When economic theory is not enough... Identification through heteroskedasticity in a likelihood-based approach

Emanuele Bacchiocchi*

Abstract

Since the seminal contribution by Rigobon (2003, *The Review of Economics and Statistics* 85, 777-792) some authors have proposed identification conditions in heteroskedastic bivariate systems of equations. None of them, however, can be generalized lo larger systems and, especially for macroeconomic applications, this represents a strong limitation. This paper shows how the analysis of identification of simultaneous equations systems with different volatility regimes can be reconciled with the conventional likelihood-based setup. We propose a new specification that explicitly models the heteroskedasticity in the residuals, and study the conditions for identification when both heteroskedasticity and traditional restrictions on the parameters are jointly considered. A Full Information Maximum Likelihood (FIML) algorithm is discussed and the small sample performances of estimators and tests on the parameters are studied through Monte Carlo simulations. Finally, this methodology is used to investigate the relationships between sovereign bond yields for some highly indebted EU countries.

Keywords: simultaneous equations model, heteroskedasticity, identification, FIML, contagion, highly indebted EU countries.
JEL codes: C01,C13,C30,C51.

I Introduction

The issue of identification of statistical and econometric models has been thoroughly debated in the literature. The seminal contributions by Haavelmo (1947), Koopmans, Rubin and Leipnik (1950), and Rothenberg (1971) represent milestones in the field[1]. Rothenberg (1971), in particular, summarized, in a single framework based on the information matrix, different approaches to identification, and proposed, as a particular case, necessary and

*DEMM, University of Milan, Via Conservatorio 7, 20122 Milan (Italy), tel +39 02 50321504, fax +39 02 50321450. Email: emanuele.bacchiocchi@unimi.it

[1]Important references for detailed surveys of identification methods in simultaneous systems of equations are Hsiao (1983), Hausman (1983), and Fuller (1987).

sufficient conditions for global and local identification of simultaneous equation systems. The main idea, that nowadays has become the traditional approach, is to restrict the parametric space by imposing restrictions on the parameters. In particular, in the simultaneous equation models, the problem of identification arises because of a non univocal correspondence between the identified parameters of the reduced form, and those of the structural form. Economic knowledge, thus, helps in considering different kinds of constraints on the parameters of the structural form that allow to draw inference based on the information contained in a sample[2]. In some cases, however, the economic knowledge of the problem might not be sufficient to impose such restrictions.

An alternative approach is to substitute the idea of restricting the parametric space with those of finding further information in the data to be included in the identification strategies. In a recent paper, Rigobon (2003) exploits the intuition in Wright (1928) to propose a solution of the identification problem based on the heteroskedasticity in the data. In particular, he provides necessary and sufficient conditions for identification of a bivariate system of simultaneous equations with two or more regimes of volatility, while a necessary condition only for more general systems. Recently, Klein and Vella (2010), Lewbel (2011), and Prono (2008) also use heteroskedasticity to identify simultaneous and mismeasured bivariate equation models. In all these papers the inspiration for identification and estimation of the parameters comes from the instrumental variable approach or generalized method of moments (GMM).

Sentana and Fiorentini (2001), instead, provide conditions for identification in a context of conditional heteroskedastic factor models. The authors discuss how the existence of time-varying heteroskedasticity in the factors has important implications for the identification of simultaneous equations systems, Markov-switching models, and structural VAR autoregressions. In the structural VAR context Lanne and Lütkepohl (2008) and Lanne et al. (2010), provide rather general conditions for identification using heteroskedasticity of the structural shocks.

In the present paper we concentrate on a particular specification for the heteroskedasticity of the structural shocks in simultaneous equation systems that enables to derive and manage the likelihood function. The identification conditions, thus, are studied following the Rothenberg (1971) approach, and the estimation of the parameters is performed through a Full Information Maximum Likelihood (FIML) procedure.

Although our definition of heteroskedasticity appears as a special case of the one discussed in Sentana and Fiorentini (2001) and Klein and Vella (2010), the novelty of the present paper is to study the identification of the structural parameters when the traditional approach, consisting in the restriction of the parametric space by means of constraints on the parameters, is enriched with the information concerning some form of heteroskedasticity contained in the data. The idea thus, is to mix the information coming from the economic theory (the restrictions of the parameters) with particular features observed in the data (heteroskedasticity) and provide necessary and sufficient conditions for the identification of the structural parameters in the case of general systems of equations.

This strategy a) allows us to obtain identification conditions for systems characterized by more than two equations (Rigobon, 2003, Klein and Vella, 2010, Lewbel, 2011, and Prono,

[2] An important stream of literature, particularly relevant for the aims of the present paper, relies on the LISREL models (Joreskog and Sorbom, 1984) where identification in multiple equation simultaneous systems is obtained via equality restrictions on the covariance matrix, sometimes joint with restrictions on structural parameters.

2008, consider bivariate systems only), b) does not require that all the variables feature heteroskedasticity, and c) does not require to restrict the covariance matrix of the structural shocks to be diagonal (or to resort to unobservable common shocks to diagonalize the residuals). In particular, point a) reveals to be an important improvement in macroeconomic applications where more than two equations generally characterize theoretical models.

The model presented and discussed in the paper reveals to be extremely useful for detecting whether, in particular periods of time, some structural shocks are transmitted to the other variables through more complicated channels than those generally used in traditional systems of equations. Such 'particular periods of time' might be represented by higher volatility of the shocks that, when appropriately considered, provides useful insights for the identification of the structural parameters, and enriches the structure of the model by proposing further channels for the propagation of the shocks.

As an example, in periods of high instability of the financial markets, a shock hitting one particular market might propagate in a different way than in relatively tranquil periods. The effects of the same shock, in different periods of time, might be completely different. The turbulences of the markets could either amplify the effects of the shocks in the same market in which it originates, or allow for a propagation to the other financial markets, or both the effects. All these aspects of the model will be thoroughly discussed in the empirical application.

The rest of the paper is organized as follows. In Section II we first present the statistical model and then derive the conditions for identification of the structural parameters, with a discussion on the relationships between this approach and the existing literature. Section III describes the statistical inference while Section IV uses this methodology to investigate about the transmission of financial shocks among some highly indebted EU countries over the last years. Section V provides some concluding remarks. All technical proofs, a generalization for the case of more than two levels of volatility for each dependent variable, and some simulation exercises are left in the Appendix.

II Identification in a simultaneous equation model with heteroskedasticity

II.1 A simultaneous equation model with heteroskedastic errors

The idea, that takes directly inspiration from Rigobon (2003), is to increase the number of relations that link the parameters in the reduced form to those in the structural form. In this section we present a new specification for simultaneous equations systems that explicitly models the heteroskedasticity of the structural shocks. As in the previous case, we first discuss the simplest case of two regimes of volatility only, while a generalization will be provided in the next sections. A simplified simultaneous equation model with different regimes of volatility can be written as:

$$By_t = (I_g + AD_t)\,\varepsilon_t \tag{1}$$

where y_t is the vector of g endogenous variables, ε_t is the vector of structural shocks, B is the $(g \times g)$ invertible matrix of simultaneous relationships among the endogenous variables. A is a $(g \times g)$ matrix capturing further transmission channels of propagation of structural shocks, while D_t is a diagonal matrix assuming only $0 - 1$ values, indicating whether, at time t, the i-th endogenous variable is in a state of high (1) or low (0) volatility. In the simplest case of only two equations, the system becomes[3]:

$$\begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \left(I_2 + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \right) \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}. \tag{2}$$

In this simple case, there are four possible volatility regimes, given by the possible combinations of d_1 and d_2.

The structural shocks ε_t are assumed to be uncorrelated (this assumption will be relaxed in the following sections) with a constant covariance matrix $E\left(\varepsilon_t \varepsilon_t'\right) = \Lambda$. When both variables are in a state of low volatility, i.e. $D_t = 0$, the model appears as a standard system of equations, without restrictions on the B parameters. When one or both variables are in a state of high volatility, instead, the a_{ii} parameters act as multiplicative factors for the structural shocks, while the off diagonal values a_{ij} allow for the propagation of shocks to other variables. These interpretations, of course, apply to the more general model in Eq. (1).

Based on the invertibility of the B matrix, the reduced form of the model simply becomes:

$$y_t = B^{-1} \left(I_g + AD_t \right) \varepsilon_t \tag{3}$$

or equivalently

$$y_t = B^{-1} C_t \varepsilon_t \tag{4}$$

where $C_t = (I_g + AD_t)$. The covariance matrix of the endogenous variables is $E\left(y_t y_t'\right) = B^{-1} C_t \Lambda C_t' B^{-1\prime}$, and changes over time because of C_t.

Example 1 Let consider the bivariate simultaneous equations model in Eq. (2), characterized by two regimes of volatility. In particular, in the first regime only the first endogenous variable y_{1t} is in a state of high volatility while, after a certain time T_B, both $(y_{1t}, y_{2t})'$ move to a state of high volatility. The model thus can be written as

$$\begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \left(I_2 + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right) \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \text{ for } t < T_B$$

and

$$\begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \left(I_2 + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \text{ for } t \geq T_B,$$

where D_1 and D_2, the matrices describing the regimes of volatility before and after the break, are defined as

$$D_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad , \text{ for } t < T_B$$

$$D_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \text{ for } t \geq T_B.$$

[3] A similar specification has been proposed by Favero and Giavazzi (2002) in which, however, the d_{it} are simple intervention dummies, and the identification problem has been solved with exclusion restrictions in the dynamic part of the model.

Actually, in terms of the identification of the model, as we will discuss in the next sections, it doesn't matter whether the two regimes come one after the other (they can be mixed in a more complicated way), what is important is to know the description of the two regimes of volatility D_1 and D_2. ■

Example 2 Suppose now that there are three regimes of volatility. In the first the two endogenous variables are in a state of low volatility, then the first moves to a state of high volatility (while the second remains in the state of low volatility), and finally, the second achieves the high volatility state too. The model will be equivalent to the one described in Eq.(2), but now there are three matrices describing the different regimes of volatility:

$$D_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad D_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \blacksquare$$

The heteroskedasticity, thus, is intended as different regimes of volatility that might apply to one or more variables in the system. As in Rigobon (2003), in this approach it is only required that some form of heteroskedasticity is present in the data, such as crisis, policy shifts, changes in collecting the data, or cross-sectional peculiarities. As we will discuss in the next section, alternative approaches use ARCH-based model for the residuals of the reduced form in order to obtain identification.

II.2 Specification and identification of a model with two regimes

In this section we study the identification of the simultaneous equation system in Eq. (1) in the simplest case of $s = 2$ regimes of volatility. The generalization is not straightforward and will be the argument of the next sections. In the case of two regimes of volatility we will have only two distinct matrices D_1 and D_2 that, at each instant t, show the state of volatility the system is. The model, thus, can be rewritten with two separate equations, one for each regime:

$$By_t = (I_g + AD_1)\,\varepsilon_t \tag{5}$$
$$By_t = (I_g + AD_2)\,\varepsilon_t \tag{6}$$

where Eq. (5) is for the observations in the first state of volatility, and Eq. (6) for those in the second. The associated covariance matrices for the error terms are:

$$E\left(C_t\varepsilon_t\varepsilon_t'C_t'\right) = (I_g + AD_1)\,\Lambda\,(I_g + AD_1)'$$
$$E\left(C_t\varepsilon_t\varepsilon_t'C_t'\right) = (I_g + AD_2)\,\Lambda\,(I_g + AD_2)'.$$

The reduced form of the model can be written as

$$y_t = B^{-1}(I_g + AD_1)\,\varepsilon_t$$
$$y_t = B^{-1}(I_g + AD_2)\,\varepsilon_t$$

with the two covariance matrices for the dependent variables in the two regimes:

$$\begin{aligned} E\left(y_t y_t'\right) &= B^{-1}\left(I_g + AD_1\right)\Lambda\left(I_g + AD_1\right)' B^{-1\prime} &= \Omega_1 \\ E\left(y_t y_t'\right) &= B^{-1}\left(I_g + AD_2\right)\Lambda\left(I_g + AD_2\right)' B^{-1\prime} &= \Omega_2. \end{aligned}$$

If we assume that the structural shocks ε_t behave as a multivariate normal variable, the identification can be studied as in the traditional simultaneous equation models, i.e. concentrating on the relationships between the parameters in the structural and reduced forms. The normality is required to impose that the distribution of y_t depends only on the parameters of the reduced form[4]. Following Rothenberg (1971), the identifiability of the structure depends on the uniqueness of the solution of the following system

$$\left(I_g + AD_1\right)^{-1} B\Omega_1 B' \left(I_g + AD_1\right)^{-1\prime} - \Lambda = 0 \qquad (7)$$

$$\left(I_g + AD_2\right)^{-1} B\Omega_2 B' \left(I_g + AD_2\right)^{-1\prime} - \Lambda = 0 \qquad (8)$$

$$R_A vec\, A - r_A = 0 \qquad (9)$$

$$R_B vec\, B - r_B = 0 \qquad (10)$$

$$R_\Lambda v\left(\Lambda\right) - r_\Lambda = 0 \qquad (11)$$

where Eqs. (9)-(11) are a set of q_A, q_B, and q_Λ linear restrictions on the parameters A, B and Λ (respectively), that can also be written in the equivalent explicit form

$$vec\, A = S_A \gamma_A + s_A \qquad (12)$$

$$vec\, B = S_B \gamma_B + s_B \qquad (13)$$

$$v\, \Lambda = S_\Lambda \gamma_\Lambda + s_\Lambda \qquad (14)$$

where

$$\begin{array}{ll} R_A S_A = 0 & R_A s_A = r_A \\ R_B S_B = 0 & R_B s_B = r_B \\ R_\Lambda S_\Lambda = 0 & R_\Lambda s_\Lambda = r_\Lambda. \end{array}$$

The vector $v\left(\Lambda\right)$ denotes the $\frac{1}{2}g\left(g+1\right)$ elements that is obtained from $vec\,\Lambda$ by eliminating the supra diagonal elements of Λ or, equivalently, $D_g v\left(\Lambda\right) = vec\,\Lambda$, with D_g the *duplication matrix*[5].

Throughout, use is made of the following notation: K_{gs} is the $g^2 s^2 \times g^2 s^2$ *commutation matrix* as defined in Magnus and Neudecker (2007), $N_{gs} = 1/2\left(I_{gs} + K_{gs}\right)$, while $\tilde{D}_g$, defined in Magnus (1988), is a $g^2 \times g(g-1)/2$ full-column rank matrix such that for any $g(g-1)/2$-dimensional vector v, it holds $\tilde{D}_g v := vec(H)$, with H a $g \times g$ skew-symmetric matrix ($H = -H'$).

Suppose to consider the simple and realistic case that in the first regime all variables are in a state of high volatility ($D_1 = I_g$), and in the second all are in a state of low volatility ($D_2 = 0_g$), which is the case investigated in Rigobon (2003). The following proposition presents the necessary and sufficient condition for identification of the structural parameters.

Proposition 1 *Consider the simultaneous equations model with $s = 2$ regimes of volatility described in Eqs. (5)-(6), with $D_1 = I_g$ and $D_2 = 0_g$. Then (A_0, B_0, Λ_0) are locally identified*

[4] Any other stochastic variable univocally defined by the first two moments will provide the same results in terms of identification.

[5] See Magnus and Neudecker (2007), pag 57.

if and only if the following $\left(2g^2 \times 2g^2 + g(g+1)/2 - q\right)$ *matrix*

$$\begin{pmatrix} -2N_g D_1^* S_A & 2N_g E_1^* S_B & -D_g S_\Lambda \\ & 2N_g E_2^* S_B & -D_g S_\Lambda \end{pmatrix} \tag{15}$$

has full column rank. The three non-singular matrices D_1^*, E_1^* *and* E_2^* *are defined as follows*

$$\begin{aligned} D_1^* &= \left[C_1^{-1} B\Omega_1 B' C_1^{-1\prime} \otimes C_1^{-1}\right] \\ E_1^* &= \left[C_1^{-1} B\Omega_1 \otimes C_1^{-1}\right] \\ E_2^* &= \left[I_g B\Omega_2 \otimes I_g\right]. \end{aligned} \tag{16}$$

A necessary condition for identification is that $q \geq g^2 + g(g-1)/2$, *where* $q = q_A + q_B + q_\Lambda$ *represents the number of restrictions in the* A, B, *and* Λ *matrices.*

Proof. The proof of Proposition 1 is discussed in the Appendix A.1.

The necessary and sufficient condition reported in Eq. (15) is specific to the case of $s = 2$ states of volatility described by the $D_1 = I_g$ and $D_0 = 0_g$ matrices. However, it provides necessary and sufficient conditions for identification of simultaneous equations models where the presence of heteroskedasticity in the data is mixed with the traditional information coming from the economic theory and expressed in the form of linear restrictions on the parameters of the model[6].

The necessary and sufficient condition in Proposition 1 generalizes those in Rigobon (2003) in different directions. First, when there are two regimes of volatility only, the rank condition in Eq. (15) is more general than the corresponding condition in Proposition 1 of Rigobon (2003), in that it presents a sufficient condition for a system of $g \geq 2$ equations, and not simply a bivariate system. Second, we don't need the covariance matrix Λ to be diagonal. This problem has been discussed in Rigobon (2003) and solved by introducing common shocks in the model, which is equivalent to relaxing the assumption on the correlation of the structural shocks. This alternative strategy, however, makes it more difficult to consider possible restrictions on the covariance matrix of the structural shocks Λ, as instead considered in our Proposition 1. Third, as already said, the necessary and sufficient condition in Eq. (15) allows to discuss identification when heteroskedasticity and linear restrictions on the parameters are jointly considered.

The price to pay, given the more complicated specification of the model, is to study the identification of the parameters only locally, and not globally as instead considered by Rigobon (2003), Lewbel (2011), Prono (2003), and Klein and Vella (2010) for (triangular or full) bivariate systems of equations.

The following corollary concentrates on the particular case in which the covariance matrix of the structural shocks is restricted to be the identity matrix, as commonly considered in the SVAR literature.

[6]If one is not interested in imposing restrictions on the parameters of the predetermined variables, the identification analysis can be restricted to the concentrated likelihood, as commonly done in the traditional SVAR literature.

Corollary 1 *Consider the simultaneous equations model with $s = 2$ regimes of volatility described in Eqs. (5)-(6), with $D_1 = I_g$ and $D_2 = 0_g$. When $\Lambda = I_g$, then the necessary and sufficient condition for (A_0, B_0) to be locally identified is that the $\left(g^2 + q_B\right) \times g^2 - q_A + g\,(g-1)$ matrix*

$$\begin{pmatrix} N_g D_1^* S_A & N_g E_1^* E_2^{*-1} \widetilde{D}_g \\ 0 & R_B E_2^{*-1} \widetilde{D}_g \end{pmatrix} \tag{17}$$

has full column rank. The three non-singular matrices D_1^*, E_1^* and E_2^* are defined as in Eq. (16).

A necessary condition for identification is that $q_A + q_B \geq g\,(g-1)$, where q_A and q_B represent the number of restrictions in the A and B matrices, respectively.

Proof. The proof of Corollary 1 is discussed in the Appendix A.2.

The results of our Corollary 1 define identification rules for the A and B matrices when the presence of two regimes of volatility is mixed with traditional restrictions on the parameters. Interestingly, if one interprets the particular specification of our model as a heteroskedastic version of the AB-SVAR formulation, as described in Amisano and Giannini (1997) and Lütkepohl (2005), the necessary order condition indicates a minimum of $g^2 - g$ restrictions, against a minimum of $g^2 + g\,(g-1)\,/2$ restrictions in the traditional AB-model without accounting for the heteroskedasticity in the data.

In two recent papers, Lanne and Lütkepohl (2008) and Lanne et al. (2010), using a more traditional specification, consider the identification of SVAR models (C-model or K-model in the terminology of Amisano and Giannini, 1997) with different levels of volatility. Using a well known result of matrix algebra in which symmetric and positive definite matrices can be simultaneously diagonalized by means of common squared matrices and specific diagonal matrices, they show that simply exploiting the presence of two or more levels of volatility is sufficient for the structural parameters (the B matrix) to be identified (without the need of parameters restrictions, like the standard Cholesky triangularization).

The following remark reconciles these results with those obtained in our Corollary 1, although with a different parametrization of the SVAR model.

Remark 1. When there are no theoretical reasons to impose restrictions on the B matrix, and thus $R_B = 0$, the necessary and sufficient condition in Corollary 1 reduces to

$$\begin{bmatrix} N_g D_1^* S_A & N_g E_1^* E_2^{*-1} \widetilde{D}_g \end{bmatrix} \tag{18}$$

and will be of full column rank only when the number of restrictions on the A matrix is at least equal to $g\,(g-1)$, as suggested by the previous necessary condition. However, as all endogenous variables show a period of high volatility, the only possibility to capture such increase in all the variances is when A is diagonal. This condition respects the necessary condition and, in a heteroskedastic SVAR framework, is equivalent to the results in Lanne and Lütkepohl (2008) and Lanne et al. (2010), although obtained from a different specification of the model. The specification adopted in the present paper, instead, is much more general in that does not require the A matrix to be diagonal.

Remark 2. Looking at the necessary and sufficient condition in Eq. (17), the lower-right block $(R_B E_2^{*-1} \widetilde{D}_g)$ is clearly equivalent to the necessary and sufficient condition for the standard K-model developed by Amisano and Giannini (1997). When the restrictions on the B matrix, R_B, are sufficient to satisfy the well known sufficient condition for the K-model (or the C-model) to be identified, i.e. $R_B E_2^{*-1} \widetilde{D}_g$ is of full column rank, then the parameters of the A matrix will be identified if and only if the $R_A D_1^{*-1} \widetilde{D}_g$ matrix is of full column rank $g\,(g-1)\,/2$, which is equivalent to the identification of the structural parameters in the traditional C-model.

In practical applications, the necessary and sufficient conditions in Eq. (15) and Eq. (17) can be numerically checked, as suggested in Giannini (1992), using random values for A and B that satisfy the restrictions in Eqs. (9)-(11), or better, as discussed later on, using random numbers for the three vectors γ_A, γ_B, and γ_Λ such that the restrictions in Eqs. (12)-(14) hold[7].

The case treated in this section, however, is limited in two directions: a) it considers two states of volatility only, b) the sufficient condition has been calculated based on the particular specifications of D_1 and D_2. In the next sections we provide a generalization that fills these two shortcomings.

II.3 Related literature

Over the last years other authors have proposed approaches to obtain identification using heteroskedasticity in the residuals. This paper is directly inspired and aims at generalizing the idea in Rigobon (2003) but, in order to relate our approach to the existing literature, it is worth discussing the main characteristics of those alternative approaches. The proposal of Rigobon (2003), originally introduced by Wright (1928), consists in using the second moments to increase the number of relations between the parameters in the reduced and structural forms. Lanne and Lütkepohl (2008) and Lanne et al. (2010) use the same idea to identify the structural parameters in structural VAR models.

Klein and Vella (2010) use the heteroskedasticity of the residuals to identify the structural parameters in bivariate triangular systems. The idea is that when the distribution (or, simply, the second moments) of the errors in the two equations does depend on the vector of exogenous variables, identification of the parameters of the controlled regression is guaranteed when it is possible to consistently estimate the non-constant conditional relations among the two residuals. The estimation approach is then based on estimating a semiparametric model of heteroskedasticity in each equation. Prono (2008) also discusses identification in linear bivariate triangular models where structural errors follow a bivariate and diagonal GARCH$(1,1)$ process.

Lewbel (2010), instead, considers bivariate models with mismeasured or endogenous regressors. Identification in triangular and fully simultaneous systems can be obtained by imposing restrictions on particular second moments involving regressors and heteroskedastic

[7]A Gauss 8.0 package for checking for identification and estimating the unknown parameters of the heteroskedastic simultaneous equations model developed in this paper can be obtained from the author under request.

residuals. The estimation strategy takes the form of modified two stage least squares or generalized method of moments. Rigobon's identification result can be interpreted as a special case of Lewbel's approach, in which the variable used for heteroskedastic identification is simply a dummy variable indicating a high versus low volatility regime.

Sentana and Fiorentini (2001), in a context of conditionally heteroskedastic factor models, provide identification conditions that can be applied in a large number of cases, like residuals following GARCH specifications[8], regime switching processes[9], or structural VAR models[10].

The first difference with respect to the existing literature lies on the specification of the model used to treat heteroskedasticity. Rigobon (2003) requires only that something has happened to justify a shift in the covariance matrix, without modeling directly the source of heteroskedasticity. In our approach, instead, we propose a specification that models such heteroskedasticity, and enables us to treat cases in which only some variables show clusters of different volatility. Clearly, the results in Rigobon (2003) are a special case of our specification when the A matrix is simply diagonal. When some dependent variables, over the investigated period, do not show evident shifts in volatility, it would be useful, in order to gain degrees of freedom for testing the stability of the structural parameters, to directly impose constant variances (simply by imposing zero values in the A matrix).

Moreover, as we will see later on, if the model is already (exactly-)identified our specification allows to test for such constant variances in the standard inference setup. The general condition for identification in Sentana and Fiorentini (2001), which are valid for a very general class of processes for the dynamics of conditional heteroskedasticity including our formulation of the model, requires that no more than one variance is constant over time. Obviously, this result remains valid in our model too, but our strategy allows for a joint analysis of identification, that involves both the parameters governing heteroskedasticity and the structural ones. In other words, restrictions in the structural parameters can be sufficient to identify the model even when more than one variance remains constant over time. In fact, although our specification for the heteroskedasticity in the residuals is a particular case of the general definition provided in Sentana and Fiorentini (2001), the effort in our strategy is to study identification when the traditional approach of restricting the parametric space is enriched with the characteristics of heteroskedasticity contained in the data.

As already mentioned, a second difference with respect to Rigobon (2003), Klein and Vella (2010), and Lewbel (2011) refers to the strategy used for estimating the parameters. The particular specification we propose in the paper allows to define the likelihood function, and derive analytically the score vector and the information matrix, that will be used in a recursive algorithm to maximize the likelihood. This full information approach, other than providing more efficient estimators than those based on the instrumental variables ones generally used in the literature, allows us to use likelihood ratio tests in order to test for overidentifying restrictions. An argument for future research, instead, is to lighten the hypothesis that the structural parameters remain constant over time, and connecting this stream of research based on heteroskedastic error terms with the traditional models accounting for structural breaks.

---

[8] See Caporale, Cipollini and Demetriades (2005), Dungey and Martin (2001), King et al. (1994), Rigobon (2002).

[9] See Caporale, Cipollini and Spagnolo (2005) and Rigobon and Sack (2003, 2004).

[10] See Normandin and Phaneuf (2004).

II.4 Specification and identification: The general case

In order to generalize the results of the previous section we need to write the model in a different way. Once we have information on the different states of volatility, we can easily build a $(T \times s)$ matrix P indicating, at each instant t, the state of volatility characterizing the y_t. As an example, let define

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{19}$$

indicating that the system is characterized by $s = 3$ states of volatility, and that for the first two periods the active state is state 1, then state 2, and when $t = 5$, state three. Thus, using the Hadamard product[11] $\odot$, we can reorganize the data as

$$Y^* = \left(i_s' \otimes Y\right) \odot \left(P \otimes i_g'\right) \tag{20}$$

where i_s and i_g are two unit vectors of dimension $(s \times 1)$ and $(g \times 1)$, respectively, while Y is the $(T \times g)$ matrix containing the data on the dependent variables. As an example where P is defined as in Eq. (19), and where $y_t = \begin{pmatrix} y_{1t} & y_{2t} \end{pmatrix}'$, the Y^* matrix, becomes

$$Y^* = \begin{pmatrix} y_{11} & y_{21} & 0 & 0 & 0 & 0 \\ y_{12} & y_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & y_{13} & y_{23} & 0 & 0 \\ 0 & 0 & y_{14} & y_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & y_{15} & y_{25} \end{pmatrix}. \tag{21}$$

In the same way, we can define the $(T \times gs)$ ε^* matrix containing the error terms

$$\varepsilon^* = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{21} & 0 & 0 & 0 & 0 \\ \varepsilon_{12t} & \varepsilon_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_{13} & \varepsilon_{23} & 0 & 0 \\ 0 & 0 & \varepsilon_{14} & \varepsilon_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & \varepsilon_{15} & \varepsilon_{25} \end{pmatrix} \tag{22}$$

which allows us to rewrite the model as:

$$\left(I_s \otimes B\right) y_t^* = A^* \varepsilon_t^* \tag{23}$$

where y_t^* and ε_t^* are vectors obtained from the t-th row of the Y^* and ε^* matrices, respectively. Furthermore, if we allow for k predetermined variables, the structural form of the model can be written as

$$\left(I_s \otimes B\right) y_t^* + \left(I_s \otimes \Gamma\right) x_t^* = A^* \varepsilon_t^*, \tag{24}$$

in which x_t^* is the $(ks \times 1)$ vector of predetermined variables expressed as in Eq. (21) and Γ is the related $(g \times k)$ matrix of coefficients. The $(gs \times gs)$ A^* and D block diagonal matrices

[11] For a general discussion on the Hadamard product, see Magnus and Neudecker (2007), pp 53-54 and 71.

are defined as

$$A^* = (I_{gs} + (I_s \otimes A)\, D) \tag{25}$$

$$D = \begin{pmatrix} D_1 & & \\ & \ddots & \\ & & D_s \end{pmatrix} \tag{26}$$

where D_i is the diagonal $(g \times g)$ matrix describing the i-th state of volatility. More precisely, for the i-th state of volatility, it presents $D_{ijj} = 1$ whether the j-th endogenous variable is in a state of high volatility and 0 if it is in a state of low volatility. The covariance matrix of the structural shocks is, as before, $E\left(\varepsilon_t \varepsilon_t'\right) = \Lambda$, but using the new notation that highlights the state of volatility, we obtain, for example

$$\Lambda_t = E\left(\varepsilon_t^* \varepsilon_t^{*\prime}\right) = \begin{pmatrix} \Lambda & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \tag{27}$$

in the case the system is in the first state of volatility at time t. The dependence on t of this matrix, however, is only apparent in that in all volatility states we impose the same covariance matrix for the structural shocks, that instead hit the endogenous variables in a different way *via* the particular combination of A^* and D. The particular specification for the data and the model allows us to select, at each t, the way the structural shocks are amplified and propagated to the different endogenous variables in the system.

The reduced form of the model can be easily obtained as

$$\begin{aligned} y_t^* &= (I_s \otimes \Pi)\, x_t^* + u_t^* \\ &= -\,(I_s \otimes B)^{-1} (I_s \otimes \Gamma)\, x_t^* + (I_s \otimes B)^{-1} A^* \varepsilon_t^* \end{aligned} \tag{28}$$

where

$$(I_s \otimes \Pi) = -\,(I_s \otimes B)^{-1} (I_s \otimes \Gamma) \tag{29}$$

and, using the same example as before,

$$\Omega_t = E\left(u_t^* u_t^{*\prime}\right) = (I_s \otimes B)^{-1} A^* E\left(\varepsilon_t^* \varepsilon_t^{*\prime}\right) A^{*\prime} \left(I_s \otimes B^{-1}\right)' = \begin{pmatrix} \Omega_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}. \tag{30}$$

When we consider all the states of volatility, the covariance matrix of the reduced form error terms can consequently be defined as

$$\Omega = (I_s \otimes B)^{-1} A^* E\left(\varepsilon_t^* \varepsilon_t^{*\prime}\right) A^{*\prime} \left(I_s \otimes B^{-1}\right)' = \begin{pmatrix} \Omega_1 & & & \\ & \Omega_2 & & \\ & & \ddots & \\ & & & \Omega_s \end{pmatrix}. \tag{31}$$

Assumption 1 *The vectors* $\{\varepsilon_t, t = 1 \ldots, T\}$ *are independent and identically distributed as* $N(0, \Lambda)$ *with* Λ *a positive definite* $(g \times g)$ *matrix of unknown parameters.*

Assumption 2 *The $(T \times k)$ matrix of predetermined variables has full column rank.*

Assumption 3 *The parameters B_0 and Γ_0 do not change among the different states of volatility.*

Given the three assumptions here above, and a set of linear restrictions on the coefficients of the structural form, the next proposition generalizes the results obtained in Section II.2 for models with $s \geq 2$ distinct regimes of volatility. Moreover, following Magnus and Neudecker (2007) p. 56, let define the matrix H such that, given two matrices $A\,(m \times n)$ and $B\,(p \times q)$ then $vec\,(A \otimes B) = (H \otimes I_p)\, vec\, B$, with $H = (I_n \otimes K_{qm})\,(vec\, A \otimes I_q)$.

Proposition 2 *Consider the simultaneous equations model in Eqs. (23)-(28) under the Assumptions 1-3 previously expressed. Assume further that prior information is available in the form of linear restrictions on A,B, Γ, and Λ (in implicit or, equivalently, in explicit form):*

$$\begin{aligned} &R_A vec\, A = r_A && vec\, A = S_A \gamma_A + s_A \\ &R_B vec\, B = r_B && vec\, B = S_B \gamma_B + s_B \\ &R_\Gamma vec\, \Gamma = r_\Gamma && vec\, \Gamma = S_\Gamma \gamma_\Gamma + s_\Gamma \\ &R_\Lambda v\, \Lambda = r_\Lambda && v\, \Lambda = S_\Lambda \gamma_\Lambda + s_\Lambda . \end{aligned}$$

Then $(A_0, B_0, \Gamma_0, \Lambda_0)$ are locally identified if and only if the matrix

$$\begin{pmatrix} 0 & (\Pi' \otimes I_g)\, S_B & S_\Gamma & 0 \\ -2N_{gs} J_{21} S_A & 2N_{gs} J_{22} S_B & 0 & -J_{23} D_g S_\Lambda \end{pmatrix} \tag{32}$$

has full column rank, where the three matrices J_{21}, J_{22}, and J_{23} are defined as follows

$$J_{21} = \left[A^{*-1}\,(I_s \otimes B)\,(I_s \otimes \Omega)\,(I_s \otimes B)'\, A^{*-1\prime} D \otimes A^{*-1}\right](H \otimes I_g) \tag{33}$$

$$J_{22} = \left[A^{*-1}\,(I_s \otimes B)\,(I_s \otimes \Omega) \otimes A^{*-1}\right](H \otimes I_g) \tag{34}$$

$$J_{23} = (H \otimes I_g) \tag{35}$$

and D_g is the duplication matrix. A necessary condition for identification is that

$$gs\,(gs+1)\,/2 + q \geq 2g^2 + g\,(g+1)\,/2 \tag{36}$$

where $q = q_A + q_B + q_\Gamma + q_\Lambda$ is the total number of restrictions on the unknown parameters A, B, Γ, and Λ, and $H = (I_s \otimes K_{gs})\,(vec\, I_s \otimes I_g)$.

Proof. The proof of Proposition 2 is discussed in the Appendix A.3.

The following corollary, instead, proposes a necessary and sufficient condition for the identification of the parameters when no linear restrictions on the parameters are imposed.

Corollary 2 *Consider the simultaneous equations model in Eqs. (23)-(28) under the Assumptions 1-3 previously expressed. Then $(A_0, B_0, \Gamma_0, \Lambda_0)$ are locally identified if and only* *if the matrix*

$$\left(\begin{array}{ccc} -2N_{gs}J_{21} & 2N_{gs}J_{22}\left[I_{g^2} - \left(\Pi' \otimes I_g\right)^{+}\left(\Pi' \otimes I_g\right)\right] & -J_{23}D_g \end{array} \right) \tag{37}$$

has full column rank, with J_{21}, J_{22}, *and* J_{23} *defined as in Eqs. (33)-(35).*

Proof. The proof of Corollary 2 is discussed in the Appendix A.4.

Remark 3. Interestingly, if $rank\left(\Pi'\right) = k \geq g$, then $\left(\Pi' \otimes I_g\right)^{+}\left(\Pi' \otimes I_g\right) = I_{g^2}$ and the necessary and sufficient condition in Corollary 2 reduces to check whether

$$\left(\begin{array}{cc} -2N_{gs}J_{21} & -J_{23}D_g \end{array} \right) \tag{38}$$

has full column rank.

The approach we follow for studying the identification is based on Rothenberg (1971) and considers, as in the specific case analyzed in the previous section, the system of homogeneous equations that links the parameters in the structural and reduced forms. The necessary condition, thus, refers to the number of equations of this system, that needs to be larger than the number of unknowns.

Including different levels of volatility can be an alternative strategy to increase the number of equations in the system of homogeneous equations. The price to pay, in our model, is to include more parameters than in standard systems of equations, due to the A matrix capturing the multiplicative (and eventually, the propagation) of the structural shocks. This, however, does not prevent the possibility of identifying the parameters without introducing restrictions as highlighted in Corollary 2.

Corollary 3 *Without any further restriction on the parameters, a necessary condition for identification is that there are at least* $s = 3$ *different states of volatility.*

Proof. The proof of Corollary 3 is discussed in the Appendix A.5.

The equivalent *order condition* for systems with different levels of volatility concerns the minimum number of states in order to have, at least, as many distinct equations as unknowns in the system. Corollary 3 states that a minimum of three different states of volatility is necessary for making the parameters identifiable. The main result of this corollary thus, is that, differently from the standard simultaneous equations models, when allowing for clusters of heteroskedasticity in the residuals, we do not need any restriction on the parameters to reach local identification. The first two assumptions are necessary in order to assume that (i) the joint distribution of the endogenous variables y_t depends on $(A_0, B_0, \Gamma_0, \Lambda_0)$ only through the reduced form parameters (Π_0, Ω_0); and (ii) Π_0 and Ω_0 are globally identified. Assumption 3, instead, is necessary to identify the (B_0, Γ_0) structural parameters. When some variances remain constant over time, in order to identify the parameters in A_0 it is necessary to restrict with zero values the parameters on the corresponding columns of the A_0 matrix.

The necessary and sufficient condition, which can be interpreted as the *rank condition* in the traditional systems of equations, is much more complicated in that it depends on the combinations of low-high volatility states as described in the D matrix. All the technical details are discussed in the Appendix A.3-A.5.

III Estimation and Inference

In this section we turn to the problem of estimating simultaneous equations models with different levels of volatility, assuming that some sufficient conditions for identification are satisfied. We propose a Full-Information Maximum Likelihood (FIML) estimator that is based on the maximization of the likelihood function of the structural form of the model. To simplify the notation let define the following matrices

$$B^* = (I_s \otimes B), \quad \Gamma^* = (I_s \otimes \Gamma), \quad \Lambda^* = (I_s \otimes \Lambda), \quad C = \begin{bmatrix} I_s \otimes B^{-1\prime} & 0 \end{bmatrix} \tag{39}$$

and

$$T^* = (T^{**} \otimes I_g) \quad \text{with} \quad T^{**} = \begin{pmatrix} T_1 & & \\ & \ddots & \\ & & T_s \end{pmatrix} \tag{40}$$

indicating the number of elements in the sample for each state of volatility and,

$$Q^* = E\left(z_t^* z_t^{*\prime}\right) = E\left[\begin{pmatrix} y_t^* \\ x_t^* \end{pmatrix} \begin{pmatrix} y_t^{*\prime} & x_t^{*\prime} \end{pmatrix}\right] \tag{41}$$

$$= \begin{bmatrix} (I_s \otimes \Pi) Q_x^* (I_s \otimes \Pi) + \Omega & (I_s \otimes \Pi) Q_x^* \\ Q_x^{*\prime} (I_s \otimes \Pi)' & Q_x^* \end{bmatrix} \tag{42}$$

with $Q_x^* = E(x_t^* x_t^{*\prime})$. Finally, let define the following H matrix as

$$H = \begin{pmatrix} \left(D \otimes A^{*-1}\right)(H_A \otimes I_g) & 0 & 0 \\ 0 & -\begin{pmatrix} \left(A^{*\prime} B^{*-1\prime} \otimes A^{*-1}\right)(H_B \otimes I_g) & 0 \\ 0 & (H_\Gamma \otimes I_k) \end{pmatrix} & 0 \\ 0 & 0 & (H_\Lambda \otimes I_g) D_g \end{pmatrix} \tag{43}$$

where, as before, the generic matrix H_M is defined such that, given two matrices $Q\,(m \times n)$ and $M\,(p \times q)$, then $vec\,(Q \otimes M) = (H_M \otimes I_p)\, vec\, M$, with $H = (I_n \otimes K_{qm})(vec\, A \otimes I_q)$.

The following proposition defines the likelihood function and finds the score vector and the information matrix for the simultaneous equations model proposed in Eq. (24).

Proposition 3 *Consider a random sample of size T from the process defined by the simultaneous equations model in Eq. (24) under the Assumptions 1-3. Let θ be an unknown vector of parameters and define θ_0 the true value of θ, such that $A_0 = A(\theta_0)$, $B_0 = B(\theta_0)$, $\Gamma_0 = \Gamma(\theta_0)$, and $\Lambda_0 = \Lambda(\theta_0)$. The log-likelihood function is*

$$\begin{aligned} l(\theta) = & -(Tg/2)\log 2\pi - \frac{1}{2}\sum_{t=1}^{T} \log\left(|I_g + AD_t|^2\right) + \frac{T}{2}\log\left(|B|^2\right) - \frac{T}{2}\log|\Lambda| \\ & - \frac{1}{2}\sum_{t=1}^{T} tr\left[(B^* y_t^* + \Gamma^* x_t^*)(B^* y_t^* + \Gamma^* x_t^*)' A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right]. \end{aligned} \tag{44}$$

The information matrix $\mathcal{F}_T(\theta_0)$, *defined as*

$$-E\left(\mathrm{d}^2 l(\theta_0)\right) = (\mathrm{d}\theta)' \mathcal{F}_T(\theta_0)\,\mathrm{d}\theta,$$

is given by

$$\mathcal{F}_T(\theta_0) = H' \begin{pmatrix} \mathcal{F}_{AA} & \mathcal{F}_{A\Psi} & \mathcal{F}_{A\Lambda} \\ \mathcal{F}_{\Psi A} & \mathcal{F}_{\Psi\Psi} & \mathcal{F}_{\Psi\Lambda} \\ \mathcal{F}_{\Lambda A} & \mathcal{F}_{\Lambda\Psi} & \mathcal{F}_{\Lambda\Lambda} \end{pmatrix} H \tag{45}$$

where

$$\mathcal{F}_{AA} = 2\left(I_{gs} \otimes \Lambda^{*-1}\right) N_{gs}\left(T^*\Lambda^* \otimes I_{gs}\right)$$

$$\mathcal{F}_{A\Psi} = 2\begin{bmatrix} \left(I_{gs} \otimes \Lambda^{*-1}\right) N_{gs}\left(T^*\Lambda^* \otimes I_{gs}\right) & 0 \end{bmatrix}$$

$$\mathcal{F}_{A\Lambda} = \left(T^* \otimes \Lambda^{*-1}\right)$$

$$\mathcal{F}_{\Psi\Psi} = \begin{pmatrix} \left(A^{*-1}B^* \otimes A^{*\prime}\right) & 0 \\ 0 & I_{gk} \end{pmatrix} \left[\left(C' \otimes C\right) K_{(g+k)s} + \left(Q^* \otimes A^{*-1\prime}\Lambda^{*-1}A^{*-1}\right)\right] \begin{pmatrix} \left(B^{*\prime}A^{*-1\prime} \otimes A^*\right) & 0 \\ 0 & I_{gk} \end{pmatrix}$$

$$\mathcal{F}_{\Psi\Lambda} = \begin{pmatrix} T^* \otimes \Lambda^{*-1} \\ 0 \end{pmatrix}$$

$$\mathcal{F}_{\Lambda\Lambda} = \left(\Lambda^* \otimes I_{gs}\right)\left(T^* \otimes \Lambda^{*-1}\right)$$

$$\mathcal{F}_{\Lambda A} = \mathcal{F}'_{A\Lambda} \qquad \mathcal{F}_{\Lambda\Psi} = \mathcal{F}'_{\Psi\Lambda} \qquad \mathcal{F}_{\Psi A} = \mathcal{F}'_{A\Psi}$$

and H *defined as in Eq. (43).*

The score vector, instead, is defined as (in row form)

$$f'(\theta) = \frac{\mathrm{d}l(\theta)}{\mathrm{d}vec\,\theta} = \begin{pmatrix} f_A(\theta), & f_\Psi(\theta), & f_\Lambda(\theta) \end{pmatrix} \tag{46}$$

where

$$\begin{aligned}
f_A(\theta) &= \sum_{t=1}^{T}\left(\left[vec\left(DA^{*-1}\Psi^* z_t^* z_t^{*\prime}\Psi^{*\prime}\Lambda^{*-1}A^{*-1\prime}\right)\right]' K_{gs}\left(H_A \otimes I_g\right) - \left[vec\left(\left(I_g + AD_t\right)' D_t\right)\right]'\right) \\
f_\Psi(\theta) &= -\sum_{t=1}^{T}\left[vec\left(A^{*-1\prime}\Lambda^{*-1}A^{*-1}\Psi^* z_t^* z_t^{*\prime}\right)\right]'\left(H_\Psi \otimes I_{(g+k)}\right) + T\left(vec\begin{bmatrix} B^{-1\prime} & 0 \end{bmatrix}\right)' \\
f_\Lambda(\theta) &= \frac{1}{2}\sum_{t=1}^{T}\left[vec\left(\Lambda^{*-1}A^{*-1}\Psi^* z_t^* z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}\right)\right]'\left(H_\Lambda \otimes I_g\right) D_g - \frac{T}{2}\left(vec\left(\Lambda\right)\right)' D_g
\end{aligned}$$

where $\Psi^* = \begin{pmatrix} I_s \otimes B & I_s \otimes \Gamma \end{pmatrix}$.

Proof. The proof of Proposition 3 is discussed in the Appendix A.6.

Using the results of Proposition 3 it becomes natural to implement the *score algorithm* to find FIML estimates of the parameters. In fact, once calculated the information matrix $\mathcal{F}_T(\theta)$ and the score vector $f(\theta)$, the score algorithm is based on the following updating formula (see for example Harvey, 1990, p. 134):

$$\theta_{n+1} = \theta_n + \lambda \left[\mathcal{F}_T(\theta_n)\right]^{-1} f(\theta_n). \tag{47}$$

where λ is a real number determining the step-size in the given direction. If the local identification does not require any restriction on the parameters, choosing accurately the starting values for θ, the recursive algorithm in Eq. (47) provides consistent estimates $\hat{\theta}$ for the true values θ_0. Once obtained, we can insert such consistent estimates into the information matrix and obtain the estimated asymptotic covariance matrix of $\hat{\theta}$:

$$\hat{\Sigma}_\theta = \mathcal{F}\left(\hat{\theta}\right)^{-1} = \left[p \lim_{T\to\infty} \frac{1}{T}\mathcal{F}_T\left(\hat{\theta}\right)\right]^{-1}. \tag{48}$$

Under the assumptions previously introduced, we obviously obtain

$$\hat{\theta} \xrightarrow{\mathcal{L}} N\left(\theta_0, \hat{\Sigma}_\theta\right) \tag{49}$$

allowing us to make inference on the parameters in the standard way.

In the more general case, in which we have both *a priori* knowledge on the parameters and different levels of volatility, and we use a combination of the two for obtaining the local identification, the FIML approach is a bit more complicated. In particular, introducing some restrictions on the parameters, both the score vector and the information matrix need to account for such restrictions. The solution, however, becomes straightforward if we consider the restrictions in the explicit form

$$\begin{pmatrix} vec\,A \\ vec\,B \\ vec\,\Gamma \\ vec\,\Lambda \end{pmatrix} = \begin{pmatrix} S_A & 0 & 0 & 0 \\ 0 & S_B & 0 & 0 \\ 0 & 0 & S_\Gamma & 0 \\ 0 & 0 & 0 & S_\Lambda \end{pmatrix} \begin{pmatrix} \gamma_A \\ \gamma_B \\ \gamma_\Gamma \\ \gamma_\Lambda \end{pmatrix} + \begin{pmatrix} s_A \\ s_B \\ s_\Gamma \\ s_\Lambda \end{pmatrix} \quad \Longrightarrow \quad \theta = S\gamma + s. \tag{50}$$

Using the standard chain of differentiation, thus, the score vector and the information matrix for the new set of parameters γ can be defined as

$$f(\gamma) = S' f(\theta) \tag{51}$$

$$\mathcal{F}_T(\gamma) = S' \mathcal{F}_T(\theta) S. \tag{52}$$

The score algorithm, at this stage, can be implemented for γ in order to obtain the FIML estimates $\hat{\gamma}$. Consistent estimates for θ and for the covariance matrix Σ_θ directly follows from the Cramer's linear transformation theorem by substituting the estimated $\hat{\gamma}$ in Eq. (50). The standard asymptotic result

$$\hat{\theta} \xrightarrow{\mathcal{L}} N\left(\theta_0, \frac{1}{T} S \mathcal{F}_T(\hat{\gamma}) S'\right) \tag{53}$$

thus applies.

IV The European Debt Crisis

This section presents an empirical analysis in which we highlight the potentiality of our methodology in relation with the existing literature on multivariate simultaneous equation systems. The analysis aims to shed light on the relationships between sovereign bond yields for some highly indebted EU countries. The data refer to 10-year bond maturity yield spreads between the so called 'PIGS' countries (Portugal, Ireland, Greece, and Spain) versus Germany, used as benchmark. We consider weekly observations over the period January 2005 - March 2011, and all the series come from Datastream.

The investigated period covers many interesting events characterizing the recent European and world-wide history, such as the global financial crises in 2007 and 2008, the 2008-2009 Spanish financial crisis, the great fear for the Greece default in 2010. All these events have led instabilities and tensions on the financial markets, which raised the problem of high public deficits and debt sustainability for the EU member states. Furthermore, given the strong interconnections between the markets, financial shocks in one country are likely propagated to other markets. Moreover, following Forbes and Rigobon (2002), Favero and Giavazzi (2002) and Bacchiocchi and Bevilacqua (2009), such mechanisms of propagation are different during tranquil or turbulent periods. It becomes fundamental, thus, to distinguish between a) the "natural" interconnections between financial markets, that we indicate as *interdependences*, from b) the propagation of financial crises hitting one or more countries, that constitutes the *pure contagion* phenomenon.

Such distinction, from an econometric point of view, leads to two serious problems of identification. The former, as already mentioned, consists in the detection of these two sources of interconnections; one which is always valid, the other which is valid during periods of crisis only. The latter, instead, is more familiar with the identification of simultaneous equation systems. If the trend of market A is important in explaining the trend of market B, it could be possible that also the contrary holds. This clearly conflicts with the traditional theory of identification in simultaneous equation systems[12]. The model we have developed in this paper, instead, starts from the idea of Rigobon (2003) to use heteroskedasticity to study and solve the problem of identification. On the other hand, as we will discuss here below, the presence of different clusters of volatility cannot be excluded, given the financial pressures characterizing the markets in the investigated sample.

The model in Eqs. (20)-(27), and all the related results for identification, estimation and inference, reveals to be perfect for distinguishing among interdependences and contagion, measured by the B and A matrices respectively, without imposing any kind of a priori restrictions on the parameters, that, in this specific case, could be hardly justifiable without any economic theoretical framework.

In Figure 1 we show the interest rates (left panel) and spreads (right panel) series for the sample period. From both graphs it emerges that the first years of the sample, at least up to the first signals of the global financial crisis such as the collapse of the U.S. housing bubble and the consequent rise in interest rates in the second half of 2007, the interest rates for all EU countries followed practically the same almost constant trend. Since that period, and up to the almost overall recognized end of the global financial crises in the late 2008, the EU interest rates started to rise and highlighted positive spreads with respect to their benchmark, the German Bund.

[12] The problem has been circumvented by Favero and Giavazzi (2002) by imposing restrictions on the dynamic part of the model, leading the contemporaneous relationships unrestricted

At the end of the global crisis, such differentials are in the order of 3 percent for Greece and Ireland, and 1.5 percent for Spain and Portugal. During the 2009 a moderate realignment appeared, but the situation became critical since the beginning of 2010, when the financial crisis turned into an even more dangerous debt crisis. Such debt crisis was mostly centered on events in Greece, where there was concern about the rising cost of financing government debt.

The global financial crisis, however, had contributed to transform other EU countries into fertile ground for financial, economic and social instabilities to occur. As an example, the Irish government officially announced it was in recession in September 2008, with a sharp rise in unemployment occurring in the following months. Ireland was the first state in the Eurozone to enter recession as declared by the Central Statistics Office. Although for different reasons, the situation was not so different for the Spanish economy, in which the collapse of the real estate boom, that included the bankruptcy of major companies, leads to a contraction of the GDP and a severe increase in the unemployment rate, that reached 13.9 percent in February 2009.

Such weak economic and financial conditions acted as a trigger for the rise in interest rates differentials realized in the markets since the beginning of 2010. Idiosyncratic policy interventions pursued by the National Governments, associated to wider rescue remedies proposed by the EU and IMF, seem to have only moderate and transitory effects on the situation of the financial markets, that up to now continue to register incredibly high spreads with respect to Germany, leading to serious problems of sustainability of the public debt for those countries.

The objectives of this sections are twofold: First, we want to estimate the simultaneous relationships between the interest rates of these countries measured by the B matrix in our general model specification. The complicated economic and financial interconnections between all these countries do not allow to follow any economic theoretical framework, imposing thus to estimate such matrix unrestrictedly. Second, the global and idiosyncratic financial crises suggest to estimate different mechanisms of propagation of shocks in periods of high volatility regimes as described by the A matrix in our formulation. Moreover, as for the case of the contemporaneous relationships, imposing restrictions on such further channel of propagation of shocks could be unrealistic, suggesting to estimate this matrix unrestrictedly too.

In Table 1 we report the different covariance and correlation matrices among the spreads over different horizons in the sample. Among the different sub periods described above, very different values for the variances and covariances among the spreads clearly emerge. Apart for the first period, characterized by stable interest rates and spreads, for all the other periods the correlations between the spreads are high and generally above 0.9.

IV.1 Interdependences, contagion, or something more

Suppose that the Data Generating Process is represented by the following model

$$B\begin{pmatrix} ir-ge_t \\ pt-ge_t \\ gr-ge_t \\ sp-ge_t \end{pmatrix} = c+\Phi(L)\begin{pmatrix} ir-ge_t \\ pt-ge_t \\ gr-ge_t \\ sp-ge_t \end{pmatrix} + \Theta(L)\begin{pmatrix} Vix_t \\ Baa-Aaa_t \end{pmatrix} + (I_4+AD_t)\begin{pmatrix} \varepsilon_t^{ir} \\ \varepsilon_t^{pt} \\ \varepsilon_t^{gr} \\ \varepsilon_t^{sp} \end{pmatrix} \tag{54}$$

where $ir - ge_t$, $pt - ge_t$, $gr - ge_t$, and $sp - ge_t$ are the interest rate spreads for Ireland, Portugal, Greece, and Spain, respectively. $Baa - Aaa_t$ is the spread between BAA and AAA corporate bonds and Vix_t measures market expectations of near term volatility conveyed by stock index option prices. Both indicators represent exogenous variables that could play a relevant role in the explanation of EU spreads. $\Phi(L)$ and $\Theta(L)$ are two matrix polynomials in the lag operator L, while c is a vector of constant terms. B and A are the two matrices of interest and represent the interdependences and contagion relationships, respectively. D_t is the (4×4) diagonal matrix indicating which variable is in a state of high volatility at time t, while the ε's are the idiosyncratic shocks, and are assumed to be uncorrelated and with constant standard deviation equal to 0.01[13].

The reduced form of the model is trivially obtained by premultiplying both sides of Eq. (54) by the invertible matrix B

$$\begin{pmatrix} ir - ge_t \\ pt - ge_t \\ gr - ge_t \\ sp - ge_t \end{pmatrix} = B^{-1}c + B^{-1}\Phi(L) \begin{pmatrix} ir - ge_t \\ pt - ge_t \\ gr - ge_t \\ sp - ge_t \end{pmatrix} + B^{-1}\Theta(L) \begin{pmatrix} Vix_t \\ Baa - Aaa_t \end{pmatrix} + u_t \quad (55)$$

where the reduced-form residuals u_t satisfy

$$B \begin{pmatrix} u_t^{ir} \\ u_t^{pt} \\ u_t^{gr} \\ u_t^{sp} \end{pmatrix} = (I_4 + AD_t) \begin{pmatrix} \varepsilon_t^{ir} \\ \varepsilon_t^{pt} \\ \varepsilon_t^{gr} \\ \varepsilon_t^{sp} \end{pmatrix} \quad (56)$$

that continue to share the same interdependence-contagion relationships as the original model in Eq. (54). Without any constraint on the parameters of the predetermined variables, maximizing the likelihood for Eq. (54) is equivalent to maximizing the concentrated likelihood in Eq. (56).

The reduced form in Eq. (55) can be seen as a standard VAR model that can be easily estimated by OLS. The residuals u_t depend on the structural matrices A and B, but also on the regimes of volatility described by D_t. In Figure 2 we report the residuals of the estimated reduced form, as well as the fitted and actual series of the spreads[14]. In Figure 3, instead, we report the absolute values for the reduced form residuals, which clearly highlight the presence of different clusters of volatility.

The recent events concerning the global financial crisis and the UE debt crisis provide a natural framework to define the regimes. As mentioned before, these events have been associated with large and persistent increases in volatility. Since June 2007 the four countries experienced global and idiosyncratic shocks that allow us to distinguish, country by country, tranquil from turbulent periods. Table 2 summarizes, for each country, the windows characterizing the high volatility regimes.

In Spain, the first and strong signals of instabilities appeared even before the 'official' start of the global financial crisis. The residential real estate bubble saw real estate prices

[13] A one standard deviation shock corresponds to one basic point change in the corresponding spread.

[14] The standard steps for the correct specification of the VAR models, based on the joint investigation of information criteria and specification tests on the residuals, suggest to include 5 lags for both the Θ and Φ polynomials.

rise 201% from 1985 to 2007. During the second half of 2007, when the real estate bubble burst, the crises immediately overcome the whole banking system that, althought credited as one of the most solid and best equipped among all Western economies to cope with the worldwide liquidity crisis, strongly relaxed his strict requirements from intending borrowers during the housing bubble, offering up to 50-year mortgages. The building market crash (included the bankruptcy of major companies), thus, rapidly resulted in a dramatic increase in unemployment and, during the third quarter of 2008, the national GDP contracted for the first time in 15 years and, in February 2009, it was confirmed that Spain had officially entered recession. Although Spain has a camparatively small debt/GDP ratio among advanced economies (60% only in 2010), tensions on both the banking and real sectors of the economy attracted speculators insofar as the Prime Minister Zapatero had to directly intervene to dismiss the rumors about a possible Spanish bail-out. Nevertheless, after few months, in May 2010, the Spanish Government had to announce extraordinary austerity measures to reduce the countrie's budget deficit, in order to dampen the financial markets and convince foreign investors about the solidity of the Spanish economy. As for many other EU countries, the Spanish real and financial economy will be under pressure until the end of the sample period used in our empirical analysis.

As for the Spanish case, the Irish crises was triggered by the 'terrible' mix of a real estate bubble from one side, and over-exposure of many large banks that financed the property market, from the other side. The situation for the banking sector became critical in September 2007, with the explosion of the global financial crises. The Irish Government, in April 2009, proposed a National Asset Management Agency (NAMA) to take over large loans from the banks, enabling them to return to normal liquidity to assist in the economic recovery. Such policy intervention had the effect to reassure the financial markets but, in September 2010, a new government help to refinance the banking sector had a negative impact on Irish sovereign bonds, leading the government to start negotiations with the ECB and IMF.

The recent historical events were substantially different for Greek and Portugal. Given the limited exposure of these countries with respect to international financial markets, they were only marginally touched by the global financial crisis. The Portuguese financial crisis was mainly an economic and political crisis and started during the first weeks of 2010. In fact, it is largely recognized that Portugal fell victim to successive waves of speculation by pressure from bond traders, rating agencies and speculators.

In Greece, instead, in only few days the situation became completely out of control. Facing the growing increase of the public debt, on April 23 the Greek government asks an initial loan of 45 billion euro to the EU and IMF to cover its financial needs for the remaining part of 2010. Only few days later Standard & Poor's decided to relegate the sovereign debt rating to 'junk'. This announcement did slump all EU and worldwide financial markets. The Greek government announced a series of austerity measures to reduce the country's deficit, which however lead to massive protests, riots and social unrest throughout Greece.

The combination of these events indicates 5 different regimes of volatility, as the result of the combination of Low-High volatility regimes for each country involved in the analysis. The windows of turbulent periods in each country are shown in Figure 4 (left panel), while the volatility regimes are shown in Figure 4 (right panel). Table 3 (column 1) summarizes the number of observations falling in each state of volatility.

The estimation of the D_t matrices allows to find the FIML estimation of the concentrated likelihood parameters A and B in Eq. (56). Before that, however, it is important to highlight that the $s = 5$ different levels of volatility, associated with the diagonal (with constant variance) structure of the structural residuals ε_t guarantee the sufficient rank and necessary order conditions for local identification as described in Proposition 2.

The estimated parameters, with associated standard errors, are reported in Table 4. Column (1) reports the estimates (and related standard errors) of A and B when no further restrictions have been imposed. Interestingly, this new procedure allows to consistently estimate the parameters of the contemporaneous relationships among the endogenous variables (the B matrix) without imposing exclusion restrictions as generally required in the traditional approach for the simultaneous equation models. Completely new in the literature, is the identification and estimation, even in this case without any kind of restrictions, of the elements in the A matrix, accounting for amplification and propagation of the structural shocks. Interestingly, a LR test for the presence of heteroskedastic residuals based on the coefficients of the A matrix, strongly rejects the null hypothesis $H_0 : A = 0$ with a p-value, obtained from a χ^2_{16} distribution, practically equal to zero.

In Table 4, column (2), we report the estimated coefficients when imposing restrictions on non-significant coefficients in the A matrix. The likelihood ratio test statistic, with associated p-values for such overidentifying restrictions, is reported in the last row of the table. The test suggests to not reject the null hypotheses for all standard significant levels. For a more comprehensible reading of the results we show, in the following equation, the estimated A and B parameters in a matrix notation, as reported in Table 4:

$$\begin{pmatrix} 1 & -0.232 & -0.184 & -1.060 \\ -0.132 & 1 & -0.364 & -0.239 \\ -0.191 & -0.408 & 1 & -0.739 \\ 0.267 & 0.390 & -0.849 & 1 \end{pmatrix} \begin{pmatrix} u_t^{ir} \\ u_t^{pt} \\ u_t^{gr} \\ u_t^{sp} \end{pmatrix} = \left[I_4 + \begin{pmatrix} 3.550 & 0 & 0 & 1.040 \\ 0 & 4.070 & 0 & 1.520 \\ 0 & -5.090 & 10.600 & -5.870 \\ 0 & 4.240 & -7.530 & 4.740 \end{pmatrix} D_t \right] \begin{pmatrix} \varepsilon_t^{ir} \\ \varepsilon_t^{pt} \\ \varepsilon_t^{gr} \\ \varepsilon_t^{sp} \end{pmatrix} .$$

The interpretation of the results is straightforward. The first equation, for example, can be read as an equation for the Irish spread, that depends on the other contemporaneous spreads and, during periods of instability, on possible combinations of domestic and internationals structural shocks. The same applies for the other equations.

The first comment highlights that there are bi-directional contemporaneous relations among all the spreads. In particular, the Irish spread is positively correlated with all other spreads, with a remarkable high coefficient related to the Spanish spread. The Portuguese spread, instead, is positively related, and with similar magnitude, to all the other markets.

The Greek spread positively depends on all other spreads and, as for Ireland, particularly with the Spanish one. The Spanish spread is contemporaneously and positively related with Greek one, while negatively with the Portuguese and Irish ones. Such structure would be clearly not identified in the standard simultaneous equation models. This problem is solved by dealing with the heteroskedasticity present in the data.

The second comment refers to the amplification and propagation of shocks, that we identify as the pure contagion effect. The estimated A matrix shows that all the structural shocks are significantly amplified when the market is in a situation of high instability. The highest value is represented by the Greek shock that is amplified by more than 10 times with respect to the tranquil periods. Positive and highly significant values have been found for all the diagonal elements of the A matrix, denoting that the Irish, Portuguese and Spanish shocks are also amplified in periods of high volatility.

The transmission or propagation of the structural shocks from one market to another are described by the off diagonal estimated elements of the A matrix. As an example, Spanish structural shocks are positively transmitted to the Irish and Portuguese spreads, while negatively to the Greek one. In other words, a positive structural shock that increases the Spanish spread by one basis point[15], also contribute to widen the Irish spread, by 1.040 basis points. Favero and Giavazzi (2002) interpret as "flight-to-quality" effects the negative propagation of shocks from one market to another. Given that we consider only highly indebted and critical countries, a similar interpretation would be questionable.

In general, the estimated results denote strong interconnections between the financial markets but, at the same time, strong evidence of non-linearities in the transmission of shocks when one or more countries exhibit periods of high instabilities. Such non-linearities can be interpreted as contagion in the case of positive coefficients of the off diagonal elements of A. In the case of negative coefficients, instead, a similar interpretation is more complicate. In order to help to comment on these empirical findings, in Table 5 (column 1 - 'Historical Events') we report the simultaneous relations among the spreads over the different regimes of volatility[16]. In particular, in the upper panel, we report the coefficients of the B matrix, related to the situation of low volatility regime for all variables. The following four panels, instead, show the simultaneous relations when some or all (last) spreads are in a state of high volatility.

As an example, when the Spanish spread faces a period of high turbulences (second panel - LLLH), the impact of the Spanish spread in the first two equations (ir-ge and pt-ge) increases, while in the Greek spread equation there is a drop in the coefficient, indicating a strong negative relation about the Greek spread and the Spanish one. In a similar manner we can interpret all the other coefficients of the table. Interestingly, apart few exceptions, when one or more countries experience periods of instability, there is a clear evidence of a change in the relationships among the markets. In some cases, this is in line with the literature of contagion. However, our results prove that the mechanism of propagation of the recent financial and debt crises is much more complicated than the findings in Forbes and Rigobon (2002), who attribute to what we (and they) have defined as *interdependences* the high levels of market comovements over the whole sample for the Asian and Mexican crises during the '90s.

IV.2 Robustness

In this section we report a robustness check. As already discussed in the previous sections, the way the crises are detected might have an influence on the estimation of the parameters. In the empirical analysis reported above, the D_t matrices have been determined by considering the historical events characterizing turbulences on local and global financial markets. In this sensitivity analysis, instead, we determine the high volatility periods as in Sack and Rigobon (2003), i.e. when the 9-weeks rolling variance of the residuals of the reduced form

[15]Corresponding to one standard deviation.

[16]These coefficients are simply obtained by calculating the matrix $\left[(I_g + AD_t)^{-1} B\right]$, for the different regimes of volatility described by D_t. The case of all spreads in a state of low volatility trivially corresponds to the B matrix, given that $D_t = 0$.

is more than one standard deviation above its average.

Table 3 (column 2 - Rolling Variances) reports a description of the volatility regimes in these sensitivity analyses. Differently with the previous case, the combination of the Low-High variances for each country leeds to $s = 13$ different volatility regimes. Moreover, the number of periods of high volatility is enormously reduced for all spreads, providing only 52 observations for estimating the amplification and propagation effects collected in the A matrix.

Table 4 (column 3 - Rolling Variances) shows the estimated coefficients and related standard errors obtained with this alternative description of the volatility regimes, that can be compared with the original results reported in the first two columns. The point estimates are rather similar to those obtained in the analysis discussed in the previous section. The main difference is that all coefficients are significantly different from zero. Concerning the coefficients on the main diagonal of the A matrix, the general result is that, apart from Spain, the amplification effects increase as the number of high volatility regimes reduces. This is clearly expected since we retain only the most pronounced cases of instabilities. About the interdependences, almost all coefficients are very similar to those obtained in the previous analysis. However, in this sensitivity analysis, the Greek spread is more related to the Irish and Portuguese spreads, while less to the Spanish one. In terms of the propagation of shocks, the consequences of such stronger interdependences are of higher negative coefficients in the A matrix that, in a certain sense, must mitigate the augmented amplification effects and the propagation of the shocks through the higher coefficients in the B matrix.

In Table 5 (column 2 - Rolling Variances) we report the contemporaneous relations among the spreads as for the previous case. The new results highlight very similar contemporaneous relations for the different regimes of volatility, with respect to those obtained in the previous analysis..

V Conclusion

In this paper we have presented a theoretical framework for identifying and estimating the parameters of heteroskedastic simultaneous equations models. In particular, we proposed a specification of the system that explicitly allows for different states of volatility. We suppose that the structural shocks hitting the economy present a constant covariance matrix, but in particular periods, such shocks might have amplified, generating thus clusters of higher volatility. The knowledge of such periods of high instability can represent a useful source of information for identifying the system, especially when a priori restrictions on the parameters of the model cannot be justified.

Under the assumption that the parameters remain constant over different states of volatility, we provide an order and a rank condition for solving the problem of local identification, both in the cases with and without restrictions on the parameters. The order condition, in particular, states that without any constraint, it is necessary to have at least three different levels of heteroskedasticity to reach local identification. The rank condition, instead, depends on the combination of high and low levels of volatility present in the data.

Concerning the estimation framework, we have developed a Full Information Maximum Likelihood approach that directly estimates the parameters of the structural form. We have also provided an analytical formulation for both the score function and the information matrix that allow us to implement an iterative procedure, the score algorithm, to maximize the likelihood. The classical inference, based on the ML estimators, can thus be applied.

Given the particular specification of the model, a fertile ground for possible empirical applications can be found in the literature of contagion, where, as highlighted in Forbes and Rigobon (2002), the distinction between interdependences (relations between endogenous variables) and pure contagion (transmission of structural shocks) is crucial. In this context, we have proposed an empirical analysis focusing on the transmission of financial shocks within four highly indebted EU countries; Ireland, Portugal, Greece and Spain. The particular specification of the model provides a useful tool for modeling the higher volatility of the interest rates on sovereign bonds observed during the recent turbulences on the financial markets all over the world. The results highlight that there are strong bilateral interdependences between the markets. Moreover, during periods of high volatility, the effect of a financial shock on one market is amplified and propagated to the other markets, regardless the standard interconnections among the markets, that remain unchanged. These findings are robust to different detections of the instability periods, and highlight that the mechanism of propagation of shocks is much more complicated than those usually detected by standard systems of equations or structural VAR modeling.

References

Bacchiocchi, E. and Bevilacqua, M. (2009), International crises, instability periods and contagion: the case of the ERM, *International Review of Economics* 56, 105-122.

Berndt, E.K., Hall, B.H., Hall, R.E. and Hausman, J.A. (1974), Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement* 3, 653-665.

Caporale, G.M., Cipollini, A. and Demetriades, P. (2005), Monetary policy and the exchange rate during the Asian crisis: Identification through heteroskedasticity, *Journal of International Money and Finance* 24, pages 39-53,.

Caporale, G.M., Cipollini, A. and Spagnolo, N. (2005), Testing for contagion: A conditional correlation analysis, *Journal of Empirical Finance* 12, 476-489.

Davidson, R. and MacKinnon, J.G. (1998), Graphical Methods for Investigating the Size and Power of Hypothesis Tests, *The Manchester School of Economics and Social Studies* 66, 1-26.

Dungey, M. and Martin, V.L. (2001), Contagion across financial markets: An empirical assessment, Australian National University mimeograph.

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

Forbes, K. and Rigobon, R. (2002), No contagion, only interdependence: measuring stock market co-movements, *Journal of Finance* 57, 2223-2261.

Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley.

Gourieroux, C., Monfort, A. and Trognon, A. (1984), Pseudo maximum likelihood methods: theory, *Econometrica* 52, 681-700.

Haavelmo, T. (1947), Methods of measuring the marginal propensity to consume, *Journal of the American Statistical Association* 42, 105-122.

Harvey, A.C. (1990), *The econometric analysis of time series*, LSE Handbooks of Economics, Philip Allan, London.

Hausman, J. (1983), Specification and Estimation of Simultaneous Models, in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics* 1, North-Holland, 391-448.

Hsiao, C. (1983), Identification, in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics* 1, North-Holland, 223-283.

Joreskog, K. G. and Sorbom D. (1984), *Advances in Factor Analysis and Structural Equation Models*, Rowman and Littlefield, Maryland.

King, M., Sentana, E. and Wadhwani, S. (1994), Volatility and links between national stock markets, *Econometrica* 62, 901-933.

Klein, R. and Vella, F. (2010), Estimating a Class of Triangular Simultaneous Equations Models without Exclusion Restrictions, *Journal of Econometrics* 154, 154-164.

Koopmans, T., Rubin, H. and Leipnik, R. (1950), Measuring the equation systems of dynamic economics, in Cowles Commission for Research in Economics (Ed.), *Statistical Inference in Dynamic Economic Models*, New York: John Wiley and Sons, 53-237.

Lanne, M. and Lütkepohl, H. (2008), Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* 40, 1131-1149.

Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010), Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics & Control* 34, 121-131.

Lewbel, A. (2011), Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models, *Journal of Business and Economic Statistics*, forthcoming.

Lütkepohl, H. (2005) *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin Heidelberg.

Magnus, J.R. (1988), *Linear structures*, Griffin, London.

Magnus, J.R. and Neudecker, H. (2007), *Matrix differential calculus with applications in statistics and econometrics*, John Wiley and Sons.

Normandin, M. and Phaneuf, B. (2004), Monetary policy shocks: Testing identification conditions under time-varying conditional volatility, *Journal of Monetary Economics* 51, 1217-43.

Prono, T. (2008), GARCH-Based Identification and Estimation of Triangular Systems, Federal Reserve Bank of Boston Working Paper QAU08-4.

Rigobon, R. (2002), The curse of non-investment grade countries, *Journal of Development Economics* 69, 423-449.

Rigobon, R. (2003), Identification through heteroskedasticity, *The Review of Economics and Statistics* 85, 777-792.

Rigobon, R. and Sack B. (2003), Measuring the reaction of monetary policy to the stock market, *Quarterly Journal of Economics* 118, 639-669.

Rigobon, R. and Sack B. (2004), The impact of monetary policy on asset prices, *Journal of Monetary Economics* 51, 1553-1575.

Rothenberg, T.J. (1971), Identification in parametric models, *Econometrica* 39, 577-591.

Sentana, E. and Fiorentini, G. (2001), Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models, *Journal of Econometrics* 102, 143-164.

White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1-25.

White, H. (1983), Corrigendum, *Econometrica* 51, 513.

Wright, P.G. (1928), *The tariff on animal and vegetable oils*, New York, Macmillan.

A Appendix: Proofs

A.1 Proof of Proposition 1

Equations (7)-(11) form a system of non-linear equations (because of Eq. (7) and Eq. (8)) in A, B and $v(\Lambda)$, whose solutions provide a fundamental indication for the identifiability of the parameters. Differentiating Eqs. (7)-(11) gives

$$\begin{aligned}
-C_1^{-1}\mathrm{d}AD_1C_1^{-1}B\Omega_1B'C_1^{-1\prime}+C_1^{-1}\mathrm{d}B\Omega_1B'C_1^{-1\prime}+C_1^{-1}B\Omega_1\mathrm{d}B'C_1^{-1}&\\
-C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\mathrm{d}A'C_1^{-1\prime}-\mathrm{d}\Lambda&=0\\
-C_2^{-1}\mathrm{d}AD_2C_2^{-1}B\Omega_2B'C_2^{-1\prime}+C_2^{-1}\mathrm{d}B\Omega_2B'C_2^{-1\prime}+C_2^{-1}B\Omega_2\mathrm{d}B'C_2^{-1}&\\
-C_2^{-1}B\Omega_2B'C_2^{-1\prime}D_2\mathrm{d}A'C_2^{-1\prime}-\mathrm{d}\Lambda&=0\\
R_A vec\,\mathrm{d}A&=0\\
R_B vec\,\mathrm{d}B&=0\\
R_\Lambda \mathrm{d}v(\Lambda)&=0
\end{aligned}$$

Using the property $vec\,(ABC)=(C'\otimes A)\,vec\,B$, the system of equations can be written

$$\begin{aligned}
-\left[C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\otimes C_1^{-1}\right]\mathrm{d}vec\,A+\left[C_1^{-1}B\Omega_1\otimes C_1^{-1}\right]\mathrm{d}vec\,B+\left[C_1^{-1}\otimes C_1^{-1}B\Omega_1\right]K_g\mathrm{d}vec\,B&\\
-\left[C_1^{-1}\otimes C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\right]K_g\mathrm{d}vec\,A-\mathrm{d}vec\,\Lambda&=0\\
-\left[C_2^{-1}B\Omega_2B'C_2^{-1\prime}D_2\otimes C_2^{-1}\right]\mathrm{d}vec\,A+\left[C_2^{-1}B\Omega_2\otimes C_2^{-1}\right]\mathrm{d}vec\,B+\left[C_2^{-1}\otimes C_2^{-1}B\Omega_2\right]K_g\mathrm{d}vec\,B&\\
-\left[C_2^{-1}\otimes C_2^{-1}B\Omega_2B'C_2^{-1\prime}D_2\right]K_g\mathrm{d}vec\,A-\mathrm{d}vec\,\Lambda&=0\\
R_A vec\,\mathrm{d}A&=0\\
R_B vec\,\mathrm{d}B&=0\\
R_\Lambda \mathrm{d}v(\Lambda)&=0
\end{aligned}$$

where K_{mn} is the *commutation matrix* previously defined. Using the property of the commutation matrix and duplication matrix, we rewrite the system as

$$\begin{aligned}
-\left(I_{g^2}+K_g\right)\left[C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\otimes C_1^{-1}\right]\mathrm{d}vec\,A+\left(I_{g^2}+K_g\right)\left[C_1^{-1}B\Omega_1\otimes C_1^{-1}\right]\mathrm{d}vec\,B&\\
-D_g\mathrm{d}v(\Lambda)&=0\\
-\left(I_{g^2}+K_g\right)\left[C_2^{-1}B\Omega_2B'C_2^{-1\prime}D_2\otimes C_2^{-1}\right]\mathrm{d}vec\,A+\left(I_{g^2}+K_g\right)\left[C_2^{-1}B\Omega_2\otimes C_2^{-1}\right]\mathrm{d}vec\,B&\\
-D_g\mathrm{d}v(\Lambda)&=0\\
R_A vec\,\mathrm{d}A&=0\\
R_B vec\,\mathrm{d}B&=0\\
R_\Lambda \mathrm{d}v(\Lambda)&=0
\end{aligned}$$

The Jacobian matrix, thus, can be written as

$$J=\begin{pmatrix}
-2N_g\left[C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\otimes C_1^{-1}\right] & 2N_g\left[C_1^{-1}B\Omega_1\otimes C_1^{-1}\right] & -D_g\\
-2N_g\left[C_2^{-1}B\Omega_2B'C_2^{-1\prime}D_2\otimes C_2^{-1}\right] & 2N_g\left[C_2^{-1}B\Omega_2\otimes C_2^{-1}\right] & -D_g\\
R_A & 0 & 0\\
0 & R_B & 0\\
0 & 0 & R_\Lambda
\end{pmatrix}. \tag{57}$$

with $N_g = \frac{1}{2}\left(I_{g^2} - K_g\right)$, a $\left(g^2 \times g^2\right)$ matrix with reduced rank $g(g+1)/2$. We note that the Jacobian matrix only depends on A and B, and not on Λ (since the non-linearity in Eqs. (7)-(9) are on A and B). Following Rothenberg (1971), a necessary and sufficient condition for (A_0, B_0, Λ_0) to be locally identifiable is that J, evaluated at Λ_0 has full column rank. A necessary only condition, however, is that the number of rows needs to be, at least, as large as the number of columns. In the present case, the sub matrix composed by the first two rows in Eq. (57) is of dimension $\left(2g^2 \times \left[2g^2 + \frac{1}{2}g(g+1)\right]\right)$, even if, as can be easily seen from Eqs. (7)-(8), the number of independent rows is equal to $g(g+1)/2 + g(g+1)/2$. In other words, the necessary condition for identification indicates that $g(g+1) + q_A + q_B + q_\Lambda \geq 2g^2 + g(g+1)/2$, that, after some simple algebra, leads to $q_A + q_B + q_\Lambda \geq g^2 + g(g-1)/2$, which is the necessary condition reported in Proposition 1.

Concerning the rank condition, using the definitions of the D_1^*, E_1^*, and E_2^* as described in Eqs. (16), and substituting $D_1 = I_g$ and $D_2 = 0_g$, the Jacobian matrix in Eq. (57) becomes

$$J = \begin{pmatrix} -2N_g D_1^* & 2N_g E_1^* & -D_g \\ 0 & 2N_g E_2^* & -D_g \\ R_A & 0 & 0 \\ 0 & R_B & 0 \\ 0 & 0 & R_\Lambda \end{pmatrix}. \tag{58}$$

The condition of full column rank of this matrix is equivalent to the condition that the following homogeneous system of $\left(2g^2 + q\right)$ equations in $2g^2 + g(g+1)/2$ unknowns

$$\begin{bmatrix} -2N_g D_1^* & 2N_g E_1^* & -D_g \\ 0 & 2N_g^* & -D_g \\ R_A & 0 & 0 \\ 0 & R_B & 0 \\ 0 & 0 & R_\Lambda \end{bmatrix} x = [0] \tag{59}$$

has only one admissible solution $x = [0]$. The system can be split into five systems of equations that are connected because they share the same unknowns

$$\begin{cases} -2N_g D_1^* x_1 + 2N_g E_1^* x_2 - D_g x_3 &= 0 \\ 2N_g E_2^* x_2 - D_g x_3 &= 0 \\ R_A x_1 &= 0 \\ R_B x_2 &= 0 \\ R_\Lambda x_3 &= 0 \end{cases} \tag{60}$$

Using the explicit notation for the restrictions as described in Eqs. (12)-(14), allows to solve the last three equations as

$$\begin{cases} x_1 &= S_A q_1 \\ x_2 &= S_B q_2 \\ x_3 &= S_\Lambda q_3 \end{cases} \tag{61}$$

for q_1, q_2, q_3 vectors of appropriate dimensions. Substituting in the first two equations leads to

$$\begin{cases} -2N_g D_1^* S_A q_1 + 2N_g E_1^* S_B q_2 - D_g S_\Lambda q_3 &= 0 \\ 2N_g E_2^* S_B q_2 - D_g S_\Lambda q_3 &= 0 \end{cases} \tag{62}$$

which admits the unique solution $(q_1', q_2', q_3')' = [0]$ if the matrix

$$\begin{pmatrix} -2N_g D_1^* S_A & 2N_g E_1^* S_B & -D_g S_\Lambda \\ 0 & 2N_g E_2^* S_B & -D_g S_\Lambda \end{pmatrix} \tag{63}$$

has full column rank.■

A.2 Proof of Corollary 1

When the covariance matrix of the structural shocks Λ is completely constrained, i.e. $R_\Lambda = I_{g(g+1)/2}$, the attention will be paid to the first two block columns of the Jacobian in Eq. (58), i.e.

$$J = \begin{pmatrix} -2N_g D_1^* & 2N_g E_1^* \\ 0 & 2N_g E_2^* \\ R_A & 0 \\ 0 & R_B \end{pmatrix} \tag{64}$$

whose rank does not change if we post-multiply by a non-singular matrix as follows

$$J^* = \begin{pmatrix} -2N_g D_1^* & 2N_g E_1^* \\ 0 & 2N_g E_2^* \\ R_A & 0 \\ 0 & R_B \end{pmatrix} \begin{pmatrix} -D_1^{*-1} & 0 \\ 0 & E_2^{*-1} \end{pmatrix} = \begin{pmatrix} 2N_g & 2N_g E_1^* E_2^{*-1} \\ 0 & 2N_g \\ -R_A D_1^{*-1} & 0 \\ 0 & R_B E_2^{*-1} \end{pmatrix}. \tag{65}$$

The condition of full column rank of this matrix is equivalent to the condition that the following homogeneous system of $\left(2g^2 + q_A + q_B\right)$ equations in $2g^2$ unknowns

$$\begin{bmatrix} N_g & 2N_g E_1^* E_2^{*-1} \\ 0 & 2N_g \\ -R_A D_1^{*-1} & 0 \\ 0 & R_B E_2^{*-1} \end{bmatrix} x = [0] \tag{66}$$

has only one admissible solution $x = [0]$. The system can be split into four systems of equations that are connected because they share the same unknowns

$$\begin{cases} N_g x_1 + N_g E_1^* E_2^{*-1} x_2 &= 0 \\ N_g x_2 &= 0 \\ -R_A D_1^{*-1} x_1 &= 0 \\ R_B E_2^{*-1} x_2 &= 0. \end{cases} \tag{67}$$

Following Magnus (1988), the second matrix equation can be solved as

$$x_2 = \widetilde{D}_g v_2$$

where $\widetilde{D}_g$, as defined above, is a $g^2 \times g(g-1)/2$ full column rank matrix and v_2 is a $g(g-1)/2$ vector of free elements. The third equation , instead, using the restrictions in the A matrix in its explicit notation S_A, can be solved as $x_1 = D_1^* S_A v_1$ for any $\left(g^2 - q_A \times 1\right)$ vector v_1. Substituting the second and third matrix equations into the first, the system becomes

$$\begin{cases} N_g D_1^* S_A v_1 + N_g E_1^* E_2^{*-1} \widetilde{D}_g v_2 &= 0 \\ \widetilde{D}_g v_2 &= x_2 \\ D_1^* S_A v_1 &= x_1 \\ R_B E_2^{*-1} x_2 &= 0. \end{cases} \tag{68}$$

The solution of the system reduces to verify whether the $\left(g^2 + q_B\right) \times g^2 - q_A + g\left(g-1\right)/2$ matrix related to the first and forth equations, i.e.

$$\begin{pmatrix} N_g D_1^* S_A & N_g E_1^* E_2^{*-1} \widetilde{D}_g \\ 0 & R_B E_2^{*-1} \widetilde{D}_g \end{pmatrix} \tag{69}$$

has full column rank.

This matrix will have full column rank only if the number of rows is at least equal to the number of columns. More precisely, given that the N_g matrix, by construction, when premultiplies another matrix generates only $g\left(g+1\right)/2$ distinct rows (instead of the original g^2), the upper part of the matrix in Eq. (69), i.e. $\begin{pmatrix} N_g D_1^* S_A & N_g E_1^* E_2^{*-1} \widetilde{D}_g \end{pmatrix}$, will be made of only $g\left(g+1\right)/2$ distinct rows.

The necessary condition, thus, requires that

$$g\left(g+1\right)/2 + q_B \geq g^2 - q_A + g\left(g-1\right)/2 \quad \Longrightarrow \quad q_A + q_B \geq g\left(g-1\right) \tag{70}$$

which proves the result.■

A.3 Proof of Proposition 2

Following Rotenberg (1971), the identifiability of the parameters of the structural form depends on the uniqueness of solutions of the system linking the parameters of the structural and reduced forms. This system of matrix equations can be written as

$$\left(I_s \otimes B\right)\left(I_s \otimes \Pi\right) + \left(I_s \otimes \Gamma\right) = 0 \tag{71}$$

$$A^{*-1}\left(I_s \otimes B\right)\left(I_s \otimes \Omega\right)\left(I_s \otimes B\right)' A^{*-1\prime} - \left(I_s \otimes \Lambda\right) = 0 \tag{72}$$

$$R_A vec\,\mathrm{d}A = 0 \tag{73}$$

$$R_B vec\,\mathrm{d}B = 0 \tag{74}$$

$$R_\Gamma vec\,\mathrm{d}\Gamma = 0 \tag{75}$$

$$R_\Lambda \mathrm{d}v\left(\Lambda\right) = 0. \tag{76}$$

where Eqs. (73)-(76) indicate possible restrictions on the parameters (A, B, Γ, Λ). The first differential is

$$\begin{aligned}
(I_s \otimes \mathrm{d}B)(I_s \otimes \Pi) + (I_s \otimes \mathrm{d}\Gamma) &= 0 \\
-A^{*-1}(I_s \otimes \mathrm{d}A) DA^{*-1}(I_s \otimes B)(I_s \otimes \Omega)(I_s \otimes B)' A^{*-1\prime} + & \\
A^{*-1}(I_s \otimes \mathrm{d}B)(I_s \otimes \Omega)(I_s \otimes B)' A^{*-1\prime} + & \\
A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)\left(I_s \otimes \mathrm{d}B'\right) A^{*-1\prime} + & \\
-A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)\left(I_s \otimes \mathrm{d}B'\right) A^{*-1\prime} D (I_s \otimes \mathrm{d}A) A^{*-1\prime} - (I_s \otimes \mathrm{d}\Lambda) &= 0 \\
R_A vec\, \mathrm{d}A &= 0 \\
R_B vec\, \mathrm{d}B &= 0 \\
R_\Gamma vec\, \mathrm{d}\Gamma &= 0 \\
R_\Lambda \mathrm{d}v(\Lambda) &= 0.
\end{aligned}$$

Simple algebra allows us to rewrite the system as

$$\begin{aligned}
\left[\left(I_s \otimes \Pi'\right) \otimes I_{gs}\right] vec(I_s \otimes \mathrm{d}B) + vec(I_s \otimes \mathrm{d}\Gamma) &= 0 \\
-\left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)(I_s \otimes B)' A^{*-1} \otimes A^{*-1}\right] vec(I_s \otimes \mathrm{d}A) + & \\
\left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega) \otimes A^{*-1}\right] vec(I_s \otimes \mathrm{d}B) + & \\
\left[A^{*-1} \otimes A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)\right] vec\left(I_s \otimes \mathrm{d}B'\right) + & \\
\left[A^{*-1} \otimes A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)\left(I_s \otimes B'\right) A^{*-1\prime} D\right] vec\left(I_s \otimes \mathrm{d}A'\right) - vec(I_s \otimes \mathrm{d}\Lambda) &= 0 \\
R_A vec\, \mathrm{d}A &= 0 \\
R_B vec\, \mathrm{d}B &= 0 \\
R_\Gamma vec\, \mathrm{d}\Gamma &= 0 \\
R_\Lambda \mathrm{d}v(\Lambda) &= 0.
\end{aligned}$$

Using the properties of the Kronecker product, the system becomes

$$\begin{aligned}
\left[\left(I_s \otimes \Pi'\right) \otimes I_{gs}\right](H_B \otimes I_g) vec\, \mathrm{d}B + (H_\Gamma \otimes I_g) vec\, \mathrm{d}\Gamma &= 0 \quad (77) \\
-\left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)(I_s \otimes B)' A^{*-1} \otimes A^{*-1}\right](H_A \otimes I_g) vec\, \mathrm{d}A + & \\
\left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega) \otimes A^{*-1}\right](H_B \otimes I_g) vec\, \mathrm{d}B + & \\
\left[A^{*-1} \otimes A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)\right] K_{gs}(H_B \otimes I_g) vec\, \mathrm{d}B + & \\
-\left[A^{*-1} \otimes A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)\left(I_s \otimes B'\right) A^{*-1\prime} D\right] K_{gs}(H_A \otimes I_g) vec\, \mathrm{d}A + & \\
-(H_\Lambda \otimes I_g) vec\, \mathrm{d}\Lambda &= 0 \quad (78) \\
R_A vec\, \mathrm{d}A &= 0 \quad (79) \\
R_B vec\, \mathrm{d}B &= 0 \quad (80) \\
R_\Gamma vec\, \mathrm{d}\Gamma &= 0 \quad (81) \\
R_\Lambda \mathrm{d}v(\Lambda) &= 0 \quad (82)
\end{aligned}$$

where, following Magnus and Neudecker (2007) p. 56, the matrix H is defined such that, given two matrices $A(m \times n)$ and $B(p \times q)$ then $vec(A \otimes B) = (H \otimes I_p) vecB$, with $H = (I_n \otimes K_{qm})(vec\, A \otimes I_q)$. Using the properties of the commutation matrix K_{gs}, the matrix equation in (78) can be simplified as:

$$\begin{aligned} -2N_{gs}\left[A^{*-1}\left(I_s\otimes B\right)\left(I_s\otimes\Omega\right)\left(I_s\otimes B\right)'A^{*-1}\otimes A^{*-1}\right]\left(H_A\otimes I_g\right)vec\,\mathrm{d}A+ & \\ 2N_{gs}\left[A^{*-1}\left(I_s\otimes B\right)\left(I_s\otimes\Omega\right)\otimes A^{*-1}\right]\left(H_B\otimes I_g\right)vec\,\mathrm{d}B-\left(H_\Lambda\otimes I_g\right)D_g v\,\mathrm{d}\Lambda &= 0 \end{aligned} \quad (83)$$

with $N_{gs}=1/2\left(I_{gs}+K_{gs}\right)$, as before. From Eqs. (77)-(82) we obtain the Jacobian matrix

$$J=\begin{pmatrix} 0 & \left[\left(I_s\otimes\Pi'\right)\otimes I_{gs}\right]\left(H_B\otimes I_g\right) & \left(H_\Gamma\otimes I_g\right) & 0 \\ -2N_{gs}J_{21} & 2N_{gs}J_{22} & 0 & -J_{23}D_g \\ R_A & 0 & 0 & 0 \\ 0 & R_B & 0 & 0 \\ 0 & 0 & R_\Gamma & 0 \\ 0 & 0 & 0 & R_\Lambda \end{pmatrix}. \quad (84)$$

However, the non-zero blocks in the first row do not change for the different regimes of volatility. The Jacobian matrix thus, simplifies to

$$J=\begin{pmatrix} 0 & \Pi'\otimes I_g & I_{gk} & 0 \\ -2N_{gs}J_{21} & 2N_{gs}J_{22} & 0 & -J_{23}D_g \\ R_A & 0 & 0 & 0 \\ 0 & R_B & 0 & 0 \\ 0 & 0 & R_\Gamma & 0 \\ 0 & 0 & 0 & R_\Lambda \end{pmatrix} \quad (85)$$

where

$$J_{21}=\left[A^{*-1}\left(I_s\otimes B\right)\left(I_s\otimes\Omega\right)\left(I_s\otimes B\right)'A^{*-1\prime}D\otimes A^{*-1}\right]\left(H_A\otimes I_g\right) \quad (86)$$

$$J_{22}=\left[A^{*-1}\left(I_s\otimes B\right)\left(I_s\otimes\Omega\right)\otimes A^{*-1}\right]\left(H_B\otimes I_g\right). \quad (87)$$

$$J_{23}=\left(H_\Lambda\otimes I_g\right). \quad (88)$$

The necessary and sufficient condition depends thus on the kind of combinations of high volatility states as highlighted in the D matrix. The necessary and sufficient condition in Proposition 2 is simply obtained by referring to the solution of the system of equations associated to the Jacobian in Eq. (85), where the restrictions on the parameters are substituted with their explicit form. This completes the proof. ■

A.4 Proof of Corollary 2

If we do not want to include restrictions as in Eq. (73), we can concentrate on the following partitioned matrix

$$J=\begin{pmatrix} 0 & \Pi'\otimes I_g & I_{gk} & 0 \\ -2N_{gs}J_{21} & 2N_{gs}J_{22} & 0 & -J_{23}D_g \end{pmatrix} \quad (89)$$

and verify for the full column rank condition. However, when no restrictions are imposed on the parameters, the analysis of identification can be carried out on the concentrated model with respect to Γ, as in the standard SVAR literature on identification (when unrestricted, the parameters related to the predetermined variables, Γ, do not influence the identification of the structural form). The Jacobian for the concentrated model becomes

$$J^* = \begin{pmatrix} 0 & \Pi' \otimes I_g & 0 \\ -2N_{gs}J_{21} & 2N_{gs}J_{22} & -J_{23}D_g \end{pmatrix} \tag{90}$$

where J_{21} and J_{22} are defined as before. The related system of equations is

$$\begin{aligned} \left(\Pi' \otimes I_g\right) x_2 &= 0 && (91) \\ -2N_{gs}J_{21}x_1 + 2N_{gs}J_{22}x_2 - J_{23}D_g x_3 &= 0. && (92) \end{aligned}$$

However, the first matrix equation is an homogeneous equation that admits solutions as

$$x_2 = \left[I_{g^2} - \left(\Pi' \otimes I_g\right)^+ \left(\Pi' \otimes I_g\right)\right] q_2 \tag{93}$$

for a general vector q_2 of appropriate dimension. Substituting the first into the second equation, it becomes

$$\begin{aligned} \left[I_{g^2} - \left(\Pi' \otimes I_g\right)^+ \left(\Pi' \otimes I_g\right)\right] q_2 &= x_2 && (94) \\ -2N_{gs}J_{21}x_1 + 2N_{gs}J_{22}\left[I_{g^2} - \left(\Pi' \otimes I_g\right)^+ \left(\Pi' \otimes I_g\right)\right] q_2 - J_{23}D_g x_3 &= 0. && (95) \end{aligned}$$

which admits the null vector as the unique possible solution if and only if the matrix

$$\begin{pmatrix} -2N_{gs}J_{21} & 2N_{gs}J_{22}\left[I_{g^2} - (\Pi' \otimes I_g)^+ (\Pi' \otimes I_g)\right] & -J_{23}D_g \end{pmatrix} \tag{96}$$

has full column rank. This condition, of course, can be easily verified numerically and represents a necessary and sufficient condition for the identifiability of the parameters of the structural form.

A sufficient condition for $(A_0, B_0, \Gamma_0, \Lambda_0)$ to be locally identifiable is that the J matrix in Eq. (89), that depends only on A and B, when evaluated at A_0 and B_0 has full column rank. A necessary condition, thus, is clearly that $rows\,(J) \geq cols\,(J)$. Considering that the set of g^2s^2 equations obtained from Eq. (72) refers to symmetric matrices, the number of effectively distinct rows in the necessary and sufficient condition in Eq. (32) is equal to $gk + gs\,(gs+1)\,/2$. The number of columns instead, is equal to the number of parameters to be estimated under the linear restrictions described in Proposition 2, i.e. $\left(g^2 - q_A\right) + \left(g^2 - q_B\right) + (gk - q_\Gamma) + [g\,(g+1)\,/2 - q_\Lambda]$.

The necessary condition, thus, states that

$$gs\,(gs+1)\,/2 + q \geq 2g^2 + g\,(g+1)\,/2 \tag{97}$$

where $q = q_A + q_B + q_\Gamma + q_\Lambda$ is the total number of restrictions on the unknown parameters A, B, Γ, and Λ. ■

A.5 Proof of Corollary 3

Including different levels of volatility is a way to increase the number of rows in the Jacobian matrix in Eq. (89). The *order condition* for systems with different levels of volatility concerns the minimum number of states in order to have, at least, as many rows as columns in the J matrix.

When there are no restrictions on the parameters, the necessary condition corresponds to Eq. (97) with $q = 0$. After some simple algebra it can be shown that for $s \geq 3$ the inequality is always verified, indicating that a minimum of three states of volatility is necessary for making the parameters identifiable. ■

A.6 Proof of Proposition 3

The log-likelihood function in Eq. (44) can also be written as

$$\begin{aligned} l(\theta) = \quad & const \quad -\frac{1}{2}\sum_{t=1}^{T}\left|I_g + AD_t\right|^2 + \frac{T}{2}\log\left|B\right|^2 - \frac{T}{2}\left|\Lambda\right| \\ - \quad & \frac{1}{2}\sum_{t=1}^{T} tr\left[\Psi^* z_t^* z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}\left(I_s \otimes \Lambda^{-1}\right)A^{*-1}\right] \end{aligned} \tag{98}$$

where

$$z_t^* = \begin{pmatrix} y_t^* \\ x_t^* \end{pmatrix} \quad , \quad \Psi^* = \begin{pmatrix} I_s \otimes B & I_s \otimes \Gamma \end{pmatrix} . \tag{99}$$

The first differential is

$$\begin{aligned} \mathrm{d}l(\theta) = \quad - \quad & \sum_{t=1}^{T} tr\left(\left(I_g + AD_t\right)^{-1}\mathrm{d}AD_t\right) + T\,tr\left(B^{-1}\mathrm{d}B\right) - \frac{T}{2}tr\left(\Lambda^{-1}\mathrm{d}\Lambda\right) + \\ + \quad & \sum_{t=1}^{T} tr\left[\left(\Lambda^{*-1}A^{*-1}\Psi^* z_t^* z_t^{*\prime}\Psi^{*\prime}\right)A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A'\right)A^{*-1\prime}\right] + \\ + \quad & \frac{1}{2}\sum_{t=1}^{T} tr\left(A^{*-1}\Psi^* z_t^* z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}\right) + \\ - \quad & \sum_{t=1}^{T} tr\left(z_t^* z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}A^{*-1}\mathrm{d}\Psi^*\right) \end{aligned} \tag{100}$$

and the second differential is

$$
\begin{aligned}
\mathrm{d}^2 l(\theta) = \quad & + \sum_{t=1}^{T} tr\left((I_g + AD_t)^{-1} \mathrm{d}AD_t \left(I_g + AD_t\right)^{-1} \mathrm{d}AD_t \right) + \\
& - T tr\left(B^{-1}\mathrm{d}BB^{-1}\mathrm{d}B\right) + \frac{T}{2} tr\left(\Lambda^{-1} d\Lambda\Lambda^{-1}\mathrm{d}\Lambda\right) + \\
& - \sum_{t=1}^{T} tr\left(\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}A^{*-1}\Psi^{*}z_t^{*}z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A^{\prime}\right)A^{*-1\prime}\right) + \\
& - \sum_{t=1}^{T} tr\left(\Lambda^{*-1}A^{*-1}\left(I_s \otimes \mathrm{d}A\right)DA^{*-1}\Psi^{*}z_t^{*}z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A^{\prime}\right)A^{*-1\prime}\right) + \\
& + \sum_{t=1}^{T} tr\left(\Lambda^{*-1}A^{*-1}\mathrm{d}\Psi^{*}z_t^{*}z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A^{\prime}\right)A^{*-1\prime}\right) + \\
& + \sum_{t=1}^{T} tr\left(\Lambda^{*-1}A^{*-1}\Psi^{*}z_t^{*}z_t^{*\prime}\mathrm{d}\Psi^{*\prime}A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A^{\prime}\right)A^{*-1\prime}\right) + \\
& - 2\sum_{t=1}^{T} tr\left(\Lambda^{*-1}A^{*-1}\Psi^{*}z_t^{*}z_t^{*\prime}\mathrm{d}\Psi^{*\prime}A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A^{\prime}\right)A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A^{\prime}\right)A^{*-1\prime}\right) + \\
& - \frac{1}{2}\sum_{t=1}^{T} tr\left(A^{*-1}\left(I_s \otimes \mathrm{d}A\right)DA^{*-1}\Psi^{*}z_t^{*}z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}\right) + \\
& + \frac{1}{2}\sum_{t=1}^{T} tr\left(A^{*-1}\mathrm{d}\Psi^{*}z_t^{*}z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}\right) + \\
& + \frac{1}{2}\sum_{t=1}^{T} tr\left(A^{*-1}\Psi^{*}z_t^{*}z_t^{*\prime}\mathrm{d}\Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}\right) + \\
& - \frac{1}{2}\sum_{t=1}^{T} tr\left(A^{*-1}\Psi^{*}z_t^{*}z_t^{*\prime}\mathrm{d}\Psi^{*\prime}A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A^{\prime}\right)A^{*-1\prime}\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}\right) + \\
& - \sum_{t=1}^{T} tr\left(A^{*-1}\Psi^{*}z_t^{*}z_t^{*\prime}Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}\right) + \\
& - \sum_{t=1}^{T} tr\left(\mathrm{d}\Psi^{*}z_t^{*}z_t^{*\prime}\mathrm{d}\Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}A^{*-1}\right) + \\
& + \sum_{t=1}^{T} tr\left(\mathrm{d}\Psi^{*}z_t^{*}z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}D\left(I_s \otimes \mathrm{d}A^{\prime}\right)A^{*-1\prime}\Lambda^{*-1}A^{*-1}\right) + \\
& + \sum_{t=1}^{T} tr\left(\mathrm{d}\Psi^{*}z_t^{*}z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}\left(I_s \otimes \mathrm{d}\Lambda\right)\Lambda^{*-1}A^{*-1}\right) + \\
& + \sum_{t=1}^{T} tr\left(\mathrm{d}\Psi^{*}z_t^{*}z_t^{*\prime}\Psi^{*\prime}A^{*-1\prime}\Lambda^{*-1}A^{*-1}\left(I_s \otimes \mathrm{d}A\right)DA^{*-1}\right) +
\end{aligned}
\tag{101}
$$

After some algebra, it becomes

$$\begin{aligned}
\mathrm{d}^2 l(\theta) &= vec\left(I_s \otimes \mathrm{d}A\right)' \left[DA^{*-1} \otimes A^{*-1\prime} T^* D\right] K_{gs} vec\left(I_s \otimes \mathrm{d}A\right) - T vec\left(\mathrm{d}B\right)' \left[B^{-1} \otimes B^{-1\prime}\right] K_g vec\,\mathrm{d}B \\
&+ \frac{1}{2}\left(\mathrm{d}v\left(\Lambda\right)\right)' D_g' \left[\Lambda^{-1} \otimes \Lambda^{-1}\right] D_g \mathrm{d}v\left(\Lambda\right) \\
&- \sum_{t=1}^{T} vec\left(\mathrm{d}A\right)' \left(H_A \otimes I_g\right)' \left[D' A^{*-1} \Psi^* z_t^* z_t^{*\prime} Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1}\right] \left(H_\Lambda \otimes I_g\right) D_g \mathrm{d}v\left(\Lambda\right) \\
&- \sum_{t=1}^{T} vec\left(\mathrm{d}A\right)' \left(H_A \otimes I_g\right)' \left[D' A^{*-1} \Psi^* z_t^* z_t^{*\prime} Psi^{*\prime} A^{*-1\prime} D \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right] \left(H_A \otimes I_g\right) vec\,\mathrm{d}A \\
&+ \sum_{t=1}^{T} vec\,\mathrm{d}A' \left(H_A \otimes I_g\right)' \left[D' A^{*-1} \Psi^* z_t^* z_t^{*\prime} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right] \left(H_\Psi \otimes I_{gk}\right) vec\,\mathrm{d}\Psi \\
&+ \sum_{t=1}^{T} vec\,\mathrm{d}A' \left(H_A \otimes I_g\right)' \left[D' A^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime}\right] K_{(g+k)s} \left(H_\Psi \otimes I_{(g+k)}\right) vec\,\mathrm{d}\Psi \\
&- 2\sum_{t=1}^{T} vec\,\mathrm{d}A' \left(H_A \otimes I_g\right)' \left[D' A^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} D\right] K_{gs} \left(H_A \otimes I_g\right) vec\,\mathrm{d}A \\
&+ \sum_{t=1}^{T} vec\,\mathrm{d}\Psi' \left(H_\Psi \otimes I_{(g+k)}\right)' \left[z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} \otimes \Lambda^{*-1}\right] \left(H_\Lambda \otimes I_g\right) D_g \mathrm{d}v\left(\Lambda\right) \\
&- \sum_{t=1}^{T} \mathrm{d}v\left(\Lambda\right)' D_g' \left(H_\Lambda \otimes I_g\right)' \left[\Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} \otimes \Lambda^{*-1}\right] \left(H_\Lambda \otimes I_g\right) D_g \mathrm{d}v\left(\Lambda\right) \\
&- \sum_{t=1}^{T} vec\,\mathrm{d}\Psi' \left(H_\Psi \otimes I_{(g+k)}\right)' \left[z_t^* z_t^{*\prime} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right] \left(H_\Psi \otimes I_{gk}\right) vec\,\mathrm{d}\Psi.
\end{aligned}$$

The Hessian matrix $\mathcal{H}_T\left(\theta_0\right)$, defined as

$$\mathrm{d}^2 l\left(\theta_0\right) = \left(\mathrm{d}\theta\right)' \mathcal{H}_T\left(\theta_0\right) \mathrm{d}\theta,$$

can be then easily obtained.

As already introduced in Eq. (40), the diagonal T^* matrix, of dimension $(gs \times gs)$, is defined as

$$T^* = \begin{pmatrix} g\left\{\begin{matrix} T_1 & & \\ & T_1 & \\ & & T_1 \end{matrix}\right. & & \\ & \ddots & \\ & & g\left\{\begin{matrix} T_s & & \\ & T_s & \\ & & T_s \end{matrix}\right. \end{pmatrix} \tag{102}$$

where $T_1, \ldots T_s$ indicate the number of observations in each state of volatility.

Given the particular definition of y_t^*, x_t^*, and as a consequence z_t^*, the following expected values take the form

$$\begin{aligned}
E\left(\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime}\right) &= A^* E\left(\varepsilon_t^* \varepsilon_t^{*\prime}\right) A^{*\prime} = A^* \Lambda_t^* A^{*\prime} \\
\Rightarrow \quad \textstyle\sum_{t=1}^{T} E\left(\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime}\right) &= A^* \textstyle\sum_{t=1}^{T} E\left(\varepsilon_t^* \varepsilon_t^{*\prime}\right) A^{*\prime} = A^* T^* \left(I_* \otimes \Lambda\right) A^{*\prime} = A^* T^* \Lambda^* A^{*\prime}
\end{aligned} \tag{103}$$

and, with some algebra

$$\begin{aligned} & E\left(\Psi^{*} z_t^{*} z_t^{*\prime}\right) = A^{*}\Lambda_t^{*} A^{*\prime} \left[\begin{array}{cc} I_s \otimes B^{-1\prime} & 0 \end{array} \right] = A^{*}\Lambda_t^{*} A^{*\prime} C \\ \Rightarrow \quad & \textstyle\sum_{t=1}^{T} E\left(\Psi^{*} z_t^{*} z_t^{*\prime}\right) = A^{*} T^{*}\left(I_{*} \otimes \Lambda\right) A^{*\prime} C = A^{*} T^{*} \Lambda^{*} A^{*\prime} C \end{aligned} \tag{104}$$

The expected value of the second differential thus becomes

$$\begin{aligned} -E\left(\mathrm{d}^2 l\left(\theta\right)\right) \; = \; & vec\left(I_s \otimes \mathrm{d}A\right)^{\prime}\left[D A^{*-1} \otimes A^{*-1\prime} T^{*} D\right] K_{gs} vec\left(I_s \otimes \mathrm{d}A\right) \\ + \; & T\, vec\left(\mathrm{d}B\right)^{\prime}\left[B^{-1} \otimes B^{-1\prime}\right] K_g vec\left(\mathrm{d}B\right) \\ - \; & \frac{T}{2}\left(\mathrm{d}v\left(\Lambda\right)\right)^{\prime} D_g^{\prime}\left[\Lambda^{-1} \otimes \Lambda^{-1}\right] D_g\left(\mathrm{d}v\left(\Lambda\right)\right) \\ + \; & vec\left(\mathrm{d}A\right)^{\prime}\left(H_A \otimes I_g\right)^{\prime}\left[D T^{*} \otimes A^{*-1\prime} \Lambda^{*-1}\right]\left(H_\Lambda \otimes I_g\right) D_g\left(\mathrm{d}v\left(\Lambda\right)\right) \\ + \; & vec\left(\mathrm{d}A\right)^{\prime}\left(H_A \otimes I_g\right)^{\prime}\left[D T^{*} \Lambda^{*-1} D \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right]\left(H_A \otimes I_g\right) vec\left(\mathrm{d}A\right) \\ - \; & vec\left(\mathrm{d}A\right)^{\prime}\left(H_A \otimes I_g\right)^{\prime}\left[D T^{*} \Lambda^{*-1} A^{*} C \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right]\left(H_\Psi \otimes I_{gk}\right) vec\left(\mathrm{d}\Psi\right) \\ - \; & vec\left(\mathrm{d}A\right)^{\prime}\left(H_A \otimes I_g\right)^{\prime}\left[D A^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1} T^{*} \Lambda^{*} A^{*} C\right] K_{(g+k)s}\left(H_\Psi \otimes I_{(g+k)}\right) vec\left(\mathrm{d}\Psi\right) \\ + \; & 2\left(\mathrm{d}A\right)^{\prime}\left(H_A \otimes I_g\right)^{\prime}\left[D A^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1} T^{*} \Lambda^{*} D\right] K_{gs}\left(H_A \otimes I_g\right) vec\left(\mathrm{d}A\right) \\ - \; & vec\left(\mathrm{d}\Psi\right)^{\prime}\left(H_\Psi \otimes I_{(g+k)}\right)^{\prime}\left[C^{\prime} A^{*} \Lambda^{*} T^{*} \otimes A^{*-1\prime} \Lambda^{*-1}\right]\left(H_\Lambda \otimes I_g\right) D_g\left(\mathrm{d}v\left(\Lambda\right)\right) \\ + \; & \left(\mathrm{d}v\left(\Lambda\right)\right)^{\prime} D_g^{\prime}\left(H_\Lambda \otimes I_g\right)^{\prime}\left[\Lambda^{*-1} T^{*} \otimes \Lambda^{*-1}\right]\left(H_\Lambda \otimes I_g\right) D_g\left(\mathrm{d}v\left(\Lambda\right)\right) \\ + \; & vec\left(\mathrm{d}\Psi\right)^{\prime}\left(H_\Psi \otimes I_{(g+k)}\right)^{\prime}\left[Q^{*} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right]\left(H_\Psi \otimes I_{gk}\right) vec\left(\mathrm{d}\Psi\right) \end{aligned} \tag{105}$$

Finally, since $\mathrm{d}\left(\theta\right) = \left(\mathrm{d}vec\, A^{\prime}, \mathrm{d}vec\, B^{\prime}, \mathrm{d}vec\, \Gamma^{\prime}, \mathrm{d}v\ \left(\Lambda\right)^{\prime}\right)^{\prime}$, with some algebra the result follows.

The score vector, instead, can be derived using the properties of the *vec* and trace operators in the first differential in Eq. (100) as follows

$$\begin{aligned} \mathrm{d}l\left(\theta\right) = \; - \; & \sum_{t=1}^{T} vec\left(\left(I_g + A D_t\right)^{\prime} D_t\right)^{\prime} vec\, \mathrm{d}A + T\, vec\left(B^{-1\prime}\right)^{\prime} vec\, \mathrm{d}B - \frac{1}{2} vec\left(\Lambda^{-1}\right)^{\prime} D_g \mathrm{d}v\left(\Lambda\right) \\ - \; & \sum_{t=1}^{T} vec\left(D A^{*-1} \Psi^{*} z_t^{*} z_t^{*\prime} \Psi^{*\prime} \Lambda^{*-1} A^{*-1\prime}\right)^{\prime} K_{gs}\left(H_A \otimes I_g\right) vec\, \mathrm{d}A \\ + \; & \frac{1}{2} \sum_{t=1}^{T} vec\left(\Lambda^{*-1} A^{*-1} \Psi^{*} z_t^{*} z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime}\right)^{\prime}\left(H_\Lambda \otimes I_g\right) D_g \mathrm{d}v\left(\Lambda\right) \\ - \; & \sum_{t=1}^{T} vec\left(A^{*-1\prime} \Lambda^{*-1} A^{*-1} \Psi^{*} z_t^{*} z_t^{*\prime}\right)^{\prime}\left(H_\Psi \otimes I_{(g+k)}\right) vec\, \mathrm{d}\Psi \end{aligned} \tag{106}$$

and, since the score vector, in row form, is defined as

$$f^{\prime}\left(\theta\right) = \frac{\mathrm{d}l\left(\theta\right)}{\mathrm{d}vec\, \theta} \tag{107}$$

with simple algebra the result follows. ■

B Appendix: Tables and Figures

Table 1: Covariances and Correlations among different sub periods

	ir-ge	pt-ge	gr-ge	sp-ge	ir-ge	pt-ge	gr-ge	sp-ge
	0.006	-0.002	-0.006	-0.001	1	-0.128	-0.324	-0.166
Jan 2005 - Sept 2007	-0.002	0.016	0.026	0.004	-0.128	1	0.886	0.744
t = 143	-0.006	0.026	0.056	0.007	-0.324	0.886	1	0.661
	-0.001	0.004	0.007	0.002	-0.166	0.744	0.661	1
	0.232	0.222	0.410	0.178	1	0.963	0.990	0.978
Oct 2007 - Dec 2008	0.222	0.230	0.405	0.179	0.963	1	0.982	0.985
t = 66	0.410	0.405	0.739	0.320	0.990	0.982	1	0.986
	0.178	0.179	0.320	0.143	0.978	0.985	0.986	1
	3.380	1.730	3.530	1.300	1	0.985	0.983	0.987
Jan 2009 - Dec 2009	1.730	0.916	1.830	0.680	0.985	1	0.982	0.989
t = 52	3.530	1.830	3.810	1.380	0.983	0.982	1	0.987
	1.300	0.680	1.380	0.516	0.987	0.989	0.987	1
	14.40	11.20	26.30	6.420	1	0.985	0.971	0.979
Jan 2010 - Mar 2011	11.20	8.980	21.10	5.110	0.985	1	0.990	0.987
t = 62	26.30	21.10	50.70	12.20	0.971	0.990	1	0.988
	6.420	5.110	12.20	2.990	0.979	0.987	0.988	1

Table 2: Tranquil and crises windows

Definition of the windows	Start	End
Tranquil Periods	2005-01-05	2007-05-30
Irish Turbulences	2008-09-17	2009-08-26
	2010-04-07	2011-03-09
Portuguese Turbulences	2010-02-03	2011-03-09
Greek Turbulences	2010-04-21	2011-03-09
Spanish Turbulences	2007-06-01	2011-03-09

Table 3: Number of observation for each volatility regimes

Ireland	Portugal	Greece	Spain	Historical Events (1)	Rolling Variances (2)
L	L	L	L	118	258
H	L	L	L	0	4
L	H	L	L	0	2
L	L	H	L	0	0
L	L	L	H	39	11
H	H	L	L	0	2
H	L	L	H	103	7
H	L	H	L	0	1
L	H	H	L	0	3
L	L	H	H	0	1
L	H	L	H	9	0
L	H	H	H	0	1
H	L	H	H	0	0
H	H	L	H	0	9
H	H	H	L	0	2
H	H	H	H	49	9

Table 4: FIML estimates of A and B.

	Historical events Unrestricted (1)		Historical events Restricted (2)		Rolling Variances (3)	
	param.	SE	param.	SE	param.	SE
A_{11}	3.320	0.333	3.550	0.286	4.030	0.622
A_{21}	-0.352	0.359	0	-	-1.490	0.332
A_{31}	-0.491	0.741	0	-	5.890	0.775
A_{41}	-0.456	0.638	0	-	-3.510	0.475
A_{12}	0.401	0.566	0	-	1.660	0.443
A_{22}	3.950	0.539	4.070	0.542	4.400	0.763
A_{32}	-5.140	1.270	-5.090	1.210	-9.490	1.370
A_{42}	4.230	1.120	4.240	1.100	1.890	0.432
A_{13}	-1.200	1.080	0	-	-2.340	0.840
A_{23}	0.674	1.210	0	-	3.120	0.837
A_{33}	10.500	1.910	10.600	1.540	14.700	3.100
A_{43}	-7.680	1.640	-7.530	1.360	-4.000	1.170
A_{14}	1.730	0.476	1.040	0.318	2.350	0.408
A_{24}	1.260	0.456	1.520	0.348	0.750	0.301
A_{34}	-5.940	0.499	-5.870	0.459	-7.750	0.956
A_{44}	4.700	0.558	4.740	0.520	3.940	0.590
B_{11}	1	-	1	-	1	-
B_{21}	-0.182	0.068	-0.132	0.041	-0.294	0.027
B_{31}	-0.300	0.081	-0.191	0.067	-0.558	0.030
B_{41}	0.211	0.068	0.267	0.062	0.217	0.025
B_{12}	-0.220	0.097	-0.232	0.061	-0.526	0.033
B_{22}	1	-	1	-	1	-
B_{32}	-0.377	0.101	-0.408	0.094	-0.615	0.036
B_{42}	0.366	0.103	0.390	0.093	-0.077	0.032
B_{13}	-0.255	0.065	-0.184	0.038	-0.052	0.015
B_{23}	-0.323	0.060	-0.364	0.044	-0.176	0.015
B_{33}	1	-	1	-	1	-
B_{43}	-0.837	0.056	-0.849	0.055	-0.306	0.013
B_{14}	-0.931	0.123	-1.060	0.110	-0.504	0.045
B_{24}	-0.210	0.104	-0.239	0.081	-0.200	0.046
B_{34}	-0.476	0.129	-0.739	0.095	-0.478	0.047
B_{44}	1	-	1	-	1	-
LR test for over-identified restrictions						
		LR	$\chi^2_6 =$	10.125		
		p-value		(0.119)		

Table 5: Simultaneous relations over different volatility regimes

	ir-ge	pt-ge	gr-ge	sp-ge	ir-ge	pt-ge	gr-ge	sp-ge
	Historical Events				Rolling Variances			
	LLLL[a]				LLLL			
ir-ge	1	-0.232	-0.184	-1.060	1	-0.526	-0.052	-0.504
pt-ge	-0.132	1	-0.364	-0.239	-0.294	1	-0.176	-0.2
gr-ge	-0.191	-0.408	1	-0.739	-0.558	-0.615	1	-0.478
sp-ge	0.267	0.390	-0.849	1	0.217	-0.077	-0.306	1
	LLLH				LLLH			
ir-ge	1	-0.318	-0.032	-1.300	1	-0.545	0.104	-1.090
pt-ge	-0.226	1	-0.155	-0.562	-0.323	1	-0.128	-0.348
gr-ge	0.623	-0.067	1	2.150	-0.418	-1.420	1	2.100
sp-ge	0.267	0.390	-0.849	1	0.217	-0.077	-0.306	1
	HLLH				HLLH			
ir-ge	1	-0.318	-0.032	-1.300	1	-0.545	0.104	-1.090
pt-ge	-0.226	1	-0.155	-0.562	-0.211	1	-0.123	-0.522
gr-ge	0.623	-0.070	1	2.150	-0.523	-1.380	1	2.230
sp-ge	0.267	0.390	-0.849	1	1.410	-0.686	-0.528	1
	LHLH				LHLH			
ir-ge	1	-0.140	-0.063	-1.470	1	-0.677	0.119	-0.981
pt-ge	-0.226	1	-0.155	-0.562	-0.323	1	-0.128	-0.348
gr-ge	0.412	1.540	1	1.780	-1.790	1.560	1	1.810
sp-ge	0.315	-0.372	-0.454	1	0.299	-0.399	-0.228	1
	HHHH				HHHH			
ir-ge	1	-0.217	-0.112	-1.580	1	-0.667	0.153	-0.979
pt-ge	-0.271	1	-0.229	-0.720	-0.078	1	-0.260	-0.770
gr-ge	0.412	1.540	1	1.780	-1.410	1.290	1	1.430
sp-ge	0.297	-0.105	-0.235	1	0.606	-0.596	-0.058	1

a: LLLL indicates (Ir = Low, Pt = Low, Gr = Low, Sp = Low), and so forth for all other regimes.

Figure 1: Interest rates (left panel) and spreads (right panel), January 2005 - March 2011.

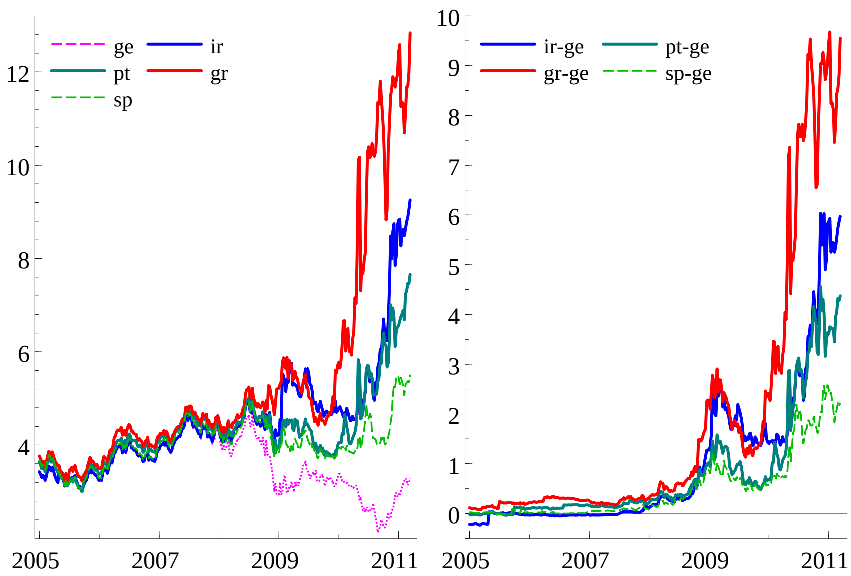


Figure 2: Actual and fitted spreads (left panel), and residuals (right panel), January 2005 - March 2011.

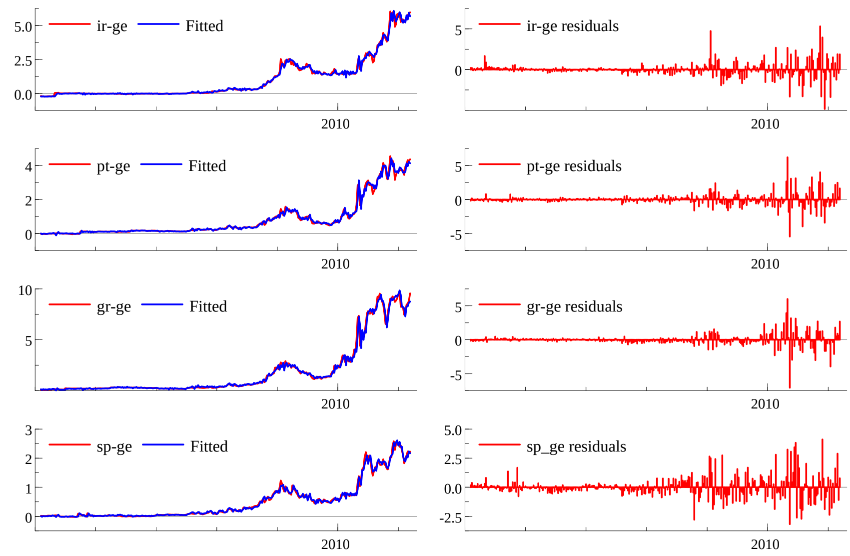


Figure 3: Absolute value of the residuals, January 2005 - March 2011.

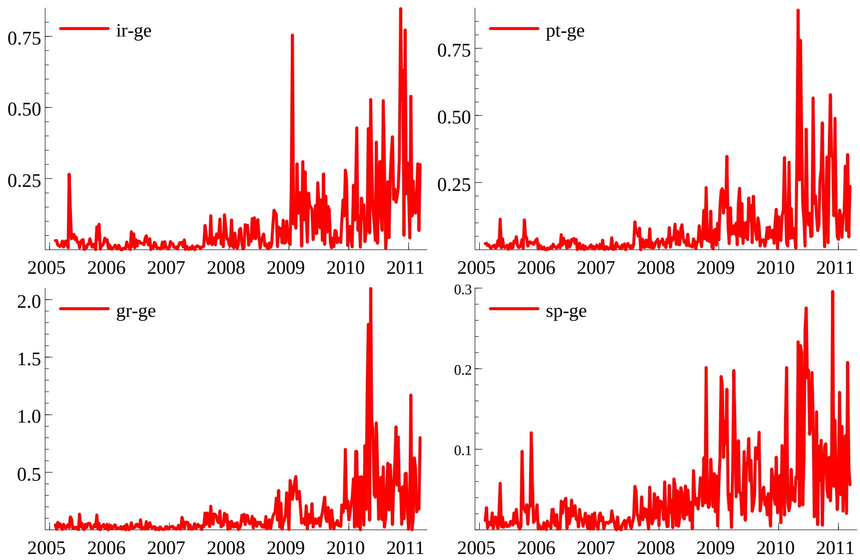
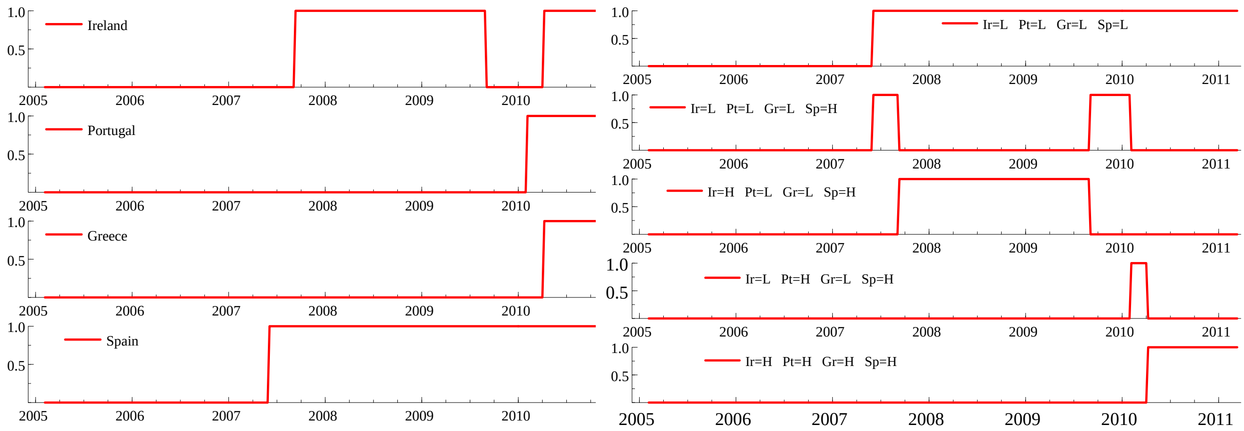


Figure 4: High-low volatility regimes for each country (left panel), and combination of the 5 different regimes (right panel).



C Appendix: Low, Mid or High volatility for each endogenous variable

Suppose that there are two different regimes of volatility governed by the two $0-1$ diagonal matrices D_1 and D_2, as considered in Section II.2. Moreover, let consider the more general case in which, when there is an increase in the volatility, this can happen either in a state of mid volatility, or in a state of high volatility[17].

Furthermore, as in the specification considered in Proposition 1, let the two diagonal matrices describing the volatility regimes be defined as $D_1 = 0_g$ and $D_2 = I_g$. Under these further assumptions, the structural representation of the model (for simplicity, without considering the predetermined variables) becomes

$$\begin{aligned} By_t &= \varepsilon_t \\ By_t &= (I_g + A)\,\varepsilon_t \\ By_t &= \left(I_g + A + \bar{A}\right)\varepsilon_t \end{aligned} \tag{108}$$

and the system of equations connecting the parameters of the reduced and structural forms can be written as

$$\begin{aligned} B\Omega_1 B' - \Lambda &= 0 \\ (I_g + A)^{-1}\, B\Omega_2 B'\,(I_g + A)^{-1\prime} - \Lambda &= 0 \\ \left(I_g + A + \bar{A}\right)^{-1} B\Omega_3 B' \left(I_g + A + \bar{A}\right)^{-1\prime} - \Lambda &= 0. \end{aligned} \tag{109}$$

Using the standard derivation rules, and some properties of the *vec* operator, the Jacobian

[17] Actually, the possible combinations of low-medium-high volatility regimes can be mixed in a much more complicated way. The strategy followed in this appendix is an illustrative example on how to generalize the results obtained in the previous sections of the paper.

is

$$J=\begin{pmatrix} 2N_gD_1^* & 0 & 0 & -D_g \\ 2N_gD_2^* & -2N_gE_1^* & 0 & -D_g \\ 2N_gD_3^* & -2N_gE_2^* & -2N_g\bar{E}_1^* & -D_g \\ R_B & 0 & 0 & 0 \\ 0 & R_A & 0 & 0 \\ 0 & 0 & R_{\bar{A}} & 0 \\ 0 & 0 & 0 & R_\Lambda \end{pmatrix} \tag{110}$$

where N_g and D_g are defined as before, and the matrices D_1^*, D_2^*, D_3^*, E_1^*, E_2^*, $\bar{E}_1^*$ are defined as follows

$$\begin{aligned} D_1^* &= B\Omega_1 \otimes I_g \\ D_2^* &= (I_g+A)^{-1} B\Omega_2 \otimes (I_g+A)^{-1} \\ D_3^* &= \left(I_g+A+\bar{A}\right)^{-1} B\Omega_3 \otimes \left(I_g+A+\bar{A}\right)^{-1\prime} \\ E_1^* &= (I_g+A)^{-1} B\Omega_2 B' (I_g+A)^{-1\prime} \otimes (I_g+A)^{-1} \\ E_2^* &= \left(I_g+A+\bar{A}\right)^{-1} B\Omega_2 B' \left(I_g+A+\bar{A}\right)^{-1\prime} \otimes (I_g+A)^{-1} \\ \bar{E}_1^* &= \left(I_g+A+\bar{A}\right)^{-1} B\Omega_2 B' \left(I_g+A+\bar{A}\right)^{-1\prime} \otimes (I_g+A)^{-1} \end{aligned}$$

The matrices R_A, R_B, $R_{\bar{A}}$ and R_Λ describe the restrictions on the parameters in the implicit form, with associated explicit form defined by S_A, S_B, $S_{\bar{A}}$ and S_Λ. Following Rothenberg (1971), the model is identified when the Jacobian matrix in Eq. (110) has full column rank. After some simple algebra, the identification condition refers to the following $3g^2 \times \left(3g^2+g\,(g+1)\,/2-q\right)$ matrix

$$\begin{pmatrix} 2N_gD_1^*S_B & 0 & 0 & -D_gS_\Lambda \\ 2N_gD_2^*S_B & -2N_gE_1^*S_A & 0 & -D_gS_\Lambda \\ 2N_gD_3^*S_B & -2N_gE_2^*S_A & -2N_g\bar{E}_1^*S_{\bar{A}} & -D_gS_\Lambda \end{pmatrix} \tag{111}$$

where $q=q_A+q_B+q_{\bar{A}}+q_\Lambda$ indicates the total number of restrictions. The rank condition for identification, thus, is satisfied if and only if the matrix in Eq. (111) has full column rank $3g^2+g\,(g+1)\,/2-q$.

The necessary condition can be obtained by counting the number of distinct rows in Eq. (111), which must be at least as large as the number of columns. This easily leads to the following necessary condition

$$\frac{3}{2}g\,(g+1) \geq 3g^2+\frac{1}{2}g\,(g+1)-q \implies q \geq g^2-g. \tag{112}$$

Remark 4. Under the conditions of Corollary 1 of a unit diagonal covariance matrix for the structural shocks, i.e. $\Lambda=I_g$, and the further assumption that $R_B=0$, then the necessary condition becomes

$$q_A+q_{\bar{A}} \geq 2g^2-g-\frac{g\,(g+1)}{2} \implies q_A+q_{\bar{A}} \geq \frac{3}{2}g^2-\frac{3}{2}g. \tag{113}$$

This result can be compared with the identification strategy found in Lanne and Lütkepohl

(2008) and Lanne et al. (2010) where, for three levels of volatility for each endogenous variable, is based on imposing a total of $\left(g^2 - g\right) + g\,(g+1)\,/2$ restrictions, i.e. the diagonal structure of the Λ_1 and Λ_2 matrices in the relations $\Omega_1 = W\Lambda_1 W'$ and $\Omega_2 = W\Lambda_2 W'$, plus the unit diagonal structure for the covariance matrix of the structural shocks in the low volatility regime. This structure is clearly not necessary for the identification of the parameters and provides $\frac{3}{2}g\,(g+1) - g^2 - 2g = g\,(g-1)\,/2$ degrees of freedom in a LR test of hypothesis with asymptotic χ^2 distribution[18].

Equivalent results can be easily re-obtained in our specification by imposing both A and $\bar{A}$ to be diagonal. Our framework, however, is much more flexible, being not constrained on the diagonal structure for the A and $\bar{A}$ matrices.

As an example, in the case of a three-dimensional system of equations the following structure

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \text{ and } \bar{A} = \begin{pmatrix} \bar{a}_{11} & 0 & 0 \\ 0 & \bar{a}_{22} & 0 \\ 0 & 0 & \bar{a}_{33} \end{pmatrix} \tag{114}$$

imposes a set of $g\,(g-1)\,/2 + g^2 - g = \frac{3}{2}g^2 - \frac{3}{2}g$ restrictions, that corresponds to the necessary condition reported in Eq. (113).

[18]See discussion on identification conditions in Lanne et al. (2010), pag. 124.

D Monte Carlo Evidence

In this section we provide some Monte Carlo simulations to evaluate the performances of the FIML estimator in estimating the unknown parameters of the model and the associated asymptotic covariance matrix. We shall also analyze the finite-sample properties of hypothesis testing procedures for evaluating the significance of the estimated coefficients and the presence of high volatility regimes.

In all Monte Carlo simulations we refer to a three-equation system for $y_t = (y_{1t}, y_{2t}, y_{3t})'$ with one common regressor x_t. The Data Generating Process (DGP) is

$$By_t = \Gamma x_t + \Big[I_{gs} + (I_s \otimes A)\, D\Big]\varepsilon_t$$

where $g = 3$ and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$ is the vector of structural uncorrelated shocks with common variances $\sigma_i^2 = 1$, $i = 1, 2, 3$. The system is governed by $s = 4$ different regimes of volatility described by the (12×12) matrix $D = diag\,(D_1, D_2, D_3, D_4)$ where

$$D_1 = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix} \quad D_2 = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 1 \end{pmatrix} \quad D_3 = \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix} \quad D_4 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}.$$

The first regime is characterized by low volatility for all endogenous variables. In the second and third only one variable is in a state of high volatility, y_{3t} and y_{1t} respectively. In the fourth regime, instead, all the variables are in a state of high volatility. The matrices of parameters have been fixed as

$$A = \begin{pmatrix} 1.5 & 0 & 0 \\ 0.5 & 3 & 0 \\ 0.5 & 0 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0.6 & 0.5 \\ 0 & 1 & -0.3 \\ -0.4 & 0 & 1 \end{pmatrix} \quad \Gamma = \begin{pmatrix} 0.7 \\ 0.5 \\ 0.5 \end{pmatrix}$$

while the exogenous variable x_t is randomly generated (differently for each replication) from a standard normal variable. All the different states of volatility are represented in the sample by an equivalent number of observations $T_i = 0.25T$, $i = 1, 2, 3, 4$. We consider three different values for the number of observations, $T = 100, 250, 1500$, that correspond roughly to twenty five years of quarterly data as in hypothetical macroeconomic analyses for the first, six years weekly or daily data as in financial applications for the second and third. The estimation of the parameters has been performed by considering the reduced form parameters first, and then by maximizing the concentrated likelihood with respect to the structural parameters in A and B. Without any constraint in the parameters of the reduced form, this two steps procedure is completely equivalent to the maximization of the general likelihood function[19]. Experiments are based on 5,000 replications of the Monte Carlo process for $T = 100$ and $T = 250$, while 2,000 replications for $T = 1500$.

[19]Given that the estimation of the model is computationally rather intensive, the likelihood function has been maximized through the OPTMUM Gauss procedure, that provides exactly the same results as the scoring algorithm discussed in Eq. (47), but reaching the maximum much more quickly.

D.1 Parameters and Covariance Matrix Estimators

Although the general asymptotic theory for the FIML estimator, as the one reported in Section III, focuses on the primary role of the Fisher information matrix, in practical works alternative methods have been proposed to estimate the covariance matrix of the maximum likelihood estimates, generally based on numerical or analytical first and second order differentiation. A first alternative is the matrix of outer products (OP) of the first order derivatives of the log-likelihood that, in the context of simultaneous systems of equations, is generally referred to as the BHHH method (Berndt et al., 1974). Another possibility, instead, is represented by the Hessian matrix ($\mathcal{H}$), obtained from the second order derivatives of the log-likelihood function. Two more alternatives, that represent two forms of the (robust) quasi-maximum likelihood covariance estimator are worth mentioning. The first, proposed in a very general framework by White (1982, 1983) and Gourieroux et al. (1984), is defined as

$$QMLH = \left(\mathcal{H}^{-1}\right)(OP)\left(\mathcal{H}^{-1}\right) \tag{115}$$

while the second substitutes the Hessian matrix with the estimated information matrix $\mathcal{F}$ as

$$QMLF = \left(\mathcal{F}^{-1}\right)(OP)\left(\mathcal{F}^{-1}\right). \tag{116}$$

Given the results obtained in Proposition 3 concerning analytical formulation for the Gradient, the Hessian matrix, and the Information matrix, it becomes straightforward to compare the behavior of these matrices as estimator of the asymptotic covariance matrix.

In Table 6 we report the true value of the parameters, the mean of their estimates, the mean squared error, and the estimated standard errors provided from the five alternatives mentioned above. For each parameter we report the results for the different sample length T. As expected, the bias of the estimated coefficients reduces as the dimension of the sample increases. However, even for relatively small samples ($T = 100$), the approximation seems to be extremely good. The Monte Carlo results indicate that the five covariance matrix estimators are, indeed, asymptotically similar, but some of them exhibit some systematic inequalities in small samples. In general, the H and F covariance estimates are very similar, for both the A and B parameters, and for the different sample sizes. Concerning the parameters of the A matrix, instead, the different estimators behave differently according to whether they are on or off the main diagonal. For the parameters on the main diagonal, the OP provides the smallest values, while those obtained with the two quasi-maximum likelihood estimators QMLH and QMLF tend to be the largest ones. For the off diagonal elements of the A matrix exactly the opposite happens. Such differences are much less pronounced for the parameters in the B matrix.

D.2 Tests for Volatility Regimes

The present section, and the next one, are dedicated to investigate the behavior of likelihood ratio-type tests (LR) on the coefficients of the A matrix. In particular, in this section we discuss a test for detecting whether one or more structural shocks, in particular periods of time, are amplified by the system and originate clusters of higher volatility. This kind of tests refers to the coefficients on the main diagonal of the A matrix. In fact, if the a_{jj} element of the A matrix is not significantly different from zero, the ε_j never contributes to generate different regimes of volatility.

The Monte Carlo experiment is based on the same structure of the previous one, but with different values for the A matrix, defined as

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{117}$$

and with only $s = 2$ regimes governed by the following D_1 and D_2 matrices

$$D_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad D_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{118}$$

The DGP indicates that only the first structural shock, for the second half of the sample, is amplified, thus generating heteroskedasticity. Using the results of Proposition 2, the zero restrictions on the A and B matrices, together with the two regimes of volatility described by D_1 and D_2, satisfy the necessary and sufficient conditions for the local identification of the parameters.

Given this DGP we study the performances of a LR test whose null hypothesis is of no multiplicative effect for the second and third structural shock in periods of high volatility, i.e. $a_{22} = 0$ and $a_{33} = 0$. In other words, under the null hypothesis, the two structural shocks ε_{2t} and ε_{3t} don't contribute to generate high volatility periods. The size properties of the LR test are evaluated through the P-value plots proposed by Davidson and MacKinnon (1998), which are plots of empirical versus nominal size for all possible test sizes. If the asymptotic distribution is correct, the P-value plot should be close to the 45° line. Figure 5 (left panel) reports such plots for the three different dimensions of the sample, $T = 100$, 250 and 1500. It is clearly shown that the actual size of the LR test behaves correctly for all values of T.

In order to check for the power of the test we employ the size-power curve defined in Davidson and MacKinnon (1998), that is constructed using two empirical distributions of the test statistic, one for an experiment in which the null hypothesis is true, and one for an experiment in which it is false, possibly using the same sequence of random numbers. The distribution of the test statistics under the alternative hypothesis is obtained using a DGP in which the A matrix is defined as

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}, \tag{119}$$

which is very close to the DGP for obtaining the distribution under the null hypothesis[20]. The difference, of course, is that under the alternative, the second and third structural shocks, during periods of high instability, are amplified through a multiplicative effect of magnitude equal to 1.2. Although this is rather close to the situation of absence of an amplification of the shock, we want to verify that the LR test rejects the null hypothesis a correct number of times. The size-power curves for the LR test on the amplification effects are reported in Figure 5 (right panel), for different dimensions of the sample. The power of the test appears to be practically perfect for very large samples ($T = 1500$), highly reliable

[20]In this new DGP, the matrix describing the behavior of the structural shocks in periods of high instability is represented by the identity matrix, i.e. $D_2 = I_3$, while $D_1 = 0_3$

for medium samples ($T = 250$), while becomes marginally satisfactory for small samples ($T = 100$). In this last case, however, the power always remains higher than the actual size of the test.

D.3 Tests for the Propagation of Shocks

As already introduced in the previous section, the following Monte Carlo exercise uses a slightly different DGP to investigate the performances of a LR test that aims to detect the propagation of shocks through the off diagonal elements of the A matrix. The only change in the DGP is represented by the A matrix used to generate the distribution of the test statistic under the null, which is defined as

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{120}$$

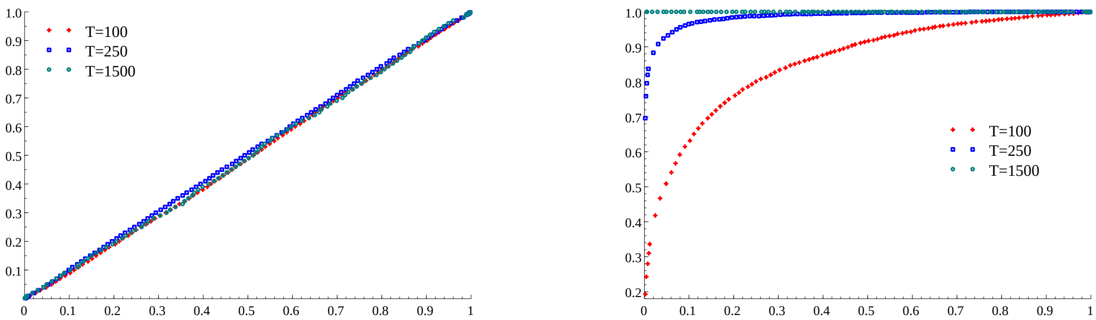
in which all the shocks, in periods of high volatility, are amplified but not transmitted to the other variables of the system[21]. Suppose to have a priori information for the A matrix to be *at most* lower triangular. The null hypothesis we want to test is that there is not at all propagation of shocks, i.e. $a_{21} = a_{31} = a_{32} = 0$. In this case, the LR test statistic is asymptotically distributed as a χ^2 distribution with three degrees of freedom. The performances of the test, in terms of size, are described by the P-value plot reported in Figure 6 (left panel), for different dimensions of the sample. As for the previous case, the empirical size of the test strongly corresponds to the nominal size, even in small samples.

In the right panel of Figure 6 we report the size-power curves, for different dimensions of the sample, obtained by considering the DGP in which the A matrix is defined as:

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.5 & 0.5 & 1 \end{pmatrix}. \tag{121}$$

The size-power curves highlight that the power is extremely good in large samples, while reduces for medium and small samples.

Figure 5: P-value plots (left panel) and size-power curve (right panel): Amplification effects



[21] As before, $D_1 = 0_3$ and $D_2 = I_3$.

Table 6: Mean estimated parameters and variances

Par.	True	T	Mean	MSE	OP	H	F	QMLH	QMLF
		100	1.465	0.224	0.269	0.271	0.269	0.299	0.298
A_{11}	1.5	250	1.492	0.155	0.160	0.171	0.171	0.192	0.192
		1500	1.498	0.060	0.063	0.070	0.070	0.080	0.080
		100	0.484	0.226	0.271	0.231	0.227	0.227	0.219
A_{21}	0.5	250	0.486	0.145	0.152	0.143	0.142	0.142	0.141
		1500	0.499	0.057	0.059	0.059	0.058	0.060	0.059
		100	0.505	0.372	0.312	0.281	0.276	0.294	0.276
A_{31}	0.5	250	0.538	0.442	0.181	0.175	0.174	0.181	0.179
		1500	0.501	0.136	0.071	0.072	0.072	0.075	0.075
		100	2.939	0.505	0.565	0.567	0.565	0.637	0.631
A_{22}	3	250	2.983	0.321	0.337	0.363	0.363	0.419	0.417
		1500	2.995	0.129	0.131	0.148	0.148	0.174	0.173
		100	1.975	0.363	0.351	0.371	0.367	0.437	0.424
A_{33}	2	250	1.988	0.230	0.210	0.235	0.234	0.276	0.275
		1500	2.002	0.093	0.083	0.096	0.096	0.113	0.113
		100	-0.386	0.205	0.149	0.142	0.140	0.154	0.147
B_{31}	-0.4	250	-0.373	0.255	0.088	0.088	0.088	0.094	0.094
		1500	-0.399	0.076	0.034	0.036	0.036	0.039	0.039
		100	0.594	0.077	0.110	0.099	0.098	0.100	0.098
B_{12}	0.6	250	0.594	0.058	0.065	0.063	0.062	0.063	0.062
		1500	0.600	0.021	0.025	0.026	0.025	0.026	0.026
		100	0.500	0.079	0.092	0.082	0.081	0.082	0.080
B_{13}	0.5	250	0.506	0.075	0.053	0.051	0.051	0.051	0.051
		1500	0.501	0.023	0.020	0.021	0.020	0.021	0.021
		100	-0.299	0.060	0.090	0.080	0.079	0.079	0.077
B_{23}	-0.3	250	-0.302	0.040	0.052	0.050	0.049	0.050	0.049
		1500	-0.300	0.015	0.020	0.020	0.020	0.021	0.020

Figure 6: P-value plots (left panel) and size-power curve (right panel): Propagation effects