Identification through heteroskedasticity: A likelihood-based approach

Emanuele Bacchiocchi*

15th June 2010

PRELIMINARY VERSION

Abstract

In this paper we show how the analysis of identification of simultaneous system of equations with different volatility regimes can be addressed is conventional likelihood-based setup, generalizing previous works in different directions. We discuss general conditions for identification and one of the results shows that an adequate number of different levels of heteroskedasticity is sufficient to identify the parameters of the structural form without the inclusion of any kind of restrictions. A Full Information Maximum Likelihood (FIML) algorithm is discussed.

Keywords: simultaneous equations model, heteroskedasticity, identification, FIML.
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 this field. 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 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. In some cases, however, such economic knowledge of the problem, might be not 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. Klein and Vella (2003) and

*Department of Economics, University of Milan. Email emanuele.bacchiocchi@unimi.it

Lewbel (2008) also uses heteroskedasticity to identify simultaneous and mismeasured equation models, imposing restrictions on the correlation between some regressors and the heteroskedastic errors. In all these papers, however, the inspiration for identification and estimation of the parameters comes from the instrumental variable approach or generalized method of moments (GMM).

In the present paper, instead, we provide a specification of simultaneous equation systems that explicitly models the heteroskedasticity of the structural shocks, and allows us to write and treat the likelihood function. The identification conditions, thus, are studied following the Rothenberg (1971) approach, and the estimation of the parameters is performed through the Full Information Maximum Likelihood (FIML) procedure.

This strategy allows us to generalize the results in Rigobon (2003) and provide necessary and sufficient conditions for identification of the structural parameters in the case of general systems of equations with non diagonal variance covariance matrix of the structural residuals. The second main result concern the conditions for identification when combining the information on the different states of volatility, with those coming from economic theory, leading to linear restrictions in the parameters of the structural form.

The rest of the paper is organized as follows: In Section II we first present the statistical model and derive the conditions for identification of the structural parameters. Section III describes the statistical inference while Section IV concludes.

II Identification in a simultaneous equation model with heteroskedasticity

The idea, directly inspired from Rigobon (2003), is to increase the number of relations that link the parameters in the reduced form to those in the structural form. Consider the following simple case:

$$p_t = \beta q_t + \epsilon_t \tag{1}$$
$$q_t = \alpha p_t + \eta_t \tag{2}$$

where ϵ_t and η_t are the uncorrelated structural shocks with variances σ_ϵ^2 and σ_η^2, respectively. The model, without restrictions, is not identified and, given the data for p_t and q_t, one can only estimate the covariance matrix of the reduced form $\hat{\Omega}$:

$$\hat{\Omega} = \frac{1}{(1-\alpha\beta)^2} \begin{bmatrix} \beta^2\sigma_\eta^2 + \sigma_\epsilon^2 & \beta\sigma_\eta^2 + \alpha\sigma_\epsilon^2 \\ \cdot & \sigma_\eta^2 + \alpha^2\sigma_\epsilon^2 \end{bmatrix} \tag{3}$$

The model is clearly not identified in that the covariance matrix of the reduced form provides only three moments faced of four unknowns in the structural form: α, β, σ_ϵ^2 and σ_η^2. Following the traditional approach, one single exclusion constraint in either α or β reaches the identification. If, instead, we suppose two regimes of volatility for the structural shocks, and that the structural parameters remain constant over the regimes, then we will have six unknowns (α, β, $\sigma_{\epsilon 1}^2$, $\sigma_{\eta 1}^2$, $\sigma_{\epsilon 2}^2$ and $\sigma_{\eta 2}^2$), but two covariance matrix with three estimated moments each. The system is thus composed by six equations in six unknowns, providing a full correspondence between the parameters of the structural and reduced forms. If such equations are independent, the model is identified. This approach has an instrumental variable interpretation, in that, in a classical supply-demand scheme as (1), the rise in the variances of one of the two structural shocks can be seen as a valid probabilistic instrument indicating a higher probability of the curve to move leaving unchanged the other.

II.1 A simultaneous equation model with heteroskedastic errors

In this section we present an alternative specification 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:

$$By_t = (I_g + AD_t)\,\epsilon_t \tag{4}$$

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 $(g \times g)$ invertible matrix that captures the different states of volatility, 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:

$$\begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \left(I_2 + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \right) \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}. \tag{5}$$

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. 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 $E(\epsilon_t \epsilon_t') = \Lambda$. 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 restriction 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. If the A matrix is restricted to be diagonal, the structural shocks do not propagate to other variables directly through the covariance matrix. These interpretations, of course, apply to the more general model in (4).

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

$$y_t = B^{-1}(I_g + AD_t)\,\epsilon_t \tag{6}$$

or equivalently

$$y_t = B^{-1} C_t \epsilon_t \tag{7}$$

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

The heteroskedasticity, thus, is intended as different regimes of volatility that might apply to one or more variables in the system. This approach to model the heteroskedasticity implicitly imposes a maximum number of regimes which, however, becomes consistent as the number of dependent variables increases. 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. Alternative approaches, instead, uses ARCH-based model for the residuals of the reduced form in order to obtain identification[1].

If we suppose that the structural shocks follow a normal random variable, the likelihood function of the structural form in (4) can be written in the following way:

$$L(B, A, \Lambda) = \prod_{t=1}^{T} \left| C_t^{-1} B \right| (2\pi)^{-g\frac{T}{2}} |\Lambda|^{-\frac{T}{2}} exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \left(C_t^{-1} B y_t \right)' \Lambda^{-1} \left(C_t^{-1} B y_t \right) \right\} \tag{8}$$

and the log likelihood function

$$\begin{aligned} l(B, A, \Lambda) &= C + \sum_{t=1}^{T} \log \left| C_t^{-1} B \right| - \frac{T}{2} \log |\Lambda| - \frac{1}{2} \sum_{t=1}^{T} \left(C_t^{-1} B y_t \right)' \Lambda^{-1} \left(C_t^{-1} B y_t \right) \\ &= C - \sum_{t=1}^{T} \log |C_t| + T \log |B| - \frac{T}{2} \log |\Lambda| - \frac{1}{2} \sum_{t=1}^{T} \left(C_t^{-1} B y_t \right)' \Lambda^{-1} \left(C_t^{-1} B y_t \right) \end{aligned} \tag{9}$$

[1] See Caporale et al. (2002), Dungey and Martin (2001), King et al. (1994), Rigobon (2002). Sentana and Fiorentini (2001) study identification in conditionally heteroskedastic factor models.

where T is the number of observation, g is the number of equations and, as before, $C_t = (I_g + AD_t)$. As we will see below, the likelihood function will be used for estimating the unknown parameters of the model by means of the FIML approach.

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

In this section we study the identification of the simultaneous equation system in (4) in the simplest case of $s = 2$ regimes of volatility. The generalization is not straightforward and will be the argument of the next section. 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 at each state of volatility the system is. The model, thus, can be rewritten with two separate equations, one for each regime:

$$By_t = (I_g + AD_1)\,\epsilon_t \tag{10}$$
$$By_t = (I_g + AD_2)\,\epsilon_t \tag{11}$$

where equation (10) is for the observations in the first state of volatility, and (11) for those in the second. The associated covariance matrices for the error terms are:

$$E\left(C_t\epsilon_t\epsilon_t'C_t'\right) = (I_g + AD_1)\,\Lambda\,(I_g + AD_1)' \tag{12}$$
$$E\left(C_t\epsilon_t\epsilon_t'C_t'\right) = (I_g + AD_2)\,\Lambda\,(I_g + AD_2)'. \tag{13}$$

The reduced form of the model can be written as

$$y_t = B^{-1}\,(I_g + AD_1)\,\epsilon_t \tag{14}$$
$$y_t = B^{-1}\,(I_g + AD_2)\,\epsilon_t \tag{15}$$

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

$$E\left(y_ty_t'\right) = B^{-1}\,(I_g + AD_1)\,\Lambda\,(I_g + AD_1)'\,B^{-1\prime} = \Omega_1 \tag{16}$$
$$E\left(y_ty_t'\right) = B^{-1}\,(I_g + AD_2)\,\Lambda\,(I_g + AD_2)'\,B^{-1\prime} = \Omega_2. \tag{17}$$

If we assume that the structural shocks ϵ_t behave as a multivariate normal variable, the identification can be studied as in the traditional 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. Following Rothenberg (1971), the identifiability of the structure depends on the uniqueness of solutions of the following system

$$(I_g + AD_1)^{-1}\,B\Omega_1B'\,(I_g + AD_1)^{-1\prime} - \Lambda = 0 \tag{18}$$
$$(I_g + AD_2)^{-1}\,B\Omega_2B'\,(I_g + AD_2)^{-1\prime} - \Lambda = 0 \tag{19}$$
$$R_A vec\,A + R_B vec\,B + R_\Lambda v\,(\Lambda) - r = 0 \tag{20}$$

where $\psi\,(A, B, \Lambda) = R_A vec\,A + R_B vec\,B + R_\Lambda v\,(\Lambda) - r = 0$ is a set of linear restrictions on the parameters A, B and Λ. The vector $v\,(\Lambda)$ denotes the $\frac{1}{2}g\,(g+1)$ elements that is obtained from $vec\Lambda$ by eliminating the supra diagonal elements of Λ or, equivalently, $D_g v\,(\Lambda) = vec\Lambda$, with D_g the *duplication matrix*[2]. Equations (18)-(20) form a system of non-linear equations (because of (18) and (19)) in A, B and $v\,(\Lambda)$. Differentiating (18)-(20) gives

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

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

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

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

where K_{mn} is the *commutation matrix*[3] and is defined such that $K_{mn}vecA=vecA'$. Using the property of the commutation matrix and duplication matrix, we rewrite the system as

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

The Jacobian matrix, thus, can be written as

$$J\left(A,B\right)=\begin{pmatrix}
-2N_g\left[C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\otimes C_1^{-1}\right] & 2N_g\left[C_1^{-1}B\Omega_1\otimes C_1^{-1}\right] & -D_g\\
-2N_g\left[C_2^{-1}B\Omega_2B'C_2^{-1\prime}D_2\otimes C_2^{-1}\right] & 2N_g\left[C_2^{-1}B\Omega_2\otimes C_2^{-1}\right] & -D_g\\
R_A & R_B & R_\Lambda
\end{pmatrix}. \tag{21}$$

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 $g\left(g+1\right)/2$. We note that the Jacobian matrix only depends on A and B, and not on Λ (since the non-linearity in (18)-(20) are on A and B). Following Rothenberg (1971), a sufficient condition for $\left(A_0,B_0,\Lambda_0\right)$ to be locally identifiable is that J, evaluated at Λ_0 has full column rank. A necessary condition, however, is that the number of row 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 (21) is of dimension $\left(2g^2\times\left[2g^2+\frac{1}{2}g\left(g+1\right)\right]\right)$, indicating the necessity of including at least $\frac{1}{2}g\left(g+1\right)$ restrictions.

If one is not interested in including restrictions on the covariance matrix Λ, i.e. $R_\Lambda=0$, the last block-column in (21) is always of full column rank (D_g is of full column rank) and the identification problem reduces to the first two block-column, i.e.

$$J\left(A,B\right)=\begin{pmatrix}
-2N_g\left[C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\otimes C_1^{-1}\right] & 2N_g\left[C_1^{-1}B\Omega_1\otimes C_1^{-1}\right]\\
-2N_g\left[C_2^{-1}B\Omega_2B'C_2^{-1\prime}D_2\otimes C_2^{-1}\right] & 2N_g\left[C_2^{-1}B\Omega_2\otimes C_2^{-1}\right]\\
R_A & R_B
\end{pmatrix} \tag{22}$$

that needs to be of full column rank.

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 ($D_2=0$), which is the case investigated in Rigobon (2003). The previous matrix reduces to:

$$J\left(A,B\right)=\begin{pmatrix}
-2N_g\left[C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\otimes C_1^{-1}\right] & 2N_g\left[C_1^{-1}B\Omega_1\otimes C_1^{-1}\right]\\
0 & 2N_g\left[C_2^{-1}B\Omega_2\otimes C_2^{-1}\right]\\
R_A & R_B
\end{pmatrix}. \tag{23}$$

For simplicity of notation, let define the following non-singular matrices

$$\begin{aligned}
D_1^* &= \left[C_1^{-1}B\Omega_1B'C_1^{-1\prime}D_1\otimes C_1^{-1}\right]\\
E_1^* &= \left[C_1^{-1}B\Omega_1\otimes C_1^{-1}\right]\\
E_2^* &= \left[C_2^{-1}B\Omega_2\otimes C_2^{-1}\right]
\end{aligned}$$

[3]See as Magnus and Neudecker (2007), pag 54. When $m=n$ it is often written K_n instead of K_{nn}.

that allows us to rewrite the J matrix (23) as

$$J(A,B) = \begin{pmatrix} -2N_g D_1^* & 2N_g E_1^* \\ 0 & 2N_g E_2^* \\ R_A & R_B \end{pmatrix} \tag{24}$$

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

$$J^*(A,B) = \begin{pmatrix} -2N_g D_1^* & 2N_g E_1^* \\ 0 & 2N_g E_2^* \\ R_A & R_B \end{pmatrix} \begin{pmatrix} -D_1^{*-1} & 0 \\ 0 & E_2^{*-1} \end{pmatrix} = \begin{pmatrix} N_g & 2N_g E_1^* E_2^{*-1} \\ 0 & 2N_g \\ -R_A D_1^{*-1} & R_B E_2^{*-1} \end{pmatrix}. \tag{25}$$

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 $2g^2$ unknowns

$$\begin{bmatrix} N_g & 2N_g E_1^* E_2^{*-1} \\ 0 & 2N_g \\ -R_A D_1^{*-1} & R_B E_2^{*-1} \end{bmatrix} x = [0] \tag{26}$$

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

$$\begin{cases} -N_g x_1 + N_g E_1^* E_2^{*-1} x_2 &= 0 \\ N_g x_2 &= 0 \\ -R_A D_1^{*-1} x_1 + R_B E_2^{*-1} x_2 &= 0 \end{cases} \tag{27}$$

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

$$x_2 = \widetilde{D}_g q_2$$

where $\widetilde{D}_g$, defined in Magnus (1988), is a $g^2 \times g(g-1)/2$ full column rank matrix and q_2 is a $g(g-1)/2$ vector of free elements. Substituting the second matrix equation into the first, the system becomes

$$\begin{cases} -N_g \left(x_1 - E_1^* E_2^{*-1} \widetilde{D}_g q_2\right) &= 0 \\ \widetilde{D}_g q_2 &= x_2 \\ -R_A D_1^{*-1} x_1 + R_B E_2^{*-1} x_2 &= 0 \end{cases} \tag{28}$$

The first equation, thus, can be solved as before

$$x_1 - E_1^* E_2^{*-1} \widetilde{D}_g q_2 = \widetilde{D}_g q_1 \tag{29}$$

and substituting into the third

$$\begin{cases} x_1 - E_1^* E_2^{*-1} \widetilde{D}_g q_2 &= \widetilde{D}_g q_1 \\ \widetilde{D}_g q_2 &= x_2 \\ -R_A D_1^{*-1} \left(E_1^* E_2^{*-1} \widetilde{D}_g q_2 + \widetilde{D}_g q_1\right) + R_B E_2^{*-1} \widetilde{D}_g q_2 &= 0 \end{cases} \tag{30}$$

The last equation, however, can be also written as

$$\left(R_A D_1^{*-1} E_1^* E_2^{*-1} - R_B E_2^{*-1}\right) \widetilde{D}_g q_2 + R_A D_1^{*-1} \widetilde{D}_g q_1 = 0 \tag{31}$$

which proves that, even if we do not have any economic knowledge about possible restrictions on the simultaneous relations among the endogenous variables, i.e. if $R_B = 0$, a suitable choice of R_A can be sufficient for allowing the parameter to be identifiable. In other words, the system of equations has the unique solution

$$\begin{Bmatrix} q_1 \\ q_2 \end{Bmatrix} = 0 \tag{32}$$

if and only if the following $(q \times g(g-1)/2)$ matrix

$$R_A \left[\begin{array}{cc} D_1^{*-1} E_1^* E_2^{*-1} & R_A D_1^{*-1} \end{array} \right] \widetilde{D}_g \tag{33}$$

has full column rank. This condition, of course, strongly depends on the number of states of volatility and on the variables that move from one state to another, as indicated by the D_i matrices. Actually, one might argue that in the previous example we simply moved the problem of identification from the B to the A matrix. However, without any restriction, the model is extremely rich, especially in the specification of the dynamics of structural shocks, that can be transmitted from one variable to another, either by the non diagonal elements of the covariance matrix Λ, or by the non diagonal elements of the A matrix, when the variable is in a state of high volatility. One reasonable possibility, thus, could be to restrict the A matrix to be simply diagonal. This solution provides $g(g-1)$ restrictions, that exceeds the necessary condition of $g(g-1)/2$ restrictions discussed above. However, whether these restrictions are sufficient for identification needs to be tested by using the rank condition in (33). The previous example 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 section we provide a generalization that fills these two shortcomings.

II.3 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

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{34}$$

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, than state 2, and when $t = 5$, state three. Thus, using the Hadamard product[4] $\odot$, we can reorganize the data as

$$Y^* = \left(i_s' \otimes Y\right) \odot \left(P \otimes i_g'\right) \tag{35}$$

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 (34), and where $y_t = \left(\begin{array}{cc} y_{1t} & y_{2t} \end{array} \right)'$, the Y^* matrix, becomes

$$Y^* = \begin{pmatrix} y_{11} & y_{21} & 0 & 0 & 0 & 0 \\ y_{12t} & y_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & y_{13} & y_{23} & 0 & 0 \\ 0 & 0 & y_{14} & y_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & y_{15} & y_{25} \end{pmatrix}. \tag{36}$$

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

$$\epsilon^* = \begin{pmatrix} \epsilon_{11} & \epsilon_{21} & 0 & 0 & 0 & 0 \\ \epsilon_{12t} & \epsilon_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \epsilon_{13} & \epsilon_{23} & 0 & 0 \\ 0 & 0 & \epsilon_{14} & \epsilon_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & \epsilon_{15} & \epsilon_{25} \end{pmatrix} \tag{37}$$

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

which allows us to rewrite the model as:

$$(I_s \otimes B)\, y_t^* = A^* \epsilon_t^* \tag{38}$$

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

$$(I_s \otimes B)\, y_t^* + (I_s \otimes \Gamma)\, x_t^* = A^* \epsilon_t^*, \tag{39}$$

in which x_t^* is the $(ks \times 1)$ vector of predetermined variables expressed as in (36) and Γ is the related $(g \times k)$ matrix of coefficients. The $(gs \times gs)$ A^* and D block diagonal matrices are defined as

$$A^* = (I_{gs} + (I_s \otimes A)\, D) \tag{40}$$

$$D = \begin{pmatrix} D_1 & & \\ & \ddots & \\ & & D_s \end{pmatrix} \tag{41}$$

where D_i is the diagonal $(g \times g)$ matrix describing the i-th state of volatility. More precisely, as described before, 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, $E(\epsilon_t \epsilon_t') = \Lambda$, but using the new notation that highlights the state of volatility, we obtain, for example

$$\Lambda_t = E\left(\epsilon_t^* \epsilon_t^{*\prime}\right) = \begin{pmatrix} \Lambda & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \tag{42}$$

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

$$\begin{array}{rcl} y_t^* & = & (I_s \otimes \Pi)\, x_t^* + u_t^* \\ & = & -(I_s \otimes B)^{-1} (I_s \otimes \Gamma)\, x_t^* + (I_s \otimes B)^{-1} A^* \epsilon_t^* \end{array} \tag{43}$$

with, using the same axample as before,

$$\Omega_t = E\left(u_t^* u_t^{*\prime}\right) = (I_s \otimes B)^{-1} A^* E\left(\epsilon_t^* \epsilon_t^{*\prime}\right) A^{*\prime} \left(I_s \otimes B^{-1}\right)' = \begin{pmatrix} \Omega & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \tag{44}$$

and

$$(I_s \otimes \Pi) = -(I_s \otimes B)^{-1} (I_s \otimes \Gamma) \tag{45}$$

Assumption 1 *The vectors* $\{\epsilon_t, t = 1 \ldots, T\}$ *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 *Each endogenous variable is at least once in a state of high volatility.*

Assumption 4 *The parameters B_0 and Γ_0 do not change among the different states of volatility.*

Proposition 1 *Consider the simultaneous equations model (38)-(43) under the Assumptions 1-4 previously expressed. Then $(A_0, B_0, \Gamma_0, \Lambda_0)$ is locally identified if and only if the matrix*

$$-2N_{gs}\left(\begin{array}{cc} J_{21} & -J_{22}\left[I_{g^2} - (\Pi' \otimes I_g)^{+}(\Pi' \otimes I_g)\right] \end{array}\right) \tag{46}$$

has full column rank. Where the two matrices J_{21} and J_{22} are defined as follows

$$J_{21} = \left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)(I_s \otimes B)' A^{*-1\prime} D \otimes A^{*-1}\right](H_A \otimes I_g) \tag{47}$$

$$J_{22} = \left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega) \otimes A^{*-1}\right](H_B \otimes I_g). \tag{48}$$

A necessary condition of identification is that there are at least $(s = 3)$ different states of volatility.

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

The approach we follow is based on Rothenberg (1971) and considers, as in the specific case analysed in the previous section, the system of 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. In the standard systems of equations, such a system is obtained from the following Jacobian matrix (without any restriction on the parameters):

$$J(\Gamma) = \left(\begin{array}{ccc} \Pi' \otimes I_g & I_{gk} & 0 \\ 2N_g\left[B\Omega \otimes I_g\right] & 0 & D_g. \end{array}\right) \tag{49}$$

which is of dimension $\left(kg + g^2\right) \times \left[g^2 + gk + g(g+1)/2\right]$. The corresponding system of matrix equations will have $\left(kg + g^2\right)$ equations with $\left[g^2 + gk + g(g+1)/2\right]$ unknowns. The necessary condition is clearly not satisfied, necessitating the inclusion of appropriate restrictions to render the parameters identifiable. More precisely, it is necessary to include at least $\left[g^2 + gk + g(g+1)/2\right] - \left(g^2 + gk\right) = \frac{1}{2}g(g+1)$ restrictions.

Including different levels of volatility can be an alternative strategy to increase the number of equations in the system of matrix equations. The price to pay, in our model, is to include more parameters than the 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 any restriction.

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 equations as unknowns in the system. Proposition 1 states that a minimum of three different states of volatility is necessary for making the parameters identifiable. The main result of this proposition 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 and Assumption 4, instead, are necessary to identify A_0, and (B_0, Γ) structural parameters respectively.

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 kind of combinations of high volatility states as described in the D matrix. All the technical details are discussed in the Appendix (A.1), as well as the necessary and sufficient condition for the identifiability of the model when combining *a priori* information on the parameters together with different states of volatility.

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. The following proposition defines the likelihood function and finds the score vector and the information matrix for the simultaneous equations model proposed in (39).

Proposition 2 *Consider a random sample of size T from the process defined by the simultaneous equations model (39) under the Assumptions 1-4. 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*

$$\begin{aligned} l(\theta) = & - (Tg/2)\log 2\pi + \sum_{t=1}^{T}\left|(I_g + AD_t)^{-1}\right| + T\log|B| - \frac{T}{2}|\Lambda| \\ & - \frac{1}{2}\sum_{t=1}^{T} tr\left[\left((I_s\otimes B)y_t^* + (I_s\otimes\Gamma)x_t^*\right)\left((I_s\otimes B)y_t^* + (I_s\otimes\Gamma)x_t^*\right)' A^{*-1\prime}\left(I_s\otimes\Lambda^{-1}\right)A^{*-1}\right] \end{aligned} \tag{50}$$

The information matrix $\mathcal{F}_T(\theta_0)$, determined by

$$E\left(d^2 l(\theta_0)\right) = (d\theta)'\,\mathcal{F}_T(\theta_0)\,d\theta, \tag{51}$$

is given by

$$\mathcal{F}_T(\theta_0) = H'\begin{pmatrix} \mathcal{F}_{AA} & \mathcal{F}_{A\Psi} & \mathcal{F}_{A\Lambda} \\ \mathcal{F}_{\Psi A} & \mathcal{F}_{\Psi\Psi} & \mathcal{F}_{\Psi\Lambda} \\ \mathcal{F}_{\Lambda A} & \mathcal{F}_{\Lambda\Psi} & \mathcal{F}_{\Lambda\Lambda} \end{pmatrix} H \tag{52}$$

where

$$\mathcal{F}_{AA} = 2\left(I_{gs}\otimes\Lambda^{*-1}\right)N_{gs}\left(T^*\Lambda^*\otimes I_{gs}\right) \tag{53}$$

$$\mathcal{F}_{A\Psi} = 2\left[\begin{array}{cc} \left(I_{gs}\otimes\Lambda^{*-1}\right)N_{gs}\left(T^*\Lambda^*\otimes I_{gs}\right) & 0 \end{array}\right] \tag{54}$$

$$\mathcal{F}_{A\Lambda} = \left(T^*\otimes\Lambda^{*-1}\right) \tag{55}$$

$$\begin{aligned} \mathcal{F}_{\Psi\Psi} = \begin{pmatrix} \left(A^{*-1}B^*\otimes A^{*\prime}\right) & 0 \\ 0 & I_{gk} \end{pmatrix}\left[\left(C'\otimes C\right)K_{(g+k)s} + \left(Q_t^*\otimes A^{*-1\prime}\Lambda^{*-1}A^{*-1}\right)\right] \\ \begin{pmatrix} \left(B^{*\prime}A^{*-1\prime}\otimes A^*\right) & 0 \\ 0 & I_{gk} \end{pmatrix} \end{aligned} \tag{56}$$

$$\mathcal{F}_{A\Lambda} = \begin{pmatrix} T^*\otimes\Lambda^{*-1} \\ 0 \end{pmatrix} \tag{57}$$

$$\mathcal{F}_{\Lambda\Lambda} = \left(\Lambda^*\otimes I_{gs}\right)\left(T^*\otimes\Lambda^{*-1}\right) \tag{58}$$

$$\mathcal{F}_{\Lambda A} = \mathcal{F}'_{A\Lambda} \qquad \mathcal{F}_{\Lambda\Psi} = \mathcal{F}'_{\Psi\Lambda} \qquad \mathcal{F}_{\Psi A} = \mathcal{F}'_{A\Psi} \tag{59}$$

$$H = \begin{pmatrix} \left(D\otimes A^{*-1}\right)\left(H_A\otimes I_g\right) & 0 & 0 \\ 0 & -\begin{pmatrix} \left(A^{*\prime}B^{*-1\prime}\otimes A^{*-1}\right)\left(H_B\otimes I_g\right) & 0 \\ 0 & \left(H_\Gamma\otimes I_k\right) \end{pmatrix} & 0 \\ 0 & 0 & \left(H_\Lambda\otimes I_g\right)D_g \end{pmatrix} \tag{60}$$

and where

$$B^* = \left(I_s\otimes B\right), \quad \Lambda^* = \left(I_s\otimes\Lambda\right), \quad C = \left[\begin{array}{cc} I_s\otimes B^{-1\prime} & 0 \end{array}\right] \tag{61}$$

$$T^* = (I_g \otimes T^{**}) \quad with \quad T^{**} = \begin{pmatrix} T_1 & & \\ & \ddots & \\ & & T_s \end{pmatrix} \tag{62}$$

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

$$Q_t^* = E\left(z_t^* z_t^{*\prime}\right) = E\left[\begin{pmatrix} y_t^* \\ x_t^* \end{pmatrix} \begin{pmatrix} y_t^{*\prime} & x_t^{*\prime} \end{pmatrix} \right] \tag{63}$$

$$= \begin{bmatrix} (I_s \otimes \Pi) Q_{xt}^* (I_s \otimes \Pi) + \Omega & (I_s \otimes \Pi) Q_{xt}^* \\ Q_{xt}^{*\prime} (I_s \otimes \Pi)' & Q_{xt}^* \end{bmatrix} \tag{64}$$

with $Q_{xt}^* = E\left(x_t^* x_t^{*\prime}\right)$.

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

$$f'(\theta) = \frac{dl(\theta)}{dvec\theta} = \begin{pmatrix} f_A(\theta), & f_\Psi(\theta), f_\Lambda(\theta) \end{pmatrix} \tag{65}$$

where

$$f_A(\theta) = \sum_{t=1}^{T} \left(\left[vec\left(DA^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} \Lambda^{*-1} A^{*-1\prime} \right) \right]' K_{gs} (H_A \otimes I_g) - \left[vec\left((I_g + AD_t)' D_t \right) \right]' \right) \tag{66}$$

$$f_\Psi(\theta) = -\sum_{t=1}^{T} \left[vec\left(A^{*-1\prime} \Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \right) \right]' \left(H_\Psi \otimes I_{(g+k)} \right) + T \left(vec \begin{bmatrix} B^{-1\prime} & 0 \end{bmatrix} \right)' \tag{67}$$

$$f_\Lambda(\theta) = \frac{1}{2} \sum_{t=1}^{T} \left[vec\left(\Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \right) \right]' (H_\Lambda \otimes I_g) D_g - \frac{T}{2} (vec(\Lambda))' D_g \tag{68}$$

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

Using the results of Proposition 2 it becomes natural to implement the *score algorithm* in order 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):

$$\theta_{n+1} = \theta_n + \left[\mathcal{F}_T(\theta_n) \right]^{-1} f(\theta_n). \tag{69}$$

If the local identification does not require any restriction on the parameters, choosing accurately the starting values for θ, the recursive algorithm (69) 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 $\hat{\theta}$:

$$\hat{\Sigma}_\theta = \mathcal{F}\left(\hat{\theta}\right)^{-1} = \left[p \lim_{x \to 0} \frac{1}{T} \mathcal{F}_T\left(\hat{\theta}\right) \right]^{-1}. \tag{70}$$

Under the assumptions previously introduced, we obviously obtain

$$\hat{\theta} \xrightarrow{\mathcal{L}} N\left(\theta_0, \hat{\Sigma}_\theta\right) \tag{71}$$

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 as follows

$$\begin{aligned} vecA &= S_A \gamma_A + s_A \\ vecB &= S_B \gamma_B + s_B \\ vec\Gamma &= S_\Gamma \gamma_\Gamma + s_\Gamma \\ vec\Lambda &= S_\Lambda \gamma_\Lambda + s_\Lambda \end{aligned} \tag{72}$$

or in more compact form

$$\begin{pmatrix} vecA \\ vecB \\ vec\Gamma \\ vec\Lambda \end{pmatrix} = \begin{pmatrix} S_A & 0 & 0 & 0 \\ 0 & S_B & 0 & 0 \\ 0 & 0 & S_\Gamma & 0 \\ 0 & 0 & 0 & S_\Lambda \end{pmatrix} \begin{pmatrix} \gamma_A \\ \gamma_B \\ \gamma_\Gamma \\ \gamma_\Lambda \end{pmatrix} + \begin{pmatrix} s_A \\ s_B \\ s_\Gamma \\ s_\Lambda \end{pmatrix} \tag{73}$$

or equivalently

$$\theta = S\gamma + s \tag{74}$$

Using the standard chain of differentiation the score vector for the new set of parameters γ can be defined as

$$f(\gamma) = S' f(\theta) \tag{75}$$

and, taking into account that the information matrix can be also defined as

$$\mathcal{F}_T(\theta) = E\left[f(\theta) \cdot f'(\theta)\right], \tag{76}$$

considering the new vector of parameters γ, it becomes

$$\mathcal{F}_T(\gamma) = S' \mathcal{F}_T(\theta) S. \tag{77}$$

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 (73). The standard asymptotic result

$$\hat{\theta} \xrightarrow{\mathcal{L}} N\left(\theta_0, \frac{1}{T} S' \mathcal{F}_T(\hat{\gamma}) S\right) \tag{78}$$

thus applies.

IV Conclusion

In this paper we have presented a theoretical framework for identifying and estimating the parameters of a simultaneous equations model with the presence of heteroskedasticity. 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, under the assumption of normally distributed structural shocks, we develop a Full Information Maximum Likelihood approach that directly estimates the parameter of the structural form. We also provide 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.

References

Caporale, G.M., Cipollini, A. and Demetriades, P. (2002), Monetary policy and the exchange rate during the Asian crisis: Identification through heteroskedasticity, CEMFE mimeograph.

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

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

Forbes, K. and Rigobon, R. (2003), No contagion, only interdependence: measuring stock market co-mevements, textitJournal of Finance 57, 2223-2261.

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

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

Klein, R. and Vella, F. (2003), Identification and estimation of the triangular simultaneous equations model in the absence of exclusion restrictions through the presence of heteroskedasticity, unpublished manuscript.

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.

Lewbel, A. (2008), Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models, mimeo.

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.

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

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

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.

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

A Appendix

A.1 Proof of Proposition 1

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

$$\begin{aligned}
(I_s \otimes B)(I_s \otimes \Pi) + (I_s \otimes \Gamma) &= 0 && (79)\\
A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)(I_s \otimes B)' A^{*-1\prime} - (I_s \otimes \Lambda) &= 0 && (80)\\
R_A vec\, A + R_B vec\, B + R_\Gamma v(\Gamma) + R_\Lambda v(\Lambda) - r &= 0 && (81)
\end{aligned}$$

where (81) indicates possible restrictions on the parameters (A, B, Γ, Λ). The first differential becomes

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

Simple algebra allows us to rewrite the system as

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

Using the properties of the Kronecker product, the system becomes

$$\begin{aligned}
\left[\left(I_s \otimes \Pi'\right) \otimes I_{gs}\right](H_B \otimes I_g) vecdB + (H_\Gamma \otimes I_g) vecd\Gamma &= 0 && (82)\\
-\left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)(I_s \otimes B)' A^{*-1} \otimes A^{*-1}\right](H_A \otimes I_g) vecdA +&\\
\left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega) \otimes A^{*-1}\right](H_B \otimes I_g) vecdB +&\\
\left[A^{*-1} \otimes A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)\right] K_{gs}(H_B \otimes I_g) vecdB +&\\
-\left[A^{*-1} \otimes A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)\left(I_s \otimes B'\right) A^{*-1\prime} D\right] K_{gs}(H_A \otimes I_g) vecdA +&\\
-(H_\Lambda \otimes I_g) vecd\Lambda &= 0 && (83)\\
R_A vec\, dA + R_B vec\, dB + R_\Gamma dv(\Gamma) R_\Lambda dv(\Lambda) &= 0. && (84)
\end{aligned}$$

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 (83) can be simplified as:

$$\begin{aligned}
-2N_{gs}\left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega)(I_s \otimes B)' A^{*-1} \otimes A^{*-1}\right](H_A \otimes I_g) vecdA +&\\
2N_{gs}\left[A^{*-1}(I_s \otimes B)(I_s \otimes \Omega) \otimes A^{*-1}\right](H_B \otimes I_g) vecdB - (H_\Lambda \otimes I_g) vecd\Lambda &= 0 && (85)
\end{aligned}$$

with $N_{gs} = 1/2(I_{gs} + K_{gs})$, as before. From (82), (85) and (84) we obtain the Jacobian matrix

$$J(A, B, \Gamma, \Omega) = \begin{pmatrix} 0 & [(I_s \otimes \Pi') \otimes I_{gs}](H_B \otimes I_g) & (H_\Gamma \otimes I_g) & 0\\ -2N_{gs}J_{21} & 2N_{gs}J_{22} & 0 & -(H_\Lambda \otimes I_g)\\ R_A & R_B & R_\Gamma & R_\Lambda \end{pmatrix} \qquad (86)$$

or equivalently

$$J(A,B,\Gamma,\Omega)=\begin{pmatrix} 0 & \Pi'\otimes I_g & I_{gk} & 0 \\ -2N_{gs}J_{21} & 2N_{gs}J_{22} & 0 & -(H_\Lambda\otimes I_g) \\ R_A & R_B & R_\Gamma & R_\Lambda \end{pmatrix} \tag{87}$$

where

$$J_{21}=\left[A^{*-1}(I_s\otimes B)(I_s\otimes\Omega)(I_s\otimes B)'A^{*-1\prime}D\otimes A^{*-1}\right](H_A\otimes I_g) \tag{88}$$

$$J_{22}=\left[A^{*-1}(I_s\otimes B)(I_s\otimes\Omega)\otimes A^{*-1}\right](H_B\otimes I_g). \tag{89}$$

A sufficient condition for $(A_0,B_0,\Gamma_0,\Lambda_0)$ to be locally identifiable is that J, 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 $rows(J)\geq cols(J)$. Including different levels of volatility is a way to increase the number of rows in the Jacobian matrix (87). 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 $J(A,B,\Gamma,\Lambda)$ matrix. Without any further restriction on the parameters, we have

$$g^2s^2+kg\geq g^2+g^2+kg+\frac{1}{2}g(g+1)\Longleftrightarrow s\geq\sqrt{\frac{5g+1}{2g}}\Longrightarrow s\geq 3 \tag{90}$$

indicating that a minimum of three states of volatility is necessary for making the parameters identifiable.

The necessary and sufficient condition depends on the kind of combinations of high volatility states as highlighted in the D matrix. If we do not want to include restrictions as in (81), we can concentrate on the following partitioned matrix

$$J(A,B)=\begin{pmatrix} 0 & \Pi'\otimes I_g & I_{gk} & 0 \\ -2N_{gs}J_{21} & 2N_{gs}J_{22} & 0 & -(H_\Lambda\otimes I_g) \end{pmatrix} \tag{91}$$

and verify for the full column rank condition. If the necessary *order condition* is satisfied, we can focus on the column rank of the sub matrix

$$J^*(A,B)=\begin{pmatrix} 0 & \Pi'\otimes I_g \\ -2N_{gs}J_{21} & 2N_{gs}J_{22} \end{pmatrix} \tag{92}$$

where J_{21} and J_{22} are defined as before. To study the column rank of the $J^*(A,B)$ we can concentrate on the following system and check whether it admits only the null vector $(x_1',x_2')'$ as possible solution

$$\begin{aligned} \left(\Pi'\otimes I_g\right)x_2 &= 0 && (93)\\ -2N_{gs}J_{21}x_1+2N_{gs}J_{22}x_2 &= 0. && (94)\end{aligned}$$

The first matrix equation is an homogeneous equation that admits solutions as

$$x_2=\left[I_{g^2}-\left(\Pi'\otimes I_g\right)^+\left(\Pi'\otimes I_g\right)\right]q_2 \tag{95}$$

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

$$\begin{aligned} \left[I_{g^2}-\left(\Pi'\otimes I_g\right)^+\left(\Pi'\otimes I_g\right)\right] &= x_2 && (96)\\ -2N_{gs}J_{21}x_1+2N_{gs}J_{22}\left[I_{g^2}-\left(\Pi'\otimes I_g\right)^+\left(\Pi'\otimes I_g\right)\right]q_2 &= 0. && (97)\end{aligned}$$

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

$$-2N_{gs}\begin{pmatrix} J_{21} & -J_{22}\left[I_{g^2}-(\Pi'\otimes I_g)^+(\Pi'\otimes I_g)\right] \end{pmatrix} \tag{98}$$

has full column rank. This condition, of course, can be easily verified numerically and represents a necessary and sufficient condition for identifiability of the parameters of the structural form. Interestingly, if the $rank\left(\Pi'\right) = k \geq g$, then $\left(\Pi' \otimes I_g\right)^{+}\left(\Pi' \otimes I_g\right) = I_{g^2}$ and the necessary and sufficient condition for J^* to have full column rank reduces to check whether

$$-2N_{gs}J_{21} = -2N_{gs}\left[A^{*-1}\left(I_s \otimes B\right)\left(I_s \otimes \Omega\right)\left(I_s \otimes B\right)' A^{*-1\prime} D \otimes A^{*-1}\right]\left(H_A \otimes I_g\right) \tag{99}$$

has full column rank.

□

A.2 Proof of Proposition 2

The log-likelihood function (50) can also be written as

$$\begin{aligned} l\left(\theta\right) = \quad const \quad & + \sum_{t=1}^{T}\left|\left(I_g + AD_t\right)^{-1}\right| + T\log\left|B\right| - \frac{T}{2}\left|\Lambda\right| \\ - \quad & \frac{1}{2}\sum_{t=1}^{T} tr\left[\Psi^* z_t^* z_t^{*\prime}\Psi^{*\prime} A^{*-1\prime}\left(I_s \otimes \Lambda^{-1}\right) A^{*-1}\right] \end{aligned} \tag{100}$$

where

$$z_t^* = \begin{pmatrix} y_t^* \\ x_t^* \end{pmatrix} \quad , \quad \Psi^* = \begin{pmatrix} I_s \otimes B & I_s \otimes \Gamma \end{pmatrix} . \tag{101}$$

The first differential is

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

and the second differential is

$$
\begin{aligned}
d^2 l(\theta) = \; & + \sum_{t=1}^{T} tr\left((I_g + AD_t)^{-1} dAD_t (I_g + AD_t)^{-1} dAD_t \right) + \\
& - Ttr\left(B^{-1} dBB^{-1} dB \right) + \frac{T}{2} tr\left(\Lambda^{-1} d\Lambda \Lambda^{-1} d\Lambda \right) + \\
& - \sum_{t=1}^{T} tr\left(\Lambda^{*-1} (I_s \otimes d\Lambda) \Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} D \left(I_s \otimes dA' \right) A^{*-1\prime} \right) + \\
& - \sum_{t=1}^{T} tr\left(\Lambda^{*-1} A^{*-1} (I_s \times dA) DA^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} D \left(I_s \otimes dA' \right) A^{*-1\prime} \right) + \\
& + \sum_{t=1}^{T} tr\left(\Lambda^{*-1} A^{*-1} d\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} D \left(I_s \otimes dA' \right) A^{*-1\prime} \right) + \\
& + \sum_{t=1}^{T} tr\left(\Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} d\Psi^{*\prime} A^{*-1\prime} D \left(I_s \otimes dA' \right) A^{*-1\prime} \right) + \\
& - 2 \sum_{t=1}^{T} tr\left(\Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} d\Psi^{*\prime} A^{*-1\prime} D \left(I_s \otimes dA' \right) A^{*-1\prime} D \left(I_s \otimes dA' \right) A^{*-1\prime} \right) + \\
& - \frac{1}{2} \sum_{t=1}^{T} tr\left(A^{*-1} (I_s \otimes dA) DA^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} (I_s \otimes d\Lambda) \Lambda^{*-1} \right) + \\
& + \frac{1}{2} \sum_{t=1}^{T} tr\left(A^{*-1} d\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} (I_s \otimes d\Lambda) \Lambda^{*-1} \right) + \\
& + \frac{1}{2} \sum_{t=1}^{T} tr\left(A^{*-1} \Psi^* z_t^* z_t^{*\prime} d\Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} (I_s \otimes d\Lambda) \Lambda^{*-1} \right) + \\
& - \frac{1}{2} \sum_{t=1}^{T} tr\left(A^{*-1} \Psi^* z_t^* z_t^{*\prime} d\Psi^{*\prime} A^{*-1\prime} D \left(I_s \otimes dA' \right) A^{*-1\prime} \Lambda^{*-1} (I_s \otimes d\Lambda) \Lambda^{*-1} \right) + \\
& - \sum_{t=1}^{T} tr\left(A^{*-1} \Psi^* z_t^* z_t^{*\prime} Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} (I_s \otimes d\Lambda) \Lambda^{*-1} \Lambda^{*-1} (I_s \otimes d\Lambda) \Lambda^{*-1} \right) + \\
& - \sum_{t=1}^{T} tr\left(d\Psi^* z_t^* z_t^{*\prime} d\Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} A^{*-1} \right) + \\
& + \sum_{t=1}^{T} tr\left(d\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} D \left(I_s \otimes dA' \right) A^{*-1\prime} \Lambda^{*-1} A^{*-1} \right) + \\
& + \sum_{t=1}^{T} tr\left(d\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} (I_s \otimes d\Lambda) \Lambda^{*-1} A^{*-1} \right) + \\
& + \sum_{t=1}^{T} tr\left(d\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} A^{*-1} (I_s \otimes dA) DA^{*-1} \right) +
\end{aligned}
\tag{102}
$$

After some algebra, it becomes

$$\begin{aligned}
d^2 l(\theta) &= vec(I_s \otimes dA)' \left[DA^{*-1} \otimes A^{*-1\prime} T^* D \right] K_{gs} vec(I_s \otimes dA) - T vec(dB)' \left[B^{-1} \otimes B^{-1\prime} \right] K_g vec(dB) \\
&+ \frac{1}{2} (dv(\Lambda))' D_g' \left[\Lambda^{-1} \otimes \Lambda^{-1} \right] D_g (dv(\Lambda)) \\
&- \sum_{t=1}^{T} vec(dA)' (H_A \otimes I_g)' \left[D' A^{*-1} \Psi^* z_t^* z_t^{*\prime} Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1} \right] (H_\Lambda \otimes I_g) D_g (dv(\Lambda)) \\
&- \sum_{t=1}^{T} vec(dA)' (H_A \otimes I_g)' \left[D' A^{*-1} \Psi^* z_t^* z_t^{*\prime} Psi^{*\prime} A^{*-1\prime} D \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1} \right] (H_A \otimes I_g) vec(dA) \\
&+ \sum_{t=1}^{T} vec(dA)' (H_A \otimes I_g)' \left[D' A^{*-1} \Psi^* z_t^* z_t^{*\prime} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1} \right] (H_\Psi \otimes I_{gk}) vec(d\Psi) \\
&+ \sum_{t=1}^{T} vec(dA)' (H_A \otimes I_g)' \left[D' A^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \right] K_{(g+k)s} \left(H_\Psi \otimes I_{(g+k)} \right) vec(d\Psi) \\
&- 2 \sum_{t=1}^{T} vec(dA)' (H_A \otimes I_g)' \left[D' A^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} D \right] K_{gs} (H_A \otimes I_g) vec(dA) \\
&+ \sum_{t=1}^{T} vec(d\Psi)' \left(H_\Psi \otimes I_{(g+k)} \right)' \left[z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} \otimes \Lambda^{*-1} \right] (H_\Lambda \otimes I_g) D_g (dv(\Lambda)) \\
&- \sum_{t=1}^{T} (dv(\Lambda))' D_g' (H_\Lambda \otimes I_g)' \left[\Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime} \Lambda^{*-1} \otimes \Lambda^{*-1} \right] (H_\Lambda \otimes I_g) D_g (dv(\Lambda)) \\
&- \sum_{t=1}^{T} vec(d\Psi)' \left(H_\Psi \otimes I_{(g+k)} \right)' \left[z_t^* z_t^{*\prime} \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1} \right] (H_\Psi \otimes I_{gk}) vec(d\Psi).
\end{aligned} \tag{103}$$

The diagonal T^* matrix, of dimension $(gs \times gs)$, is defined as

$$T^* = \begin{pmatrix}
g\left\{\begin{matrix} T_1 & & \\ & T_1 & \\ & & T_1 \end{matrix}\right. & & & & \\
& \ddots & & & \\
& & & g\left\{\begin{matrix} T_s & & \\ & T_s & \\ & & T_s \end{matrix}\right.
\end{pmatrix} \tag{104}$$

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

$$\begin{aligned}
& E(\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime}) = A^* E(\epsilon_t^* \epsilon_t^{*\prime}) A^{*\prime} = A^* \Lambda_t^* A^{*\prime} \\
\Rightarrow \quad & \textstyle\sum_{t=1}^{T} E(\Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime}) = A^* \sum_{t=1}^{T} E(\epsilon_t^* \epsilon_t^{*\prime}) A^{*\prime} = A^* T^* (I_* \otimes \Lambda) A^{*\prime} = A^* T^* \Lambda^* A^{*\prime}
\end{aligned} \tag{105}$$

and, with some algebra

$$\begin{aligned}
& E(\Psi^* z_t^* z_t^{*\prime}) = A^* \Lambda_t^* A^{*\prime} \begin{bmatrix} I_s \otimes B^{-1\prime} & 0 \end{bmatrix} = A^* \Lambda_t^* A^{*\prime} C \\
\Rightarrow \quad & \textstyle\sum_{t=1}^{T} E(\Psi^* z_t^* z_t^{*\prime}) = A^* T^* (I_* \otimes \Lambda) A^{*\prime} C = A^* T^* \Lambda^* A^{*\prime} C
\end{aligned} \tag{106}$$

The expected value of the second differential thus becomes

$$
\begin{aligned}
-E\left(d^2 l\left(\theta\right)\right) &= vec\left(I_s \otimes dA\right)' \left[DA^{*-1} \otimes A^{*-1\prime} T^* D\right] K_{gs} vec\left(I_s \otimes dA\right) \\
&+ Tvec\left(dB\right)' \left[B^{-1} \otimes B^{-1\prime}\right] K_g vec\left(dB\right) \\
&- \frac{T}{2} \left(dv\left(\Lambda\right)\right)' D_g' \left[\Lambda^{-1} \otimes \Lambda^{-1}\right] D_g \left(dv\left(\Lambda\right)\right) \\
&+ vec\left(dA\right)' \left(H_A \otimes I_g\right)' \left[DT^* \otimes A^{*-1\prime} \Lambda^{*-1}\right] \left(H_\Lambda \otimes I_g\right) D_g \left(dv\left(\Lambda\right)\right) \\
&+ vec\left(dA\right)' \left(H_A \otimes I_g\right)' \left[DT^* \Lambda^{*-1} D \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right] \left(H_A \otimes I_g\right) vec\left(dA\right) \\
&- vec\left(dA\right)' \left(H_A \otimes I_g\right)' \left[DT^* \Lambda^{*-1} A^* C \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right] \left(H_\Psi \otimes I_{gk}\right) vec\left(d\Psi\right) \\
&- vec\left(dA\right)' \left(H_A \otimes I_g\right)' \left[DA^{*-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(d\Psi\right) \\
&+ 2\left(dA\right)' \left(H_A \otimes I_g\right)' \left[DA^{*-1} \otimes A^{*-1\prime} \Lambda^{*-1} T^* \Lambda^* D\right] K_{gs} \left(H_A \otimes I_g\right) vec\left(dA\right) \\
&- vec\left(d\Psi\right)' \left(H_\Psi \otimes I_{(g+k)}\right)' \left[C' A^* \Lambda^* T^* \otimes A^{*-1\prime} \Lambda^{*-1}\right] \left(H_\Lambda \otimes I_g\right) D_g \left(dv\left(\Lambda\right)\right) \\
&+ \left(dv\left(\Lambda\right)\right)' D_g' \left(H_\Lambda \otimes I_g\right)' \left[\Lambda^{*-1} T^* \otimes \Lambda^{*-1}\right] \left(H_\Lambda \otimes I_g\right) D_g \left(dv\left(\Lambda\right)\right) \\
&+ vec\left(d\Psi\right)' \left(H_\Psi \otimes I_{(g+k)}\right)' \left[Q^* \otimes A^{*-1\prime} \Lambda^{*-1} A^{*-1}\right] \left(H_\Psi \otimes I_{gk}\right) vec\left(d\Psi\right)
\end{aligned} \tag{107}
$$

Finally, since $d\left(\theta\right) = \left(dvecA', dvecB', dvec\Gamma', dv\left(\Lambda\right)'\right)'$, 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 (102) as follows

$$
\begin{aligned}
dl\left(\theta\right) = &- \sum_{t=1}^{T} vec\left(\left(I_g + AD_t\right)' D_t\right)' vecdA + Tvec\left(B^{-1\prime}\right)' vecdB - \frac{1}{2} vec\left(\Lambda^{-1}\right)' D_g dv\left(\Lambda\right) \\
&- \sum_{t=1}^{T} vec\left(DA^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} \Lambda^{*-1} A^{*-1\prime}\right)' K_{gs} \left(H_A \otimes I_g\right) vecdA \\
&+ \frac{1}{2} \sum_{t=1}^{T} vec\left(\Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime} \Psi^{*\prime} A^{*-1\prime}\right)' \left(H_\Lambda \otimes I_g\right) D_g dv\left(\Lambda\right) \\
&- \sum_{t=1}^{T} vec\left(A^{*-1\prime} \Lambda^{*-1} A^{*-1} \Psi^* z_t^* z_t^{*\prime}\right)' \left(H_\Psi \otimes I_{(g+k)}\right) vecd\Psi
\end{aligned} \tag{108}
$$

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

$$
f'\left(\theta\right) = \frac{dl\left(\theta\right)}{dvec\theta} \tag{109}
$$

with simple algebra the result follows.

□