{}'\mathbf{x}_{2t} + \beta_2\mathbf{I}\left(-r_{1t} - c_1\right) + u_{2t}, \tag{29}$$

where

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

[24] The GARCH$(1,1)-t$ model in (27) is estimated by means of the procedure developed by Laurent and Peters (2005).

$\hat{\sigma}_{i|t-1}$ being the estimator of $\sigma_{i|t-1}$ arising from (27), and $i = 2$ always referring to the S&P 500. In this way we are left with 3736 observations. The threshold parameters c_i are estimated by grid search, while the remaining set of parameters is estimated by the CFIML procedures discussed in Section 5.3. For each stock market return the width of the grid 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 are $c_{S\&P500} \in [0.65, 3.40]$, $c_{FTSE100} \in [0.75, 3.40]$, $c_{SMI} \in [0.80, 3.60]$, $c_{DAX30} \in [0.90, 4.70]$ and $c_{CAC40} \in [0.90, 4.10]$, with a step equal to 0.01.

8.2 Results

Results from the estimation of (28) and (29) are reported in Table 4. Starting from the effect of the NYSE upon European markets, we can see that the S&P 500 affects all the other indices as soon as it return goes below -2.69%; the only expection is the DAX 30, which reacts whenever the S&P 500 goes falls below -2.44%. However, the effect of a crisis varies from market to market: it causes a fall of 0.42% in the FTSE 100, 0.58% in the DAX 30, 0.57% in the SMI and 0.64% in the CAC 40. This result can be interpreted as evidence in favour of the FTSE 100 being significantly the least vulnerable amongst the European indices to shocks generated by the S&P 500.

Turning the attention to the effect of the European markets on the S&P 500, it can be seen that the FTSE 100 seems to be the index that most often affects the S&P 500 (14% of the times), followed by the CAC 40 (4%), the DAX 30 (2%) and SMI (0.7%). 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 (that is the magnitude of the corresponding contagion coefficient) is.

However, this set of results ought to be interpreted with cautiouness, as identification issues may arise. Identification of the model specification is assessed by means of the Wald statistics W, which tests the joint significance of the equation specific explanatory variables. We can therefore see that while the models for the FTSE 100 and CAC 40 are identified (although the null of the coefficients of lagged values of the S&P 500 being simultaneously equal to zero is only rejected at 5% level), in the other two cases identification issues do arise.

Table 5 reports the results of the application of the test for contagion discussed in Section 7. The null hypothesis of no contagion from each of the European markets to the NYSE is considered; this is equivalent to testing the null hypothesis $\beta_2 = 0$ in the model specification given in (28) and (29). As it can be seen, in each case the null hypothesis cannot be rejected, meaning that the NYSE seems to be immune to crises occurring in any of the major European stock markets. This result however has to be interpreted bearing in mind the identification issues discussed above.

9 Conclusion

This paper deals with estimation and inference in the canonical model of contagion developed in Pesaran and Pick (2006), which represents a two-equation nonlinear simultaneous equations model. As far as estimation is concerned, the likelihood function is obtained, which takes into account the coherency issues of the model. Using Monte Carlo simulation, the resulting Full Information Maximum Likelihood estimator is compared to the single equation GIVE estimators proposed in Pesaran and Pick (2006): the former provides better performance than the latter, which also suffer from a weak instruments problem and face efficiency issues. Statistical inference aimed at assessing the presence of contagion turns out to be nonstandard due to the presence of an unidentified nuisance parameter under the null hypothesis.

The canonical model is then applied to stock market returns. From the empirical results the New York Stock Exchange seems to be unaffected by a crisis taking place in any of the major European stock markets. However, these results have to be interpreted with cautiouness due to the limited number of crisis periods and to the presence of identification issues.

References

[1] Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell, Oxford.

[2] Andrews, D. W. K. and W. Ploberger (1994), "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative", *Econometrica*, 62(6), 1383-1414.

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

[4] 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.

[5] 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.

[6] 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.

[7] Corsetti, G., M. Pericoli and M. Sbracia (2006), "Some Contagion, Some Interdependence': More Pitfalls in Tests of Financial Contagion" *Journal of International Money and Finance*, 24(8), 1177-1199.

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

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

[10] Diamond, D. W. and P. H. Dybvig (1983), "Bank Runs, Deposit Insurance, and Liquidity", *The Journal of Political Economy*, 91(3), 401-419.

[11] Dungey, M., R. Fry, B. Gonzalez-Hermosillo and Vance L. Martin (2005), "Empirical Modelling of Contagion: A Review of Methodologies", *Quantitative Finance*, 5(1), 9-24.

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

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

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

[15] 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.

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

[17] 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.

[18] 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.

[19] Hansen, B. E. (1992), "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics*, 7, S61-S82.

[20] Hansen, B. E. (1996*a*), "Erratum: The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics*, 11(2), 195-198.

[21] Hansen, B. E. (1996*b*), "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis", *Econometrica*, 64(2), 413-430.

[22] Hansen, B. E. (1997), "Inference in TAR Models", *Studies in Nonlinear Dynamics and Econometrics*, 2(1), 1-14.

[23] Hansen, B. E. (1999), "Testing for Linearity", *Journal of Economic Surveys*, 13(5), 551-576.

[24] Hausman, J. A. (1983), "Specification and Estimation of Simultaneous Equation Models", Ch. 7 in *Handbook of Econometrics*, Vol. I, eds. Z. Griliches and M. D. Intriligator, North-Holland, Amsterdam.

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

[26] 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.

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

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

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

[30] Krugman, P. (1979), "A Model of Balance-of-Payments Crises", *Journal of Money, Credit and Banking*, 11(3), 311-325.

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

[32] 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.

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

[34] 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.

[35] Masson, P. (1998), "Contagion: Monsoonal Effects, Spillovers, and Jumps Between Multiple Equilibria", IMF Working Paper 98/142.

[36] McAleer M. and M. H. Pesaran (1986), "Statistical Inference in Non-Nested Econometric Models", *Applied Mathematics and Computation*, 20(3-4), 271-311.

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

[38] Obstfeld, M. (1986), "Rational and Self-Fulfilling Balance-of-Payments Crises", *The American Economic Review*, 76(1), 72-81.

[39] Pesaran, M. H. (1981), "Pitfalls of Testing Non-Nested Hypothesis by the Lagrange Multiplier Method", *Journal of Econometrics*, 17(3), 323-331.

[40] Pesaran, M. H. and A. Pick (2006), "Econometric Issues in the Analysis of Contagion", *Journal of Economic Dynamics and Control*, forthcoming.

[41] 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.

[42] 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.

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

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

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

A Further analysis of the model

Consider first the case $\beta_1, \beta_2 < 0$. Taking into account (3) and (4), the model in (1) and (2) can be written as

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

The five mutually exclusive solution regions are then given by

$$\text{Region A: } \begin{cases} W_{1t} \geq 0 \\ W_{2t} \geq 0 \end{cases}$$

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

$$\text{Region C: } \begin{cases} W_{1t} < -1 \\ W_{2t} < -1 \end{cases}$$

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

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

and the reduced form by

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

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

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

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

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

the process d_t being defined in (11). The normalising term for the likelihood function is equal to

$$q_t = 1 + \Pr\left[E \,|\mathcal{F}_t\right]$$

the analytical expression for $\Pr\left[E \,|\mathcal{F}_t\right]$ is the same as that obtained in (16). Therefore, the expression for the joint density function given in (23) remains valid also for the case $\beta_1, \beta_2 < 0$.

Consider now the case $\beta_1 < 0, \beta_2 > 0$[25]. The model can be equivalently written as

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

In this case only four mutually exclusive solution regions arise

$$\text{Region A: } \begin{cases} W_{1t} \geq -1 \\ W_{2t} > 0 \end{cases}$$

$$\text{Region B: } \begin{cases} W_{1t} \geq 0 \\ W_{2t} \leq 0 \end{cases}$$

$$\text{Region C: } \begin{cases} W_{1t} < 0 \\ W_{2t} \leq -1 \end{cases}$$

$$\text{Region D: } \begin{cases} W_{1t} < -1 \\ W_{2t} > -1 \end{cases}$$

while in region E defined as

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

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

no solution arises. The reduced form is then given by

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

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

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

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

and no multiple equilibria arises. In this case the normalising term q_t is given by

$$q_t = 1 - \Pr[E \,|\mathcal{F}_t]$$

where the analytical expression for $\Pr[E\,|\mathcal{F}_t]$ is of the same magnitude but opposite sign compared to the expression provided in (16). Therefore, also for the case $\beta_1 < 0, \beta_2 > 0$ the expression for the joint density function given in (23) remains valid.

B Further mathematical results

From the reduced form equation (13) it follows that

$$\begin{aligned}\Pr[y_{1t} - c_1 > 0\,|\mathcal{F}_t] \;=\;& \Pr[A\,|\mathcal{F}_t]\Pr[y_{1t} - c_1 > 0\,|A;\mathcal{F}_t] \\ &+ \Pr[B\,|\mathcal{F}_t]\Pr[y_{1t} - c_1 > 0\,|B;\mathcal{F}_t] \\ &+ \Pr[C\,|\mathcal{F}_t]\Pr[y_{1t} - c_1 > 0\,|C;\mathcal{F}_t] \\ &+ \Pr[D\,|\mathcal{F}_t]\Pr[y_{1t} - c_1 > 0\,|D;\mathcal{F}_t] \\ &+ \Pr[E\,|\mathcal{F}_t]\Pr[y_{1t} - c_1 > 0\,|E;\mathcal{F}_t]\end{aligned}$$

where

$$\begin{aligned}\Pr[y_{1t} - c_1 > 0\,|A;\mathcal{F}_t] &= \Pr[y_{1t} - c_1 > 0\,|B;\mathcal{F}_t] = 1 \\ \Pr[y_{1t} - c_1 > 0\,|C;\mathcal{F}_t] &= \Pr[y_{1t} - c_1 > 0\,|D;\mathcal{F}_t] = 0 \\ \Pr[y_{1t} - c_1 > 0\,|E;\mathcal{F}_t] &= (1 - \pi_d)\end{aligned}$$

so that

$$\Pr[y_{1t} - c_1 > 0\,|\mathcal{F}_t] = \Pr[A\,|\mathcal{F}_t] + \Pr[B\,|\mathcal{F}_t] + (1 - \pi_d)\Pr[E\,|\mathcal{F}_t]$$

and taking into account (9)

$$\begin{aligned}\Pr[y_{1t} - c_1 > 0\,|\mathcal{F}_t] \;=\;& \Pr[W_{1t} > 0\,|\mathcal{F}_t] + \Pr[-1 < W_{1t} \leq 0, W_{2t} > 0] \\ &+ (1 - \pi_d)\Pr[-1 < W_{1t} \leq 0, -1 < W_{2t} \leq 0]\,.\end{aligned}$$

In the same way, we obtain

$$\begin{aligned}\Pr\left[y_{2t}-c_2>0\,|\mathcal{F}_t\right] &= \Pr\left[W_{2t}>0\,|\mathcal{F}_t\right]+\Pr\left[W_{1t}>0,-1<W_{2t}\leq 0\right]\\ &\quad+\left(1-\pi_d\right)\Pr\left[-1<W_{1t}\leq 0,-1<W_{2t}\leq 0\right]\end{aligned}$$

so that in general

$$\begin{aligned}\Pr\left[y_{it}-c_i>0\,|\mathcal{F}_t\right] &= \Pr\left[W_{it}>0\,|\mathcal{F}_t\right]+\Pr\left[-1<W_{it}\leq 0,W_{jt}>0\right]\\ &\quad+\left(1-\pi_d\right)\Pr\left[-1<W_{it}\leq 0,-1<W_{jt}\leq 0\right]\end{aligned}$$

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

C Test for contagion

Let '$\Rightarrow$' denote weak convergence with respect to the uniform metric and '$\|\cdot\|$' the Euclidean metric. In order to test for contagion, we apply the procedure suggested in Hansen (1992, 1996*a*). Consider the model

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

and suppose we are interested in testing

$$H_0:\beta_2=0,\qquad H_1:\beta_2>0,$$

which is equivalent to testing for linearity in the equation for y_{2t}. Define the vectors of parameters

$$\boldsymbol{\theta}=\left(\boldsymbol{\delta}_1',\boldsymbol{\alpha}_1',\beta_1,\sigma_1^2,\boldsymbol{\delta}_2',\boldsymbol{\alpha}_2',c_2,\sigma_2^2,\sigma_{12}\right)',\quad \boldsymbol{\varphi}=\left(\beta_2,\boldsymbol{\theta}'\right)';$$

the log-likelihood of the model can be written as

$$L_T\left(c_1,\boldsymbol{\varphi}\right)=\sum_{t=1}^{T}l_t\left(c_1,\boldsymbol{\varphi}\right)=\sum_{t=1}^{T}\log f\left(y_{1t},y_{2t}\,|\mathcal{F}_t\right),$$

where $f\left(y_{1t},y_{2t}\,|\mathcal{F}_t\right)$ is the joint density function of y_{1t} and y_{2t}, $\mathcal{F}_t$ is defined as

$$\mathcal{F}_t=\left(\mathbf{z}_t^{'},\mathbf{x}_{1t}^{'},\mathbf{x}_{2t}^{'}\right)^{'},$$

and the subscript T denotes the sample size. The concentrated log-likelihood function is given by

$$\widehat{L}_T^U\left(c_1\right)=L_T\left(c_1,\hat{\boldsymbol{\varphi}}^U\left(c_1\right)\right)=\sum_{t=1}^{T}l_t\left(c_1,\hat{\boldsymbol{\varphi}}^U\left(c_1\right)\right),$$

where

$$\hat{\boldsymbol{\varphi}}^U\left(c_1\right)=\left(\hat{\beta}_2^U\left(c_1\right),\hat{\boldsymbol{\theta}}^U\left(c_1\right)'\right)'=\arg\max_{\boldsymbol{\varphi}}\left[L_T\left(c_1,\boldsymbol{\varphi}\right)\right].$$

The concentrated log-likelihood function has large-sample counterpart given by

$$L_T^U(c_1) = L_T\left(c_1, \boldsymbol{\varphi}^U(c_1)\right),$$

where

$$\begin{aligned}
\boldsymbol{\varphi}^U(c_1) &= \left(\beta_2^U(c_1), \boldsymbol{\theta}^U(c_1)'\right)' \\
&= \lim_{T\to\infty}\left\{\arg\max_{\boldsymbol{\varphi}} \frac{1}{T}\mathrm{E}\left[L_T(c_1, \boldsymbol{\varphi})\right]\right\} \\
&= \lim_{T\to\infty}\left\{\arg\max_{\boldsymbol{\varphi}} \frac{1}{T}\mathrm{E}\left[\sum_{t=1}^{T} l_t(c_1, \boldsymbol{\varphi})\right]\right\} \\
&= \arg\max_{\boldsymbol{\varphi}} \mathrm{E}\left[l_t(c_1, \boldsymbol{\varphi})\right]
\end{aligned}$$

is the pseudo-true value of $\boldsymbol{\varphi}$ for fixed c_1. Note that $\hat{\boldsymbol{\varphi}}^U(c_1)$ is a consistent estimator for $\boldsymbol{\varphi}^U(c_1)$ with rate of convergence $\sqrt{T}$ and uniformly in c_1, that is

$$\sqrt{T}\sup_{c_1}\left\|\hat{\boldsymbol{\varphi}}^U(c_1) - \boldsymbol{\varphi}^U(c_1)\right\| = O_p(1).$$

The concentrated likelihood ratio (LR) as a function of c_1 can be written as

$$\widehat{LR}_T(c_1) = \widehat{L}_T^U(c_1) - \widetilde{L}_T^R(c_1), \tag{30}$$

where $\widetilde{L}_T^R(c_1)$ denotes the restricted value of the log-likelihood function under the null hypothesis $\beta_2 = 0$; $\widetilde{L}_T^R(c_1)$ is defined as

$$\widetilde{L}_T^R(c_1) = L_T\left(c_1, \tilde{\boldsymbol{\varphi}}^R(c_1)\right) = \sum_{t=1}^{T} l_t\left(c_1, \tilde{\boldsymbol{\varphi}}^R(c_1)\right),$$

where

$$\tilde{\boldsymbol{\varphi}}^R(c_1) = \left(0, \tilde{\boldsymbol{\theta}}^R(c_1)'\right)'$$

$\tilde{\boldsymbol{\theta}}^R(c_1)$ being determined as

$$\tilde{\boldsymbol{\theta}}^R(c_1) = \arg\max_{\boldsymbol{\theta}}\left[L_T(c_1, 0, \boldsymbol{\theta})\right],$$

both $L_T(c_1, 0, \boldsymbol{\theta})$ and $\tilde{\boldsymbol{\theta}}^R(c_1)$ being independent of c_1. The LR process has large-sample counterpart given by

$$LR_T(c_1) = L_T^U(c_1) - L_T^R(c_1),$$

where

$$L_T^R(c_1) = L_T\left(c_1, \boldsymbol{\varphi}^R(c_1)\right),$$

where

$$\boldsymbol{\varphi}^R(c_1) = \left(0, \boldsymbol{\theta}^R(c_1)'\right)',$$

$\boldsymbol{\theta}^R(c_1)$ being defined as

$$\boldsymbol{\theta}^R(c_1) = \arg\max_{\boldsymbol{\theta}} \mathrm{E}\left[l_t\left(c_1, 0, \boldsymbol{\theta}(c_1)\right)\right].$$

The LR statistic is then defined as

$$\widehat{LR}_T = \sup_{c_1} \widehat{LR}_T(c_1),$$

$\widehat{LR}_T(c_1)$ being defined in (30).

Define the expected value of the LR process as

$$R_T(c_1) = \mathrm{E}\left[LR_T(c_1)\right] = T\mathrm{E}\left[l_t\left(c_1, \boldsymbol{\varphi}^U(c_1)\right) - l_t\left(c_1, \boldsymbol{\varphi}^R(c_1)\right)\right].$$

This allows us to define the centred version of the LR process as

$$\widehat{Q}_T(c_1) = \widehat{LR}_T(c_1) - R_T(c_1)$$

and its large-sample counterpart as

$$Q_T(c_1) = LR_T(c_1) - R_T(c_1).$$

Notice that under the null hypothesis $R_T(c_1) \leq 0$; therefore

$$LR_T(c_1) \leq Q_T(c_1)$$

so that the empirical process $LR_T(c_1)$ has an upper bound given by the process $Q_T(c_1)$.

Assume

$$\frac{1}{\sqrt{T}} Q_T(c_1) \Rightarrow Q(c_1),$$

where $Q(c_1)$ is a zero mean Gaussian process with covariance function[26]

$$\begin{aligned} K\left(c_1^{(i)}, c_1^{(j)}\right) &= \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left[Q_T\left(c_1^{(i)}\right) Q_T\left(c_1^{(j)}\right)\right] \\ &= \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left\{\left[\sum_{t=1}^{T} q_t\left(c_1^{(i)}\right)\right]\left[\sum_{t=1}^{T} q_t\left(c_1^{(j)}\right)\right]\right\} \\ &= \mathrm{E}\left[q_t\left(c_1^{(i)}\right) q_t\left(c_1^{(j)}\right)\right], \end{aligned}$$

and associated variance function

$$V(c_1) = K(c_1, c_1)$$

[26] Note that the equality

$$K\left(c_1^{(i)}, c_1^{(j)}\right) = \mathrm{E}\left[q_t\left(c_1^{(i)}\right) q_t\left(c_1^{(j)}\right)\right]$$

does not generally hold, as it requires uncorrelated likelihood components $q_t(c_1)$. This would not be the case if the underlying model was a Markow Switching Model, as discussed in Hansen (1996) **expand with connections to Markov Switching models**.

where

$$q_t(c_1) = l_t\left(c_1, \beta_2^U(c_1), \boldsymbol{\theta}^U(c_1)\right) - l_t\left(c_1, 0, \boldsymbol{\theta}^R(c_1)\right) - \frac{1}{T} R_T(c_1).$$

Therefore, $K\left(c_1^{(i)}, c_1^{(j)}\right)$ can be estimated as

$$\widehat{K}_T\left(c_1^{(i)}, c_1^{(j)}\right) = \sum_{t=1}^{T} \hat{q}_t\left(c_1^{(i)}\right) \hat{q}_t\left(c_1^{(j)}\right),$$

where

$$\hat{q}_t(c_1) = l_t\left(c_1, \hat{\beta}_2^U(c_1), \hat{\boldsymbol{\theta}}^U(c_1)\right) - l_t\left(c_1, 0, \tilde{\boldsymbol{\theta}}^R(c_1)\right) - \frac{1}{T}\widehat{LR}_T(c_1)$$

and $\widehat{V}_T(c_1) = \widehat{K}_T(c_1, c_1)$. Since $LR_T(c_1) \leq Q_T(c_1)$, then

$$\widehat{LR}_T(c_1) \leq \widehat{Q}_T(c_1).$$

Consider now the *standardised* LR function

$$\widehat{LR}_T^*(c_1) = \frac{\widehat{LR}_T(c_1)}{\left[\widehat{V}_T(c_1)\right]^{1/2}}$$

and the corresponding standardised LR statistic

$$\widehat{LR}_T^* = \sup_{c_1} \widehat{LR}_T^*(c_1).$$

Further, define the processes

$$\widehat{Q}_T^*(c_1) = \frac{\widehat{Q}_T(c_1)}{\left[\widehat{V}_T(c_1)\right]^{1/2}}$$

and its large-sample counterpart

$$Q_T^*(c_1) = \frac{Q_T(c_1)}{\left[V_T(c_1)\right]^{1/2}},$$

where

$$Q_T^*(c_1) \Rightarrow Q^*(c_1);$$

$Q^*(c_1)$ is a zero mean Gaussian process defined as

$$Q^*(c_1) = \frac{Q(c_1)}{\left[V(c_1)\right]^{1/2}}$$

with covariance function equal to

$$K^*\left(c_1^{(i)}, c_1^{(j)}\right) = \frac{K\left(c_1^{(i)}, c_1^{(j)}\right)}{\left[V\left(c_1^{(i)}\right)\right]^{1/2}\left[V\left(c_1^{(j)}\right)\right]^{1/2}}.$$

Therefore, under fairly general regularity conditions

$$\Pr\left[\widehat{LR}_T^* \geq x\right] \leq \Pr\left[\sup_{c_1} \widehat{Q}_T^*(c_1) \geq x\right] \to \Pr\left[\text{Sup}Q^* \geq x\right]$$

where

$$\text{Sup}Q^* = \sup_{c_1} Q^*(c_1).$$

The distribution of SupQ^* can be approximated by simulation. This is because $Q^*(c_1)$ is a Gaussian process with zero mean and unit variance, and it is completely characterised by the covariance function $K^*\left(c_1^{(i)}, c_1^{(j)}\right)$, which can be estimated as

$$\widehat{K}_T^*\left(c_1^{(i)}, c_1^{(j)}\right) = \frac{\widehat{K}_T\left(c_1^{(i)}, c_1^{(j)}\right)}{\left[\widehat{V}_T\left(c_1^{(i)}\right)\right]^{1/2}\left[\widehat{V}_T\left(c_1^{(j)}\right)\right]^{1/2}}.$$

Therefore, the simulation procedure to compute the distribution of the process SupQ^*, which is the upper bound to the distribution of the standardised LR statistic $\widehat{LR}_T^*$, is made of the following steps:

1. Generate a sample of $NID(0,1)$ variables $\{u_t^r\}_{t=1}^T$, where the superscript r denotes the replication.
2. Construct
$$\widetilde{LR}^*(c_1) = \frac{\sum_{t=1}^{T} \hat{q}_t(c_1)\, u_t^r}{\left[\widehat{V}_T(c_1)\right]^{1/2}},$$
and compute $\sup_{c_1} \widetilde{LR}^*(c_1)$ for the $r-th$ replication.
3. Repeat steps 1 and 2 above R times: this will provide an approximation to the distribution of SupQ^*.

Conditional upon the data, $\widetilde{LR}^*(c_1)$ is a zero mean Gaussian process with exact covariance function equal to $\widehat{K}_T^*\left(c_1^{(i)}, c_1^{(j)}\right)$, which is an asymptotic approximation to $K^*\left(c_1^{(i)}, c_1^{(j)}\right)$: this means that $\sup_{c_1} \widetilde{LR}^*(c_1)$ will approximately have the same distribution as SupQ^*. Therefore, as $T \to \infty$ and $R \to \infty$, the simulated distribution of $\sup_{c_1} \widetilde{LR}^*(c_1)$ will be the same as the theoretical distribution of SupQ^*. The number of replications for the simulated LR test statistic is $R = 2000$.

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**
	CFIML									
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
$\alpha_1 = \alpha_2 = 1$										
	CFIML									
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**
	CFIML									
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
$\alpha_1 = \alpha_2 = 1$										
	CFIML									
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 $y_{1t} = \delta_1 + \alpha_1 x^r_{1t} + u^r_{1t}$ and $y_{2t} = \delta_2 + \alpha_2 x^r_{2t} + \beta_2 \mathbf{I}\left(y^r_{1t} - c^0_1\right) + u^r_{2t}$, with $c^0_1 = 1.64$ and $u^r_{it} = \left(\gamma_i f^r_t + \varepsilon^r_{it}\right)\left(\gamma^2_i + 1\right)^{-1/2}$, 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. The regressors are generated by $x^r_{it} = \left(\phi_i h^r_t + q^r_{it}\right)\left(\phi^2_i + 1\right)^{-1/2}$, 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 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**
	CFIML									
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
$\alpha_1 = \alpha_2 = 1$										
	CFIML									
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**
	CFIML									
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
$\alpha_1 = \alpha_2 = 1$										
	CFIML									
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 the reduced form in (13), with $c_i^0 = 1.64$ and $u_{it}^r = \left(\gamma_i f_t^r + \varepsilon_{it}^r\right)\left(\gamma_i^2 + 1\right)^{-1/2}$, 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. The regressors are generated by $x_{it}^r = \left(\phi_i h_t^r + q_{it}^r\right)\left(\phi_i^2 + 1\right)^{-1/2}$, where $q_{it}^r \sim NID(0,1)$, $h_t^r \sim NID(0,1)$ and $\phi_i \sim U(0.8,1)$, ϕ_i fixed in repeated samples.

Table 3: Daily stock market returns. Period: 06/08/1990 to 30/06/2005.

Descriptive statistics

Statistics	S&P 500	FTSE 100	DAX 30	SMI	CAC 40
Mean	0.0323	0.0173	0.0211	0.0311	0.0183
Median	0.0631	0.0250	0.0742	0.0488	0.0498
Maximum	5.7708	8.3361	7.1683	7.0489	10.3560
Minimum	-5.5327	-5.6812	-13.0580	-9.1340	-10.2510
Std. Dev.	1.0251	1.0635	1.4215	1.1597	1.3371
Skewness	-0.0292	0.1288	-0.3517	-0.1293	0.0130
Kurtosis	6.0354	6.2713	7.4717	7.0317	7.4367
Jarque-Bera	1436.7**	1678.4**	3194**	2544.1**	3068.4**
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]

Correlation matrix

	S&P 500	FTSE 100	DAX 30	SMI	CAC 40
S&P 500	1.0000	0.5869	0.5567	0.4720	0.5766
FTSE 100		1.0000	0.6575	0.6294	0.6875
DAX 30			1.0000	0.7214	0.7588
SMI				1.0000	0.6890
CAC 40					1.0000

Table 4: Empirical results, CFIML Estimation

	FTSE 100	vs	S&P 500
c	0.95		2.69
β	0.4198		0.1761
n	530		43
π	0.1419		0.0115
W	26.64		11.42
	[0.0001]		[0.0436]
$\log L$		-9935.93	
	DAX 30	**vs**	**S&P 500**
c	3.06		2.44
β	0.5823		0.5035
n	89		69
π	0.0238		0.0185
W	2.83		27.86
	[0.7269]		[0.0000]
$\log L$		-10137.40	
	SMI	**vs**	**S&P 500**
c	3.15		2.69
β	0.5731		0.7403
n	26		43
π	0.0070		0.0115
W	4.50		16.27
	[0.4799]		[0.0061]
$\log L$		-10250.70	
	CAC 40	**vs**	**S&P 500**
c	2.20		2.69
β	0.6438		0.3412
n	164		43
π	0.0439		0.0115
W	14.27		12.27
	[0.0140]		[0.0312]
$\log L$		-9955.72	

Notes: β denotes the contagion coefficient; c the threshold parameter; n the number of crisis periods and π the probability of crisis; W is the Wald statistic for identification of the model.

Table 5: Test for Contagion

	FTSE 100	vs	S&P 500
LR statistic		1.5318	
p-value		0.1910	
	DAX 30	**vs**	**S&P 500**
LR statistic		1.8115	
p-value		0.1230	
	SMI	**vs**	**S&P 500**
LR statistic		1.706	
p-value		0.2200	
	CAC 40	**vs**	**S&P 500**
LR statistic		1.9062	
p-value		0.1180	