AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Identification and estimation of non-Gaussian structural vector autoregressions

Markku Lanne, Mika Meitz and Pentti Saikkonen

CREATES Research Paper 2015-16

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Identification and estimation of non-Gaussian structural vector autoregressions*

Markku Lanne
University of Helsinki
and CREATES

Mika Meitz
University of Helsinki

Pentti Saikkonen
University of Helsinki

March 30, 2015

Abstract

Conventional structural vector autoregressive (SVAR) models with Gaussian errors are not identified, and additional identifying restrictions are typically imposed in applied work. We show that the Gaussian case is an exception in that a SVAR model whose error vector consists of independent non-Gaussian components is, without any additional restrictions, identified and leads to (essentially) unique impulse responses. We also introduce an identification scheme under which the maximum likelihood estimator of the non-Gaussian SVAR model is consistent and asymptotically normally distributed. As a consequence, additional economic identifying restrictions can be tested. In an empirical application, we find a negative impact of a contractionary monetary policy shock on financial markets, and clearly reject the commonly employed recursive identifying restrictions.

Keywords: Structural vector autoregressive model, identification, impulse responses, non-Gaussianity

JEL Classification: C32

*The authors thank the Academy of Finland, Finnish Cultural Foundation, and Yrjö Jahnsson Foundation for financial support. The first author also acknowledges financial support from CREATES (DNRF78) funded by the Danish National Research Foundation. Contact addresses: Markku Lanne, Department of Political and Economic Studies, University of Helsinki, P. O. Box 17, FI–00014 University of Helsinki, Finland; e-mail: markku.lanne@helsinki.fi. Mika Meitz, Department of Political and Economic Studies, University of Helsinki, P. O. Box 17, FI–00014 University of Helsinki, Finland; e-mail: mika.meitz@helsinki.fi. Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FI–00014 University of Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

1 Introduction

Vector autoregressive (VAR) models are widely employed in empirical macroeconomic research, and they have also found applications in other fields of economics and finance. While the reduced-form VAR model can be seen as a convenient description of the joint dynamics of a number of time series that also facilitates forecasting, the structural VAR (SVAR) model is more appropriate for answering economic questions of theoretical and practical interest. The main tools in analyzing the dynamics in SVAR models are the impulse response function and the forecast error variance decomposition. The former traces out the future effects of an economic shock on the variables included in the model, while the latter gives the relative importance of each shock for each variable. In order to apply these tools, the economic shocks (or at least the interesting subset of them) must be identified. Traditionally short-run and long-run restrictions, constraining the immediate and permanent impact of certain shocks, respectively, have been entertained, while recently alternative approaches, including sign restrictions and identification based on heteroskedasticity, have been introduced.

When SVAR models are applied the joint distribution of the error terms is almost always (either explicitly or implicitly) assumed to have a multivariate Gaussian (normal) distribution. This means that the joint distribution of the reduced-form errors is fully determined by their covariances only (their expectation is always set to zero). A well-known consequence of this is that the structural errors cannot be identified without some additional information or restrictions. This raises the question of the potential benefit of SVAR models with non-Gaussian errors whose joint distribution is not determined by the (first and) second moments only and which may therefore contain more useful information for identification of the structural shocks.

In this paper, we show that the Gaussian case is an exception in that SVAR models with (suitably defined) non-Gaussian errors are identified without any additional identifying restrictions. In the non-Gaussian SVAR model we consider, exact identification is achieved by assuming mutual independence across the non-Gaussian error processes. We obtain an identification result that is 'statistical' in the sense that it allows the computation of (essentially) unique impulse responses, but not 'economic' in the sense that the structural shocks do not carry any economic meaning as such. For interpretation, additional information is needed to give economic labels to the structural shocks of interest. We also obtain an exact identification result that makes it possible to develop an asymptotic theory of maximum likelihood (ML) estimation. A particularly useful consequence of this is that additional identifying restrictions, such as commonly used short-run and long-run restrictions, become testable. In traditional identification approaches based on economic restrictions only, exactly-identifying restrictions cannot be tested, and finding over-identifying restrictions, or even convincing exactly identifying restrictions, may be difficult (with sign restrictions, the assessment of the validity of identifying restrictions is not testable; see, e.g., Fry and Pagan (2011)).

Compared to the previous literature on statistical identification in SVAR models, our approach is quite general. Similarly to us, Hyvärinen, Zhang, Shimizu, and Hoyer (2010) and Moneta, Entner, Hoyer, and Coad (2013) also assume independence and non-Gaussianity, but, in addition, they impose a recursive structure, which in our model only obtains as a special case. Lanne and Lütkepohl (2010) assume that the error term of their SVAR model follows a mixture of two Gaussian distributions whereas our model allows for a wide variety of (non-Gaussian) distributions. Statistical identification by explicitly modeling conditional heteroskedasticity of the errors in various forms, considered by Normandin and Phaneuf (2004), Lanne, Lütkepohl, and Maciejowska (2010), and Lütkepohl and Netšunajev (2014b), is also covered by our model. In fact, identification by unconditional heteroskedasticity (see, e.g., Rigobon (2003)) is the only approach to statistical identification entertained in the previous literature that our model does not encompass.

We apply our SVAR model to examining the impact of monetary policy in financial markets. There is a large related literature that for the most part relies on Gaussian SVAR models identified by short-run restrictions. While empirical results vary depending on the data and identification schemes, typically a monetary policy shock is not found to account for a major part of the variation of stock returns. This is counterintuitive and goes contrary to recent theoretical results (see Castelnuovo (2013) and the references therein). Our model, with the errors assumed to follow independent Student's t-distributions, is shown to fit recent U.S. data well, and we find a strong negative, yet short-lived, impact of a contractionary monetary policy shock on financial conditions, as recent macroeconomic theory predicts. Moreover, the recursive identification restrictions employed in much of the previous literature are clearly rejected.

The rest of the paper is organized as follows. In Section 2, we introduce the SVAR model. Section 3 contains the identification results. First we show how identification needed for the computation of impulse responses is achieved and then how to obtain complete identification needed in Section 4 where we develop an asymptotic estimation theory and establish the consistency and asymptotic normality of the maximum likelihood (ML) estimator of the parameters of our model (for the estimation theory, the stronger assumption of temporal independence of the error terms is made). In addition, a two-step estimator is proposed that may be useful in cases where full ML estimation is cumbersome due to short time series or the high dimension of the model. As both estimators turn out asymptotically normal, standard tests (of, e.g., additional economic identifying restrictions) can be carried out in the usual manner. An empirical application to the effect of U.S. monetary policy in financial markets is presented in Section 5, and Section 6 concludes.

Finally, a few notational conventions are given. All vectors will be treated as column vectors and, for the sake of uncluttered notation, we shall write $x = (x_1, \ldots, x_n)$ for the (column) vector x where the components x_i may be either scalars or vectors (or both). For any vector or matrix x, the Euclidean norm is denoted by $\|x\|$. The vectorization operator $vec(A)$ stacks the columns of matrix A on top of one another. Kronecker and Hadamard (elementwise) products of matrices are denoted by $\otimes$ and $\odot$, respectively. Notation $\imath_i$ is used for the ith canonical unit vector of $\mathbb{R}^n$ (i.e., an n-vector with 1 in the ith coordinate and zeros elsewhere), $i = 1, \ldots, n$ (the dimension n will be clear from the context). An identity matrix of order n will be denoted by I_n.

2 Model

Consider the structural VAR (SVAR) model

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B\varepsilon_t, \tag{1}$$

where y_t is the n-dimensional time series of interest, ν $(n \times 1)$ is an intercept term, $A_1, \ldots, A_p$ and B $(n \times n)$ are parameter matrices with B nonsingular, and ε_t $(n \times 1)$ is a temporally uncorrelated stationary error term with zero mean and finite positive definite covariance matrix (more specific assumptions about the covariance matrix will be made later). As we only consider stationary (or stable) time series, we assume

$$\det A(z) \stackrel{def}{=} \det(I_n - A_1 z - \cdots - A_p z^p) \neq 0, \quad |z| \leq 1 \;\; (z \in \mathbb{C}). \tag{2}$$

Left-multiplying (1) by the inverse of B yields an alternative formulation of the SVAR model,

$$A_0 y_t = \nu^\bullet + A_1^\bullet y_{t-1} + \cdots + A_p^\bullet y_{t-p} + \varepsilon_t, \tag{3}$$

where ε_t is as in (1), $A_0 = B^{-1}$, $\nu^\bullet = B^{-1}\nu$, and $A_j^\bullet = B^{-1}A_j$ $(j = 1, \ldots, p)$. Typically the diagonal elements of A_0 are normalized to unity, so that the model becomes a conventional simultaneous-equations model. In this paper, we shall not consider formulation (3) in detail.

The literature on SVAR models is voluminous (for a recent survey, see Kilian (2013)). A central problem with these models is the identification of the parameter matrix B: without additional assumptions or prior knowledge, B cannot be identified because, for any nonsingular $n \times n$ matrix C, the matrix B and the error term ε_t in the product $B\varepsilon_t$ can be replaced by BC and $C^{-1}\varepsilon_t$, respectively, without changing the assumptions imposed above on model (1). This identification problem has serious implications on the interpretation of the model via impulse response functions that trace out the impact of economic shocks (i.e., the components of the error term ε_t) on current and future values of the variables included in the model. Impulse responses are elements of the coefficient matrices $\Psi_j B$ in the moving average representation of the model,

$$y_t = \mu + \sum_{j=0}^{\infty} \Psi_j B \varepsilon_{t-j}, \quad \Psi_0 = I_n, \tag{4}$$

where $\mu = A(1)^{-1} \nu$ is the expectation of y_t and the matrices Ψ_j $(j = 0, 1, \ldots)$ are determined by the power series $\Psi(z) = A(z)^{-1} = \sum_{j=0}^{\infty} \Psi_j z^j$. As the preceding discussion makes clear, for a meaningful interpretation of such an analysis, an appropriate identification result is needed to make the two factors in the product $B\varepsilon_t$, and hence the impulse responses $\Psi_j B$, unique.

So far we have only made very general assumptions about the SVAR model, implying uniqueness only up to linear transformations of the form $B \mapsto BC$ and $\varepsilon_t \mapsto C^{-1}\varepsilon_t$ with C nonsingular. In SVAR models of the type (1), the covariance matrix of the error term is typically restricted to a diagonal matrix so that the transformation matrix C has to be of the form $C = DO$ with O orthogonal and D diagonal and nonsingular. The diagonal elements of D are either $+1$ or -1 if the covariance matrix of ε_t is assumed an identity matrix, while in the absence of such a normalization, the diagonal elements of D are not restricted (except to be nonzero). Thus, further assumptions are needed to achieve identifiability, and probably the most common way of achieving identification is to impose short-run restrictions that restrict some of the elements of B to zero. In the best known example of this approach, the matrix B is restricted to a lower triangular matrix which can be identified as a Cholesky factor of the covariance matrix of the error term $B\varepsilon_t$. This solves the identification problem, but it imposes a recursive structure upon the variables included in y_t that may be implausible: either recursiveness is completely absent, or present only under a different ordering of the variables. This example also illustrates what seems to be an inherent difficulty in using short-run restrictions: one basically tries to solve the identification problem by using only the covariance matrix of the error term. Nevertheless, this approach has a long history in the econometrics and time series literature. For instance, the SVAR model (1) is a special case of a simultaneous vector ARMAX model where identification results based only on knowledge of second order moments have been obtained by Kohn (1979), Hannan and Deistler (1988), and others. Similarly to these previous authors, we use the term 'class of observationally equivalent SVAR processes' to refer to SVAR processes satisfying the assumptions made of (1) with the matrix B and the error term ε_t replaced by BC and $C^{-1}\varepsilon_t$ with C a nonsingular matrix (in the same way we shall speak of classes of observationally equivalent moving average representations). Then the identification problem boils down to finding conditions which imply that the only possible choice for the matrix C is an identity matrix and thus that the matrix B and the error term ε_t are unique.

As already indicated, successful identification results may be difficult to obtain without strengthening the assumptions so far imposed on the error term ε_t. In this paper, we consider model (1) where, similarly to Hyvärinen et al. (2010) and Moneta et al. (2013), the components of the error term are assumed contemporaneously independent. However, for identification, the errors need not be independent in time, but only serially uncorrelated, which covers conditional heteroskedasticity recently used to achieve identifiability in SVAR models (see Lütkepohl and Netšunajev (2014b) and the references therein).

3 Identification

3.1 Non-Gaussian errors

As already mentioned, we assume that the components of the error term $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})$ are mutually independent, non-Gaussian, and uncorrelated in time. Specifically, we make the following assumption.

Assumption 1.

(i) The error process $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})$ is a sequence of (strictly) stationary random vectors with each component $\varepsilon_{i,t}$, $i = 1, \ldots, n$, having zero mean and finite positive variance.

(ii) The component processes $\varepsilon_{i,t}$, $i = 1, \ldots, n$, are mutually independent and at most one of them has a Gaussian marginal distribution.

(iii) For all $i = 1, \ldots, n$, the components $\varepsilon_{i,t}$ are uncorrelated in time, that is, $Cov\left[\varepsilon_{i,t}, \varepsilon_{i,t+k}\right] = 0$ for all $k \neq 0$.

The conditions imposed in Assumption 1(i) are rather standard. Although it might be possible to weaken them (to require only covariance stationarity, or to allow for infinite variances) we do not pursue this matter. Assumption 1(ii) restricts the interdependence of the components of the error process. Note that complete statistical independence of the n component processes $\{\varepsilon_{i,t},\ t \in \mathbb{Z}\}$, $i = 1, \ldots, n$, is assumed; this implies, for instance, the (weaker) condition that $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$ are independent for all $i \neq j$. The vector process ε_t is assumed non-Gaussian, but the possibility that (at most) one of its components is Gaussian is permitted. Note that in this non-Gaussian case, independence is a much stronger requirement than mere uncorrelatedness (in the Gaussian case, independence and uncorrelatedness are equivalent, but here the distinction is essential). The importance of requiring independence and not only uncorrelatedness in SVAR models has recently been stressed also by Gouriéroux and Monfort (2014, Sec. 3).

The last part of Assumption 1 restricts the temporal dependence of each individual component process. Assuming only uncorrelatedness (and thus not necessarily independence) has the convenience that conditionally heteroskedastic errors are also covered (for instance, the component error processes can follow conventional GARCH processes which, with appropriate parameter restrictions, are stationary with finite second moments and necessarily non-Gaussian, so that Assumptions 1(i) and (ii) apply).

Compared with assumptions made in previous literature, Assumption 1 is similar to its counterparts in Hyvärinen et al. (2010) and Moneta et al. (2013) except that (at least) the former authors also assume independence of the errors over time.

3.2 Identification up to permutations and scalings

In this section, we explain how non-Gaussianity aids in solving the identification problem discussed in Section 2. As impulse response analysis constitutes a major application of the SVAR model, we consider the identification of the moving average representation (4). Under Assumption 1, this representation is essentially unique in the following sense (the subsequent arguments will be formalized and proved in Proposition 1 below): If the process y_t can be represented by two (potentially) different moving average representations, say,

$$y_t = \mu + \sum_{j=0}^{\infty} \Psi_j B \varepsilon_{t-j} = \mu^* + \sum_{j=0}^{\infty} \Psi_j^* B^* \varepsilon_{t-j}^*, \tag{5}$$

then necessarily $\mu^* = \mu$, $\Psi_j^* = \Psi_j$ $(j = 0, 1, \ldots)$, and $B\varepsilon_t = B^*\varepsilon_t^*$ for all t, but the choice of the matrix B and the error process ε_t is not unique: As discussed in Section 2, the choice $B^* = BC$

and $\varepsilon_t^* = C^{-1}\varepsilon_t$ will do for any nonsingular $n \times n$ matrix C. This holds under Assumption 1(i) and Assumption 1(iii) strengthened to $Cov[\varepsilon_t, \varepsilon_{t+k}] = 0$ for all $k \neq 0$. In the conventional Gaussian set-up, Assumption 1(ii) is not imposed directly, but independence of the component processes obtains because ε_t is assumed to be independent and identically normally distributed with mean zero and a diagonal covariance matrix. In this Gaussian set-up, the discussion in Section 2 applies and the aforementioned (nonsingular) matrix C is of the form $C = DO$ with O orthogonal and D diagonal, so that an identification problem remains. However, assuming non-Gaussianity and independence (in the sense of Assumption 1(ii)) we can restrict the orthogonal matrix O in the product $C = DO$ to a permutation matrix so that only permutations and scale changes in the columns of B are allowed. This constitutes a considerable improvement and forms the first step in achieving complete identification which is the topic of the next subsection.

The preceding discussion is formalized in the following proposition, whose proof is given in Appendix A.[1]

Proposition 1. *Consider the SVAR model (1) and assume that the stationarity condition (2) and Assumption 1 on the error term* ε_t *are satisfied. Suppose the two moving average representations in (5) hold true*

(i) for some parameters μ^* $(n \times 1)$ *and* B^* $(n \times n)$ *with* B^* *nonsingular,*

(ii) for some coefficient matrices Ψ_j^* $(n \times n)$*,* $j = 0, 1, \ldots,$ *that are determined by the power series* $\Psi^*(z) = A^*(z)^{-1} = \sum_{j=0}^{\infty} \Psi_j^* z^j$ *with* $A^*(z) = I_n - A_1^* z - \cdots - A_p^* z^p$ *satisfying condition (2) (with* A_j *therein replaced by* A_j^**,* $j = 1, \ldots, p$*), and*

(iii) for some error process $\varepsilon_t^* = (\varepsilon_{1,t}^*, \ldots, \varepsilon_{n,t}^*)$ *satisfying Assumption 1 (with each '*ε*' therein replaced by '*ε^**').*

Then, for some diagonal matrix $D = diag(d_1, \ldots, d_n)$ *with nonzero diagonal elements, for some permutation matrix* P $(n \times n)$*, and for all* t*,*

$$B^* = BDP, \quad \varepsilon_t^* = P'D^{-1}\varepsilon_t, \quad \mu^* = \mu, \quad \textit{and} \quad \Psi_j^* = \Psi_j \quad (j = 0, 1, \ldots). \tag{6}$$

Variants of Proposition 1 have appeared in previous literature. For instance, in the independent component literature, reference can be made to Theorem 11 and its corollaries in Comon

[1] This proposition can be specialized to formulation (3) by setting $B = A_0^{-1}$, $\nu = A_0^{-1}\nu^\bullet$, and $A_j = A_0^{-1}A_j^\bullet$ $(j = 1, \ldots, p)$ in model (1).

(1994) that are very similar, although formulated for the case corresponding to a serially uncorrelated process, i.e., $y_t = \nu + B\varepsilon$. Another related result in the statistics literature is Theorem 4 of Chan and Ho (2004) (a discussion of this theorem can also be found in Chan, Ho, and Tong (2006)). Unlike in Assumption 1, Chan and Ho (2004) assume that the components of ε_t are independent in time, but instead of the specific moving average representation (4) they allow for a general linear process with a two-sided moving average representation. Recently, also Gouriéroux and Zakoïan (2014, Proposition 7) and Gouriéroux and Monfort (2014, Proposition 2) have presented counterparts of Proposition 1 assuming independence of ε_t over time.

Proposition 1 does not provide a complete solution to the identification problem. It only shows that the moving average representation (4) and its SVAR counterpart (1) are unique apart from permutations and scalings of the columns of B and the components of ε_t; uniqueness of the expectation μ and the coefficients Ψ_j, $j = 0, 1, \ldots$, or, equivalently, the intercept term ν and the autoregressive parameters $A_1, \ldots, A_p$ obtains, however. Using the terminology introduced in Section 2, Proposition 1 characterizes a class of observationally equivalent SVAR processes and the corresponding moving average representations: The moving average representations in (5) are observationally equivalent (and hence members of this class) if they satisfy the equations in (6). The same, of course, applies to the corresponding SVAR processes, i.e., (1) and $y_t = \nu^* + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + B^* \varepsilon_t^*$, but now the last two equations in (6) are replaced by $\nu = \nu^*$ and $A_i = A_i^*$, $i = 1, \ldots, p$, respectively.

From the viewpoint of computing impulse responses and forecast error variance decompositions, identification up to permutations and scalings is sufficient, and this is all that is attained by means of statistical identification procedures of SVAR models put forth in the previous literature. Upon such identification of the SVAR model, labeling the shocks is in any case based on outside information, such as sign restrictions, or conventional identifying short-run or long-run restrictions (see Lütkepohl and Netšunajev (2014a)), and the sign and size of the shocks are chosen by the researcher. For these purposes, any permutation and scaling are equally useful. However, development of conventional statistical estimation theory, in particular, calls for a complete solution to the identification problem.

3.3 Complete identification

In this section, we provide formal identifying or normalizing restrictions that remove the indeterminacy due to scaling and permutation in Proposition 1. One set of such conditions, employed in the context of independent component analysis, can be found in Ilmonen and Paindaveine (2011) (see also Hallin and Mehta (in press)); for potential alternative conditions, see, e.g., Pham and Garat (1997) and Chen and Bickel (2005). In the case of Proposition 1 these conditions are specified as follows.

To express the result, let $\mathcal{M}_n$ denote the set of nonsingular $n \times n$ matrices. We say that two matrices B_1 and B_2 in $\mathcal{M}_n$ are equivalent, expressed as $B_1 \sim B_2$, if and only if they are related as $B_2 = B_1 DP$ for some diagonal matrix $D = diag\,(d_1, \ldots, d_n)$ with nonzero diagonal elements and some permutation matrix P.[2] The equivalence relation $\sim$ partitions $\mathcal{M}_n$ into equivalence classes, and each of these equivalence classes defines a set of observationally equivalent SVAR processes. Using this terminology, Proposition 1 and the discussion following it state that while a specific equivalence class for B is identifiable, any member from this equivalence class can be used as a B and also used to define a member from the corresponding set of observationally equivalent SVAR processes. Our next aim is to pinpoint a particular (unique) member from the equivalence class indicated by Proposition 1. We collect the description of how this can be done in the following 'Identification Scheme' (whose content is adapted from Ilmonen and Paindaveine (2011) and Hallin and Mehta (in press)).

Identification Scheme. *For each $B \in \mathcal{M}_n$, consider the sequence of transformations*

$$B \to BD_1 \to BD_1P \to BD_1PD_2,$$

where, whenever such $n \times n$ matrices D_1, P, and D_2 exist,

(i) D_1 is the positive definite diagonal matrix that makes each column of BD_1 have Euclidean norm one,

(ii) P is the permutation matrix for which the matrix $C = (c_{ij}) = BD_1P$ satisfies $|c_{ii}| > |c_{ij}|$ for all $i < j$, and

(iii) D_2 is the diagonal matrix such that all diagonal elements of BD_1PD_2 are equal to one.

Let $\mathcal{I} \subseteq \mathcal{M}_n$ be the set consisting of those $B \in \mathcal{M}_n$ for which the matrices D_1, P, and D_2 above exist, and $\mathcal{E} = \mathcal{M}_n \backslash \mathcal{I}$ the complement of this set in $\mathcal{M}_n$.[3] *Define the transformation*

[2] Note that $DP = PD_1$ for some scaling matrix D_1 so that the order of the permutation and scaling matrix does not matter for the defined equivalence; from this fact it can also be seen that the relation $B_1 \sim B_2$ is transitive and, as it is clearly symmetric and reflexive, it really is an equivalence relation.

[3] That is, $\mathcal{E}$ is the set of those matrices $B \in \mathcal{M}_n$ for which a tie occurs in step (ii) in the sense that for any choice of P we have $|c_{ii}| = |c_{ij}|$ for some $i < j$, or for which at least one diagonal element of BD_1P equals zero so that step (iii) cannot be done.

$\Pi(\cdot) : \mathcal{I} \to \mathcal{I}$ *as* $\Pi(B) = BD_1PD_2$ *with* D_1*,* P*, and* D_2 *as above*[4]*, and define the set* $\mathcal{B}$ *as*

$$\mathcal{B} = \Pi(\mathcal{I}) = \{\tilde{B} \in \mathcal{M}_n : \tilde{B} = \Pi(B) \text{ for some } B \in \mathcal{I}\}.$$

This scheme provides a recipe for picking a particular permutation and a particular scaling to identify a unique matrix B from each equivalence class corresponding to observationally equivalent SVAR processes. Therefore, the scheme provides a solution to the identification problem in the sense formalized in the following proposition (which is justified in Appendix A).

Proposition 2.

(a) Under the assumptions of Proposition 1, the matrix B *is uniquely identified in the set* $\mathcal{B}$ *defined in the Identification Scheme.*[5]

(b) The set $\mathcal{B}$ *consists of unique, distinct representatives from each* $\sim$*-equivalence class of* $\mathcal{I}$*.*

(c) The set $\mathcal{E}$ *(of matrices being excluded in the Identification Scheme) has Lebesgue measure zero in* $\mathbb{R}^{n\times n}$*, and the set* $\mathcal{I}$ *(of matrices being included in the Identification Scheme) contains an open and dense subset of* $\mathcal{M}_n$*.*

According to part (a) of Proposition 2, unique identification is achieved by restricting the permissible values of the matrix B to the set $\mathcal{B} = \Pi(\mathcal{I})$ defined in the Identification Scheme, while parts (b) and (c) of the proposition explain in further detail what exactly is achieved. According to part (b), the set $\mathcal{B}$ is suitably defined: no two observationally equivalent SVAR processes are represented in $\mathcal{B}$, while nearly all observationally non-equivalent SVAR processes are represented in $\mathcal{B}$. Part (c) explains the quantifier 'nearly all': A small number of SVAR processes, namely those corresponding to the set $\mathcal{E}$, have to be excluded from consideration, but as these processes only comprise a set of measure zero, ignoring them is hardly relevant in practice; moreover, the set $\mathcal{I}$ corresponding to those SVAR processes that are included in the Identification Scheme is 'large' in the sense that $\mathcal{I}$ contains an open and dense subset of $\mathcal{M}_n$. Some further remarks on this result are in order.

First, the set $\mathcal{E}$ having measure zero and $\mathcal{I}$ containing an open and dense subset of $\mathcal{M}_n$ indeed mean that almost all SVAR processes are being included. According to the terminology used by some authors, the matrix B would be 'generically identified' in case it were identified in this open and dense subset $\mathcal{I}$ of the parameter space of interest, $\mathcal{M}_n$; see, e.g., Anderson et

[4] The matrices D_1, P, and D_2 depend on B, but we do not make this dependence explicit.

[5] In the sense that if $B, B^* \in \mathcal{B}$ are as in Proposition 1, then necessarily $D = P = I$ in (6) so that $B = B^*$.

al. (in press) for the use of this terminology in the context of VAR models, or Johansen (1995) in a cointegrated VAR model. It is also worth noting that the excluded matrices in $\mathcal{E}$ are in no way 'ill-behaving'; their exclusion is done for purely technical reasons to make the formulation of the Identification Scheme easy (although it would be possible to devise a scheme in a way that no exlusions are needed, we do not pursue this matter, as such a scheme would be rather complex and its implementation would presumably be difficult in practice).

Second, as the preceding discussion suggests, one can similarly obtain identifiability by using some alternative formulation of the Identification Scheme. One relevant alternative is obtained if the definitions of D_1 and P in the Identification Scheme are maintained but D_2 is defined as the diagonal matrix whose diagonal elements equal either 1 or -1 and which makes the diagonal elements of BD_1PD_2 positive. The restrictions implied by this alternative identification scheme may be easier to take into account in estimation than those based on the original Identification Scheme. In our empirical application we use this alternative identification scheme, and after finding the maximum of the likelihood function, we switch to the original Identification Scheme. On the other hand, the original Identification Scheme is more convenient in deriving asymptotic distributions for estimators; in the alternative scheme just described, one would need to employ Lagrange multipliers as the columns of BD_1PD_2 would then have Euclidean norm one.

Third, as already alluded to in Section 3.2, the Identification Scheme and Proposition 2 only yield statistical identification which need not have any economic interpretation. In particular, they do not offer any information about which economic shock each component of ε_t might be. The statistical identification result obtained does, however, facilitate the development of conventional estimation theory, the topic of Section 4 below.

3.4 Discussion of previous identification results

There are a number of statistical identification procedures for SVAR models introduced in the previous literature that are more or less closely related to the procedure put forth in this paper. Hyvärinen et al. (2010) and Moneta et al. (2013) consider identification in SVAR models and, similarly to us, assume that the error terms are non-Gaussian and mutually independent. Their identification condition is explicitly stated for model (3), but it, of course, applies to model (1) as well (an analog of our Proposition 2 could also be formulated for model (3)). Compared to us, an essential difference is that they assume the matrix A_0 in model (3), or equivalently the matrix B in model (1), to be lower triangular (potentially after reordering the variables in

y_t). This is a rather stringent and potentially undesirable a priori assumption, as it imposes a recursive structure on the SVAR model. Hence, our result is more general, yet allowing for a recursive structure as a special case.

Lanne and Lütkepohl (2010) assume that the errors of model (1) are independent over time with a distribution that is a mixture of two Gaussian distributions with zero means and diagonal covariance matrices, one of which is an identity matrix and the other one has positive diagonal elements, which for identifiability have to be different from each other. Under these conditions, identifiability is obtained apart from permutations of the columns of B and multiplication by minus one. If the above-mentioned positive diagonal elements are ordered in some specific way, say from largest to smallest, the indeterminacy due to permutations of the columns of B is removed and identifiability is achieved. Thus, their identification result differs from ours mainly in that a specific non-Gaussian error distribution is employed, and its components are only contemporaneously uncorrelated, not independent.

Assuming some form of heteroskedasticity of the errors ε_t is one popular approach to identification. Lanne et al. (2010) and Lütkepohl and Netšunajev (2014b) assume Markov switching and a smooth transition in the covariance matrix of the error term ε_t in model (1), respectively, while Normandin and Phaneuf (2004) allow for GARCH-type heteroskedasticity in the errors. As already discussed, our approach also covers these cases in that the identification results hold under conditional heteroskedasticity that necessarily implies non-Gaussianity of the errors. In contrast, identification by unconditional heteroskedasticity that has also been entertained in the recent SVAR literature (see, e.g., Rigobon (2003) and Lanne and Lütkepohl (2008)) is not covered.

4 Parameter estimation

We next discuss parameter estimation in the non-Gaussian SVAR model. The first four subsections present the likelihood function, study the score vector and the Hessian matrix, and give an asymptotic normality result for the maximum likelihood (ML) estimator. The fifth subsection develops a two-step estimator useful in computationally demanding situations, and the sixth subsection discusses hypothesis testing.

4.1 Likelihood function

We consider ML estimation of the parameters in model (1), and to that end, we have to be more specific about the temporal dependence of the error term. For the estimation theory of this paper, we assume that the non-Gaussian and mutually independent error terms $\varepsilon_{i,t}$ ($i = 1, \ldots, n$) are independent also in time instead of only uncorrelated, as required in Assumption 1(iii). Specifically, we make the following assumption.

Assumption 2.

(i) Assumption 1 holds with the mutually independent error processes $\varepsilon_{i,t}$ being independent also in time and with finite and positive variances σ_i^2, $i = 1, \ldots, n$.

(ii) For each $i = 1, \ldots, n$, the distribution of the error term $\varepsilon_{i,t}$ has a (Lebesgue) density $f_{i,\sigma_i}(x; \lambda_i) = \sigma_i^{-1} f_i(\sigma_i^{-1} x; \lambda_i)$ which may also depend on a parameter vector λ_i.

Assumption 2(i) strengthens Assumption 1(iii) by requiring that each of the error processes $\varepsilon_{i,t}$ is independent and identically distributed with zero mean and variance σ_i^2 ($i = 1, \ldots, n$). Assumption 2(ii) is sufficient to construct the likelihood function of the parameters. Note that the component densities $f_i(\cdot; \lambda_i)$ are supposed to depend on their own parameter vectors, but they can (though need not) belong to the same family of densities. For instance, they can be densities of Student's t-distribution with different degrees of freedom parameters.

Next we define the parameter space of the model. First consider the parameter matrix B which we assume to belong to the set $\mathcal{B}$ introduced in the previous section. This restricts the diagonal elements of the matrix B to unity, and we collect its off-diagonal elements in the vector β ($n(n-1) \times 1$) and express this as $\beta = vecd^{\circ}(B)$ where, for any $n \times n$ matrix C, $vecd^{\circ}(C)$ signifies the $n(n-1)$–dimensional vector obtained by removing the n diagonal entries of C from its usual vectorized form $vec(C)$. Note that $vec(B(\beta)) = H\beta + vec(I_n)$, where the $n^2 \times n(n-1)$ matrix H is of full column rank and its elements consist of zeros and ones[6] (we use the notation $B(\beta)$ when we wish to make the dependence of the parameter matrix B on its unknown off-diagonal elements explicit). The parameters of the model are now contained in the vector $\theta = (\pi, \beta, \sigma, \lambda)$ where $\pi = (\pi_1, \pi_2)$ with $\pi_1 = \nu$ and $\pi_2 = vec([A_1 : \cdots : A_p])$,

[6]The matrix H can be expressed as $H = \sum_{i=1}^{n} \sum_{j=1}^{n-1} (\imath_i \imath_i' \otimes \imath_{j+I[j \geq i]} \tilde{\imath}_j')$, where $\tilde{\imath}_j$ denotes an $(n-1)$-vector with 1 in the jth coordinate and zeros elsewhere, $j = 1, \ldots, n-1$, and $I[j \geq i] = 1$ if $j \geq i$ and zero otherwise (cf. Ilmonen and Paindaveine (2011, p. 2452)).

and $\sigma = (\sigma_1, \ldots, \sigma_n)$ and $\lambda = (\lambda_1, \ldots, \lambda_n)$. We use θ_0 to signify the true parameter value (and similarly for its components) and introduce the following assumption.

Assumption 3. *The true parameter value θ_0 belongs to the permissible parameter space $\Theta = \Theta_\pi \times \Theta_\beta \times \Theta_\sigma \times \Theta_\lambda$, where (i) $\Theta_\pi = \mathbb{R}^n \times \Theta_{\pi_2}$ with $\Theta_{\pi_2} \subseteq \mathbb{R}^{n^2 p}$ such that condition (2) holds for every $\pi_2 \in \Theta_{\pi_2}$, (ii) $\Theta_\beta = vecd^\circ(\mathcal{B}) = \{\beta \in \mathbb{R}^{n(n-1)} : \beta = vecd^\circ(B) \text{ for some } B \in \mathcal{B}\}$, (iii) $\Theta_\sigma = \mathbb{R}^n_+$, and (iv) $\Theta_\lambda = \Theta_{\lambda_1} \times \cdots \times \Theta_{\lambda_n} \subseteq \mathbb{R}^d$ with $\Theta_{\lambda_i} \subseteq \mathbb{R}^{d_i}$ open for every $i = 1, \ldots, n$ and $d = d_1 + \cdots + d_n$.*

Condition (2) entails that Θ_{π_2}, the parameter space of π_2, is open whereas Θ_β is open due to the Identification Scheme and Proposition 2 (a justification is given in the Supplementary Appendix). Thus, Assumption 3 implies that the whole parameter space Θ is open so that the true parameter value θ_0 is an interior point of the parameter space, as assumed in standard derivations of the asymptotic normality of a ML estimator.

The log-likelihood function of the parameter $\theta \in \Theta$ based on model (1) and the data $y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T$ (and conditional on $y_{-p+1}, \ldots, y_0$) can now be written as

$$L_T(\theta) = T^{-1} \sum_{t=1}^{T} l_t(\theta), \tag{7}$$

where

$$l_t(\theta) = \sum_{i=1}^{n} \log f_i\left(\sigma_i^{-1} \iota_i' B(\beta)^{-1} u_t(\pi); \lambda_i\right) - \log\det(B(\beta)) - \sum_{i=1}^{n} \log \sigma_i$$

with ι_i the ith unit vector and $u_t(\pi) = y_t - \nu - A_1 y_{t-1} - \cdots - A_p y_{t-p}$. Maximizing $L_T(\theta)$ over the permissible parameter Θ space yields the ML estimate of θ.

To apply the estimator discussed above one has to choose a non-Gaussian error distribution. In economic applications departures from Gaussianity typically manifest themselves as leptokurtic behavior, and Student's t-distribution is presumably the non-Gaussian distribution most commonly employed in the previous empirical literature. Alternatives include the normal inverse Gaussian distribution, the generalized hyperbolic distribution, and their skewed versions.

4.2 Score vector

We first derive the asymptotic distribution of the score vector (evaluated at the true parameter value θ_0). We use a subscript to signify a partial derivative; for instance $l_{\theta,t}(\theta) = \partial l_t(\theta) / \partial\theta$,

$f_{i,x}(x;\lambda_i) = \partial f_i(x;\lambda_i)/\partial x$, and $f_{i,\lambda_i}(x;\lambda_i) = \partial f_i(x;\lambda_i)/\partial \lambda_i$ (an assumption which guarantees the existence of these partial derivatives will be given shortly). The score vector of a single observation, $l_{\theta,t}(\theta)$, is derived in Appendix B. When evaluated at the true parameter value, the components of $l_{\theta,t}(\theta_0) = (l_{\pi,t}(\theta_0), l_{\beta,t}(\theta_0), l_{\sigma,t}(\theta_0), l_{\lambda,t}(\theta_0))$ are

$$l_{\pi,t}(\theta_0) = -(x_{t-1} \otimes B_0^{-1\prime}\Sigma_0^{-1})e_{x,t} \tag{8a}$$

$$l_{\beta,t}(\theta_0) = -H'[(\varepsilon_t \otimes B_0^{-1\prime}\Sigma_0^{-1}e_{x,t}) + vec(B_0^{-1\prime})] \tag{8b}$$

$$l_{\sigma,t}(\theta_0) = -\Sigma_0^{-2}(\varepsilon_t \odot e_{x,t} + \sigma_0) \tag{8c}$$

$$l_{\lambda,t}(\theta_0) = e_{\lambda,t}, \tag{8d}$$

where $x_{t-1} = (1, y_{t-1}, \ldots, y_{t-p})$, $\Sigma_0 = diag(\sigma_{1,0}, \ldots, \sigma_{n,0})$, $e_{x,t} = (e_{1,x,t}, \ldots, e_{n,x,t})$, and $e_{\lambda,t} = (e_{1,\lambda_1,t}, \ldots, e_{n,\lambda_n,t})$ with

$$e_{i,x,t} = \frac{f_{i,x}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})} \quad \text{and} \quad e_{i,\lambda_i,t} = \frac{f_{i,\lambda_i}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}.$$

We also introduce the compact and convex set $\Theta_0 = \Theta_{0,\pi} \times \Theta_{0,\beta} \times \Theta_{0,\sigma} \times \Theta_{0,\lambda}$ contained in the interior of Θ that has θ_0 as an interior point. Now, we make the following assumption.

Assumption 4. *The following conditions hold for $i = 1, \ldots, n$:*

(i) For all $x \in \mathbb{R}$ and all $\lambda_i \in \Theta_{0,\lambda_i}$, $f_i(x;\lambda_i) > 0$ and $f_i(x;\lambda_i)$ is twice continuously differentiable with respect to $(x;\lambda_i)$.

(ii) The function $f_{i,x}(x;\lambda_{i,0})$ is integrable with respect to x, i.e., $\int |f_{i,x}(x;\lambda_{i,0})|\,dx < \infty$.

(iii) For all $x \in \mathbb{R}$, the functions

$$x^2 \frac{f_{i,x}^2(x;\lambda_{i,0})}{f_i^2(x;\lambda_{i,0})} \quad \text{and} \quad \frac{\|f_{i,\lambda_i}(x;\lambda_{i,0})\|^2}{f_i^2(x;\lambda_{i,0})}$$

are dominated by $c_1(1+|x|^{c_2})$ with $c_1, c_2 \geq 0$ and $\int |x|^{c_2} f_i(x;\lambda_{i,0})\,dx < \infty$.

(iv) For all $x \in \mathbb{R}$ and $\lambda_i \in \Theta_{0,\lambda_i}$, $\int \sup_{\lambda_i \in \Theta_{0,\lambda_i}} \|f_{i,\lambda_i}(x;\lambda_i)\|\,dx < \infty$.

Moreover,

(v) The matrix $E[l_{\theta,t}(\theta_0)l'_{\theta,t}(\theta_0)]$ is positive definite.

Assumption 4(i) guarantees that the log-likelihood function satisfies conventional differentiability assumptions of ML estimation by imposing differentiability assumptions on the density functions $f_i(x;\lambda_i)$. Assumptions 4(ii)–(iv) require that the partial derivatives of the density functions $f_i(x;\lambda_i)$ satisfy suitable integrability conditions that are needed to ensure that the score function (evaluated at the true parameter value) has zero mean and a finite covariance matrix. Assumption 4(v) ensures that this covariance matrix, and hence the covariance matrix of the (normal) limiting distribution of the ML estimator of θ, is positive definite. The conditions in Assumption 4 (as well as those in Assumption 5 below) are similar to those previously used in the estimation theory of noncausal and noninvertible ARMA models (see, e.g., Breidt, Davis, Lii, and Rosenblatt (1991), Andrews, Davis, and Breidt (2006), Lanne and Saikkonen (2011), Meitz and Saikkonen (2013), and the references therein), although their formulation is somewhat different. Most common density functions satisfy these assumptions.

The limiting distribution of the score vector is given in the following lemma which is proved in Appendix B.

Lemma 1. *If Assumptions 2–4 hold,* $T^{-1/2}\sum_{t=1}^{T} l_{\theta,t}(\theta_0) \xrightarrow{d} N(0,\mathcal{I}(\theta_0))$, *where* $\mathcal{I}(\theta_0) = E[l_{\theta,t}(\theta_0)\, l'_{\theta,t}(\theta_0)]$ *is positive definite.*

As shown in Appendix B, $l_{\theta,t}(\theta_0)$ is a stationary and ergodic martingale difference sequence with covariance matrix $\mathcal{I}(\theta_0)$ and, consequently, the limiting distribution can be obtained by applying a standard central limit theorem. An explicit expression of the covariance matrix $\mathcal{I}(\theta_0)$ is given in Appendix B.

4.3 Hessian matrix

We next consider the Hessian matrix. Expressions for the required second partial derivatives are given in Appendix C. Similarly to the first partial derivatives, we use notations such as $l_{\theta\theta,t}(\theta) = \partial^2 l_t(\theta)/\partial\theta\partial\theta'$, $f_{i,xx}(x,\lambda_i) = \partial^2 f_i(x;\lambda_i)/\partial x^2$, and $f_{i,x\lambda_i}(x;\lambda_{i,0}) = \partial^2 f_{i,x\lambda_i}(x;\lambda_i)/\partial x\partial\lambda_i'$. The following assumption complements Assumption 4 by providing further regularity conditions on the partial derivatives of the density functions $f_i(x;\lambda_i)$.

Assumption 5. *The following conditions hold for* $i = 1,\ldots,n$*:*

(i) The functions $f_{i,xx}(x;\lambda_{i,0})$ *and* $f_{i,x\lambda_i}(x;\lambda_{i,0})$ *are integrable with respect to* x*, i.e.,* $\int |f_{i,xx}(x;\lambda_{i,0})|\,dx < \infty$ *and* $\int \|f_{i,x\lambda_i}(x;\lambda_{i,0})\|\,dx < \infty$.

(ii) For all $x \in \mathbb{R}$ *and* $\lambda_i \in \Theta_{0,\lambda_i}$, $\int \sup_{\lambda_i\in\Theta_{0,\lambda_i}} \|f_{i,\lambda_i\lambda_i}(x;\lambda_i)\|\,dx < \infty$.

(iii) For all $x \in \mathbb{R}$ *and all* $\lambda_i \in \Theta_{0,\lambda_i}$*, the functions*

$$\frac{f_{i,x}^2(x;\lambda_i)}{f_i^2(x;\lambda_i)} \quad \text{and} \quad \left|\frac{f_{i,xx}(x;\lambda_i)}{f_i(x;\lambda_i)}\right| \quad \text{are dominated by } a_0\left(1+|x|^{a_1}\right),$$
$$\left\|\frac{f_{i,x\lambda_i}(x;\lambda_i)}{f_i(x;\lambda_i)}\right\| \quad \text{and} \quad \left\|\frac{f_{i,x}(x;\lambda_i)}{f_i(x;\lambda_i)}\frac{f_{i,\lambda_i}(x;\lambda_i)}{f_i(x;\lambda_i)}\right\| \quad \text{are dominated by } a_0\left(1+|x|^{a_2}\right),$$
$$\left\|\frac{f_{i,\lambda_i}(x;\lambda_i)}{f_i(x;\lambda_i)}\right\|^2 \quad \text{and} \quad \left\|\frac{f_{i,\lambda_i\lambda_i}(x;\lambda_i)}{f_i(x;\lambda_i)}\right\| \quad \text{are dominated by } a_0\left(1+|x|^{a_3}\right),$$

with $a_0, a_1, a_2, a_3 \geq 0$ *such that* $\int(|x|^{2+a_1}+|x|^{1+a_2}+|x|^{a_3})f_i(x;\lambda_{i,0})\,dx < \infty$ *(*$i = 1, \ldots, n$*).*

These conditions are similar to those in Assumptions 4(ii)–(iv) and again impose suitable integrability conditions on partial derivatives of the density functions $f_i(x;\lambda_i)$. Assumptions 5(i) and (ii) are needed to ensure that, when evaluated at the true parameter value, the expectation of the Hessian matrix has the usual property $E[l_{\theta\theta,t}(\theta_0)] = -Cov[l_{\theta,t}(\theta_0)]$, whereas Assumption 5(iii) guarantees that the (standardized) Hessian matrix obeys an appropriate uniform law of large numbers. These results are given in the following lemma which is proved in Appendix C.

Lemma 2. *If Assumptions 2–5 hold,* $\sup_{\theta\in\Theta_0}\left\|T^{-1}\sum_{t=1}^{T} l_{\theta\theta,t}(\theta) - E\left[l_{\theta\theta,t}(\theta)\right]\right\| \to 0$ *a.s., where* $E[l_{\theta\theta,t}(\theta)]$ *is continuous at* θ_0 *and* $E[l_{\theta\theta,t}(\theta_0)] = -\mathcal{I}(\theta_0)$*.*

In addition to enabling us to establish the asymptotic normality of the ML estimator, Lemma 2 can also be used to obtain a consistent estimator for the covariance matrix of the limiting distribution needed to conduct statistical inference.

4.4 Maximum likelihood estimator

The results of Lemmas 1 and 2 provide the basic ingredients needed to derive the consistency and asymptotic normality of a local ML estimator stated in the following theorem.

Theorem 1. *If Assumptions 2–5 hold, there exists a sequence of solutions* $\hat{\theta}_T$ *to the likelihood equations* $L_{\theta,T}(\theta) = 0$ *such that* $T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$ *as* $T \to \infty$*.*

Theorem 1 shows that the usual result on consistency and asymptotic normality of a local maximizer of the log-likelihood function applies. The proof of Theorem 1, given in Appendix C, is based on arguments used in similar proofs in the previous literature.

A consistent estimator of the covariance matrix $\mathcal{I}(\theta_0)^{-1}$ in Theorem 1 can be obtained by using the ML estimator $\hat{\theta}_T$ and the Hessian matrix of the log-likelihood function. Specifically,

$$-L_{\theta\theta,T}^{-1}(\hat{\theta}_T) \overset{def}{=} -\left(T^{-1}\sum_{t=1}^{T} l_{\theta\theta,t}(\hat{\theta}_T)\right)^{-1} \rightarrow \mathcal{I}(\theta_0)^{-1} \text{ (a.s.).} \tag{9}$$

We omit the proof of this result, which follows from Lemma 2 and Theorem 1 with standard arguments.

4.5 Two-step estimation

The ML estimator $\hat{\theta}_T$ can be computationally rather demanding when the dimension n is not small and relatively short time series are considered. In this section, we therefore discuss two-step alternatives which may be of interest in such cases. We partition the parameter vector θ as $\theta = (\pi, \gamma)$, where π contains the autoregressive parameters (ν and $A_1, \ldots, A_p$) and $\gamma = (\beta, \sigma, \lambda)$ the parameters related to the error term $B\varepsilon_t$.

A computationally convenient and widely used estimator of π is the least squares (LS) estimator denoted by $\tilde{\pi}_T$. In the present non-Gaussian set-up, the LS estimator is consistent (although not efficient), and can therefore serve as a useful estimator for the autoregressive parameters. Thus, a natural way to obtain an estimator of γ, the parameter vector related to the error term, is to replace the parameter π in the log-likelihood function $L_T(\pi, \gamma)$ by the LS estimator $\tilde{\pi}_T$ and maximize the resulting function

$$\tilde{L}_T(\gamma) = L_T(\tilde{\pi}_T, \gamma) = T^{-1}\sum_{t=1}^{T} l_t(\tilde{\pi}_T, \gamma) \tag{10}$$

with respect to γ. Here

$$l_t(\tilde{\pi}_T, \gamma) = \sum_{i=1}^{n} \log f_i\left(\sigma_i^{-1}\iota_i' B(\beta)^{-1} u_t(\tilde{\pi}_T); \lambda_i\right) - \log\det(B(\beta)) - \sum_{i=1}^{n}\log\sigma_i,$$

where $u_t(\tilde{\pi}_T) = y_t - \tilde{\nu}_T - \tilde{A}_{1,T}y_{t-1} - \cdots - \tilde{A}_{p,T}y_{t-p}$ are the LS residuals (here $\tilde{\nu}_T$ and $\tilde{A}_{1,T}, \ldots, \tilde{A}_{p,T}$ denote the appropriate LS estimators). The resulting estimator, denoted by $\tilde{\gamma}_T$, hence uses the LS residuals to estimate the parameters related to the error term $B\varepsilon_t$. As is shown in Theorem 2 below, the estimator $\tilde{\gamma}_T$ is consistent. Moreover, if the components of the error term ε_t are symmetric in a certain sense (for details, see Theorem 2 below), the estimator $\tilde{\gamma}_T$ has the same asymptotic distribution as the ML estimator, so that $\tilde{\gamma}_T$ is even asymptotically efficient and hence a suitable final estimator for γ.

In general, if the components of the error distribution are not symmetric, a second stage estimator is needed to achieve asymptotic efficiency. Let $\tilde{\theta}_T = (\tilde{\pi}_T, \tilde{\gamma}_T)$ denote the preliminary estimator of θ with $\tilde{\pi}_T$ and $\tilde{\gamma}_T$ as described above. An asymptotically efficient second stage estimator of θ can be obtained by a single Newton-Raphson iteration from $\tilde{\theta}_T$, that is, by

$$\tilde{\theta}_T^{(2)} = \tilde{\theta}_T - L_{\theta\theta,T}^{-1}(\tilde{\theta}_T) L_{\theta,T}(\tilde{\theta}_T),$$

where $L_{\theta,T}(\cdot)$ and $L_{\theta\theta,T}(\cdot)$ denote the first and second partial derivatives of $L_T(\cdot)$.

The preceding discussion is summarized in the following theorem. To present the result, we partition the matrix $\mathcal{I}(\theta_0)^{-1}$ conformably with the partition of $\theta = (\pi, \gamma)$ as

$$\mathcal{I}(\theta_0)^{-1} = \begin{bmatrix} \mathcal{I}^{\pi\pi}(\theta_0) & \mathcal{I}^{\gamma\pi}(\theta_0) \\ \mathcal{I}^{\pi\gamma}(\theta_0) & \mathcal{I}^{\gamma\gamma}(\theta_0) \end{bmatrix}.$$

Theorem 2. *Suppose Assumptions 2–5 hold. Then*

(i) The first stage estimator $\tilde{\theta}_T = (\tilde{\pi}_T, \tilde{\gamma}_T)$ is a strongly consistent estimator of θ_0 and $T^{1/2}(\tilde{\theta}_T - \theta_0) = O_p(1)$.

(ii) If for each $i = 1, \ldots, n$, the distribution of $\varepsilon_{i,t}$ is symmetric in the sense that $f_i(x; \lambda_i) = f_i(-x; \lambda_i)$ for all $\lambda_i \in \Theta_{0,\lambda_i}$, then the first stage estimator $\tilde{\gamma}_T$ is asymptotically efficient, i.e., $T^{1/2}(\tilde{\gamma}_T - \gamma_0) \xrightarrow{d} N(0, \mathcal{I}^{\gamma\gamma}(\theta_0))$ as $T \to \infty$.

(iii) The second stage estimator $\tilde{\theta}_T^{(2)}$ is asymptotically efficient, i.e., $T^{1/2}(\tilde{\theta}_T^{(2)} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$ as $T \to \infty$.

The result given in (9) applies with the ML estimator $\hat{\theta}_T$ replaced by the first stage estimator $\tilde{\theta}_T$ so that consistent estimators of the covariance matrices of the limiting distributions in parts (ii) and (iii) of Thereom 2 can be obtained from $L_{\theta\theta,T}^{-1}(\tilde{\theta}_T)$. Finally, it should be noted that the conclusion of Theorem 2(iii) remains valid even if the first stage estimator $\tilde{\theta}_T$ is replaced with some other estimator having the properties listed in part (i) of Theorem 2.

4.6 Testing hypotheses

A major advantage of the non-Gaussian SVAR model is the ability to test restrictions that are partly or exactly identifying in its Gaussian counterpart. Such restrictions, often obtained from the previous literature, may also prove useful in interpretation. Short-run restrictions typically come in the form of zero restrictions on certain elements of the matrix B (or, equivalently, on the elements of β); for instance, in a four-variable SVAR model, B could take one of the following forms:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & * & * & 0 \\ * & 1 & * & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} 1 & * & * & * \\ * & 1 & * & * \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix},$$

where $*$ denotes an arbitrary value. The first matrix implies a recursive structure on the SVAR model, while the second restricts the fourth shock to have an immediate impact on the fourth variable only, and the third precludes the immediate impact of the fourth shock on the third variable. Note that, in the Gaussian SVAR model, only the first set of restrictions is exactly identifying, while the other two do not suffice for identification of the structural shocks.

As the parameter vector θ is fully identified in Θ and the ML (as well as the two-step) estimator has a conventional asymptotic normal distribution, hypothesis tests can be carried out in the usual manner, using standard Wald, likelihood ratio, or Lagrange multiplier tests. In the case of short-run restrictions discussed above, testing is straightforward. Also long-run restrictions imposing zero restrictions on the sum of certain element(s) of the matrices $\Psi_j B$, $j = 0, 1, \ldots$, can be tested by standard tests.

When performing and interpreting tests, one should keep in mind that the straightforward conventional tests require the parameter vector under the null hypothesis to belong to the parameter space considered. In particular, it is required that the assumed value of the matrix B under the null hypothesis belongs to the set $\mathcal{B}$ defined in the Identification Scheme (see Section 3.3). One implication of this is that not all restrictions can be straightforwardly tested (an example is the restriction that a diagonal element of B equals zero). Another, more subtle, implication to be kept in mind is that the particular permutation (of the columns of B and the elements of ε_t) being considered is fixed to the one defined by step (ii) of the Identification Scheme. For instance, one might be tempted to interpret a test of the second set of restrictions above as a test of whether there exists a shock with no immediate impact on the other three variables. However, it should only be interpreted as a test of whether, with this particular ordering, the fourth structural shock has no immediate impact on the first three variables. Therefore, prior to testing restrictions, we recommend labeling the shocks by inspection of impulse response functions, as illustrated in Section 5.

5 Empirical application

The interdependence of monetary policy and the stock market is an issue that has recently awoken a lot of interest and that has been addressed by means of SVAR analysis. Intuitively, one would expect the dynamics of monetary policy actions and the stock market to be closely linked. Movements of stock prices are driven by expectations of future returns that are connected to the business cycle and monetary policy decisions. On the other hand, because of the close interconnections between financial markets and the real economy, policymakers monitor asset prices, and presumably use them as indicators when making monetary policy decisions.

Given the plausibly close connections between financial markets and monetary policy, it is somewhat surprising that typical new-Keynesian models of the business cycle mostly ignore stock prices, as Castelnuovo and Nisticò (2010), among others, have pointed out. They put forth a dynamic stochastic general equilibrium (DSGE) model where the stock market is allowed to play an active role in the determination of the business cycle, and their empirical results with postwar U.S. data indeed lend support to reciprocal effects between financial markets and monetary policy. Specifically, they find an on-impact negative reaction in the stock-price gap following a contractionary monetary policy shock, and an interest rate increase following a positive stock market shock.

While the theoretical literature on interactions between monetary policy and the stock market is scant, empirically this issue has been addressed in a number of papers by means of SVAR analysis using different identification schemes. Examples include Lastrapes (1998) and Rapach (2001) who rely on long-run restrictions for identification, Li, Isca, and Xu (2010) who use non-recursive short-run restrictions, Bjørnland and Leitemo (2009) who consider identification by a combination of short-run and long-run restrictions, and Rigobon and Sack (2004), who base identification on the heteroskedasticity of shocks in high-frequency data. However, short-run recursive restrictions have probably been the most commonly employed approach to identification in this literature; see, e.g., Patelis (1997), Thorbecke (1997), and Cheng and Jin (2013). Empirical results depend on the data and identification scheme used, but typically a monetary policy shock is found not to account for a major part of the variation of stock returns.

However, recursive identification by the Cholesky decomposition has been strongly criticized by Bjørnland and Leitemo (2009) on the grounds that in their U.S. data set (from 1983 to 2002), such identification yields counterintuitive impulse responses. In particular, they found a permanent positive effect on stock returns following a contractionary monetary policy shock, while on economic grounds a temporary negative response is expected. Moreover, recursive ordering, by construction, precludes the immediate impact of a monetary policy (stock market) shock on the stock price (policy rate) if the interest rate (stock return) is placed last in the ordering of the variables as is usually done. This is not theoretically well founded, and it does not conform to Castelnuovo and Nisticò's (2010) DSGE model. According to Castelnuovo's (2013) simulation results, the impulse response functions of a monetary policy shock of a Cholesky-identified SVAR model estimated on data generated from their DSGE model are quite different from those implied by the actual DSGE model. Specifically, the DSGE model predicts a significant negative reaction of financial conditions to a contractionary monetary policy shock, which is necessarily overlooked by the recursive SVAR model.

In this paper, we estimate a four-variable SVAR model with recent U.S. data. Identification is achieved by assuming that the components of the error term are independently t-distributed. Given that financial market data are involved, a distributional assumption allowing for fat-tailed errors seems useful. Moreover, t-distributed shocks have also recently been implemented in DSGE models (see, e.g., Chib and Ramamurthy (2014), and Cúrdia, Del Negro, and Greenwald (2014)). To facilitate direct interpretation of our results in terms of Castelnuovo's (2013) DSGE model, we use the same data set as he did. As discussed in Section 4.6, our identification scheme facilitates testing additional identification restrictions, and we test a number of such restrictions in order to understand the importance of the monetary policy shock for financial markets.

5.1 Data

Our quarterly U.S. data set comprises the same four time series on which Castelnuovo (2013) based the estimates of the parameters of his DSGE model discussed above. The output gap is computed as the log-deviation of the real GDP from the potential output estimated by the Congressional Budget Office. Inflation is measured by the growth rate of the GDP deflator. Instead of a stock return, we include the Kansas City Financial Condition Index (KCFCI) that combines information from a variety of financial indexes (see Hakkio and Keeton (2009) for details, and Castelnuovo (2013, Appendix 4) for further discussion). Federal funds rate (average of monthly values) is the policy interest rate in the model. The output gap (x_t), inflation (π_t), and federal funds rate (R_t) are measured as percentages. Our sample period runs from the beginning of 1990 until the second quarter of 2008. Hence, the time series consist of only 74 observations, but there are a number of reasons to prefer this relatively short

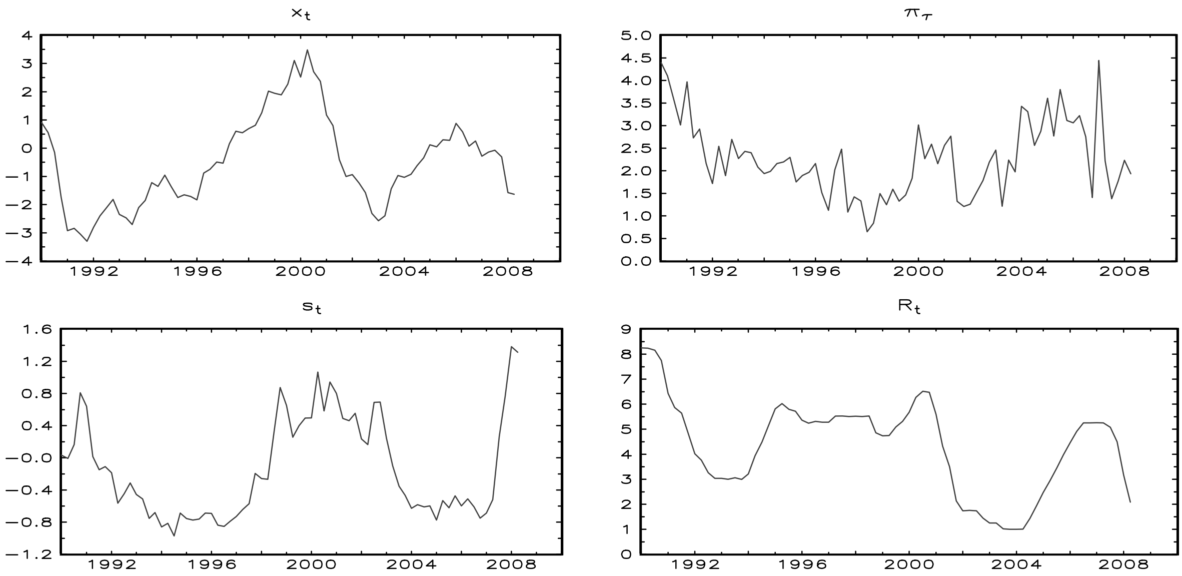

Figure 1: The time series included in the SVAR model.

sample period. First, observations of the KCFCI are not available before 1990, and, second, as Castelnuovo (2013), we also do not want to include earlier data to avoid the plausible policy break prior to the Greenspan-Bernanke regime. Moreover, the most recent data are excluded to avoid having to deal with the acceleration of the financial crisis. The KCFCI series (s_t) is downloaded from the website of the Federal Reserve Bank of Kansas City, while the rest of the data are extracted from FRED database of the Federal Reserve Bank of St. Louis. The time series are depicted in Figure 1.

5.2 Results

We start out by selecting an adequate reduced-form VAR(p) model for the data vector $y_t = (x_t, \pi_t, s_t, R_t)$. The Bayesian and Akaike information criteria select models with one and two lags, respectively. However, according to the multivariate Portmanteau test (with eight lags), only the latter produces serially uncorrelated residuals. Moreover, the solution of Castelnuovo and Nisticò's (2010) DSGE model has a VAR(2) representation. The multivariate Jarque-Bera test soundly rejects normality at the 1% level, and all residual series seem leptokurtic. Thus, we proceed to a second-order SVAR model with errors following independent t-distributions.

Given the short sample period, we estimate the SVAR(2) model by the two-step procedure discussed in Section 4.5. In estimation, the identification restrictions on the matrix B mentioned

Table 1: Estimation results of the unrestricted SVAR(2) model.

B						Equation			
						x_t	π_t	s_t	R_t
	1.000	–0.231	–1.362	–0.760					
	·	(0.115)	(0.592)	(0.954)					
	0.310	1.000	–0.008	0.013					
	(0.259)	·	(0.231)	(0.408)	σ_i	0.293	0.657	0.211	0.199
	0.334	–0.043	1.000	–0.467		(0.082)	(0.203)	(0.051)	(0.065)
	(0.198)	(0.056)	·	(0.336)					
	0.501	–0.049	–0.335	1.000	λ_i	10.480	3.142	4.070	14.375
	(0.358)	(0.070)	(0.293)	·		(9.521)	(1.473)	(2.535)	(18.886)

Notes: The model is estimated by the two-step method described in Section 4.5. The intercept term and the coefficient matrices of the lags are estimated in the first step by ordinary least squares, and kept fixed in the second-step maximization of (10). The figures in parentheses are standard errors computed from the Hessian of the log-likelihood function.

in Section 3.3 are imposed. The estimation results are presented in Table 1. We only report the estimates of B and the scale (σ_i) and degree-of-freedom (λ_i) parameters corresponding to the errors of each equation i. They are conditional on the OLS estimates of the intercept term and the coefficient matrices estimated by OLS in the first step and kept fixed in the second-step maximization of (10).[7] The fit of the SVAR(2) model to the data appears quite good. According to the Ljung-Box test with eight lags, there is no evidence of remaining autocorrelation in the residuals (the p-values for the four residual series are 0.17, 0.16, 0.66, and 0.50). Also, no remaining conditional heteroskedasticity is detected (the p-values of the McLeod-Li test with eight lags for the four residual series equal 0.10, 0.92, 0.87, and 0.95). The residuals are virtually uncorrelated, and do not exhibit any significant cross correlations[8], lending support to the independence assumption underlying identification. The estimates of the degree-of-freedom parameters suggest clear deviations from normality, which is required for identification. The fit of the error distributions is also reasonable as shown by the quantile-quantile plots in Figure 2.

In order to interpret the estimation result, we compute the implied impulse response functions. However, as discussed in Section 3, the identified shocks do not, as such, carry any economic interpretation despite exact identification. Therefore, along the lines of Lütkepohl and Netšunajev (2014a), we use sign restrictions to help in economic identification. It is es-

[7] The first-step OLS estimation results are available upon request.

[8] To save space, the detailed results are not reported, but they are available upon request.

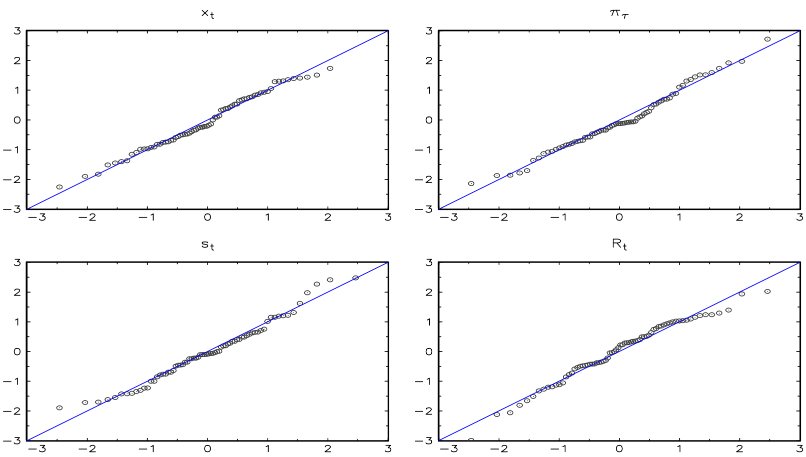

Figure 2: Quantile-quantile plots of the residuals of the unrestricted SVAR(2) model.

pecially the monetary policy shock that we are interested in, and its qualitative properties on which there is considerable agreement in the established literature, are summarized by Christiano, Eichenbaum, and Evans (1999), among others. As far as the variables included in our SVAR model are concerned, these properties are as follows: after a contractionary monetary policy shock, the short-term interest rate rises, output (gap) increases, and inflation responds very slowly. Because of the arguments presented at the beginning of this section, there should be an immediate negative effect on the financial condition index.

The impulse response functions of one standard deviation shocks up to 16 quarters ahead are depicted in Figure 3. Each row contains the impulse responses of all variables to one shock. Following the common practice in the literature, 68% (pointwise Hall's percentile) confidence bands are plotted to facilitate the assessment of the significance of the impulse reponses. They are obtained by residual-based bootstrap (1,000 replications). In bootstrapping, the intercept term and the coefficient matrices of the VAR model were estimated by OLS, and the ML estimates of B, σ_i, and λ_i $(i = 1, 2, 3, 4)$ reported in Table 1 were used as starting values.

Judged by the confidence bands, only the shocks on the top and bottom rows have a nonzero (positive) immediate impact on the interest rate, and they are thus possible candidates for a (contractionary) monetary policy shock. Both have a significantly negative impact on inflation over time as would be expected of a monetary policy shock, but the impact of the shock on the top row on the output gap is (significantly) positive, while that of the other shock is (insignificantly) negative. A contractionary monetary policy shock should not have a positive impact on output, and therefore, the shock on the bottom row can be labeled as the monetary

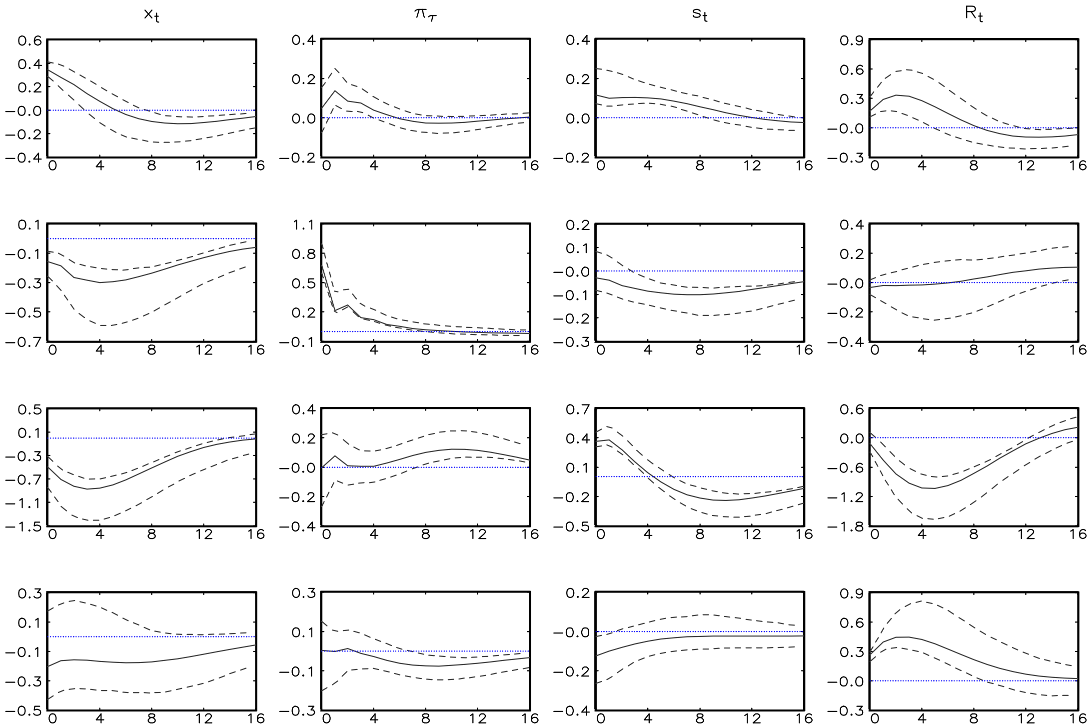

Figure 3: Impulse response functions implied by the unrestricted SVAR model. Each row contains the impulse responses of all variables to one shock. The dashed lines are the pointwise 68% Hall's percentile confidence bands.

policy shock. Interestingly, it also has a significant negative immediate impact on financial conditions, and the negative effect lasts for one to two years. With the exception of inflation, the magnitudes of the impact effects and the time it takes for the impulse responses to revert to zero are quite well in line with those implied by the DSGE model of Castelnuovo (2013).

In order to gauge the importance of the restriction involved in recursive identification that a monetary policy shock cannot have an immediate impact on financial conditions s_t, we next consider two restricted models. First, we estimate a model where the matrix B is restricted to be lower-triangular with the exception that the $b_{3,4}$ element is left unrestricted to allow the monetary policy shock to have an immediate impact on s_t. Then, we restrict also that element to zero to obtain a model corresponding to the Cholesky identification commonly employed in the previous literature. The p-value of the likelihood ratio (LR) test of the former restriction is 0.253, indicating nonrejection at conventional significance levels. The p-values of the LR and Wald tests of the restriction $b_{3,4} = 0$ in this model, in turn, equal 0.052 and 0.013, respectively, indicating rejection (at least at the 10% level) of the hypothesis that the monetary

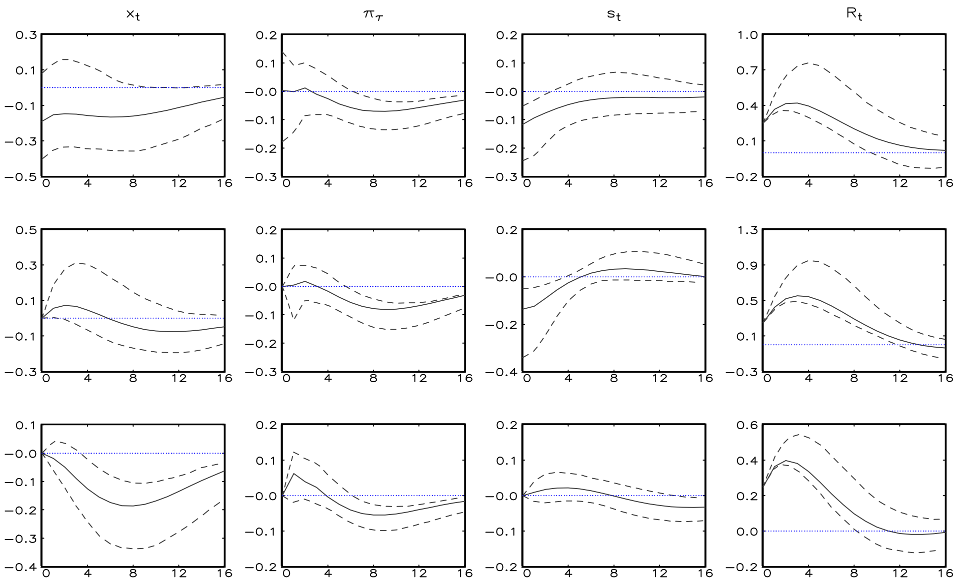

Figure 4: Impulse response functions of the monetary policy shock in the unrestricted (upper panel) and the restricted SVAR models (lower panel: recursive model; middle panel: model allowing for immediate impact of the monetary policy shock on financial conditions).

policy shock has no immediate impact on financial markets. This evidence against Cholesky identification is in line with the results of Lütkepohl and Netšunajev (2014b), who achieved exact identification in a similar SVAR model for U.S. data by introducting a smooth transition in the error covariance matrix.

The impulse response functions of the monetary policy shock implied by the unrestricted and restricted models along with their 68% bootstrap confidence intervals are depicted in Figure 4. From the middle panel it is seen that, when not restricted to zero, the immediate impact of the monetary policy shock on financial conditions is negative, and its effect dies out quickly in the same way as in the unrestricted model. As far as the rejected recursive model is concerned, the impulse response functions in the lower panel show no significant impact on financial conditions. Given that our unrestricted model seems to fit the data reasonably well and the recursive restrictions are rejected, we can thus conclude that a contractionary monetary policy shock indeed has a negative albeit short-lived effect on financial conditions.

6 Conclusion

In this paper, we have considered identification and estimation of SVAR models with non-Gaussian errors. Specifically, we considered a SVAR model where the components of the error process were assumed non-Gaussian and independent. Deviations from Gaussianity, especially fat-tailed error distributions, are often encountered in VAR analysis, and therefore we expect the model to be useful in a large number of applications. Our first identification result showed that, together with standard VAR assumptions, the non-Gaussianity and independence assumptions are sufficient for identification up to permutation and scaling of the structural shocks, which facilitates impulse response analysis. We also presented an Identification Scheme yielding complete identification, a prerequisite for the development of conventional estimation theory.

Under mild technical conditions, we showed consistency and asymptotic normality of the maximum likelihood estimator and a two-step estimator devised for computationally demanding situations. Due to complete statistical identification and standard asymptotic estimation theory, additional economic identifying restrictions, such as commonly used short-run and long-run restrictions, can be tested, which is a particularly convenient feature of the non-Gaussian SVAR model.

We illustrated the new methods in an empirical application to the relationship between the U.S. stock market and monetary policy. In previous studies, the instantaneous impact of a monetary policy shock on the stock market has either been precluded at the outset or found relatively minor or insignificant. In contrast, we found the monetary policy shock to have a negative significant instantaneous impact on the stock market. Moreover, we were able to clearly reject the recursive identification scheme precluding an instantaneous impact of the monetary policy shock on the stock market, employed in part of the previous literature.

Appendix A: Technical details for Section 3

The proof of Proposition 1 makes use of a well-known result referred to as the Skitovich-Darmois theorem (see, e.g., Theorem 3.1.1 in Kagan, Linnik, and Rao (1973)). A variant of this theorem has also been used by Comon (1994) to obtain identifiability in the context of an independent component model. For ease of reference, we first provide this result as the following lemma.

Lemma A.1. *(Kagan et al. (1973), Theorem 3.1.1). Let $X_1, \ldots, X_n$ be independent (not necessarily identically distributed) random variables, and define $Y_1 = \sum_{i=1}^{n} a_i X_i$ and $Y_2 = \sum_{i=1}^{n} b_i X_i$ where a_i and b_i are constants. If Y_1 and Y_2 are independent, then the random variables X_j for which $a_j b_j \neq 0$ are all normally distributed.*

Now we can prove Proposition 1. The proof is straigthforward with the most essential part being based on arguments already used by Comon (1994).

Proof of Proposition 1. First note that (5) can be expressed as $y_t = \mu + A(L)^{-1} B\varepsilon_t = \mu^* + A^*(L)^{-1} B^*\varepsilon_t^*$, where L denotes the lag operator (e.g., $Ly_t = y_{t-1}$). Taking expectations this implies that $E[y_t] = \mu = \mu^*$. Without loss of generality we can continue by assuming that $\mu = \mu^* = 0$ (alternatively, we can replace y_t below by $y_t - \mu$). From the preceding equation we then obtain $y_t - A_1 y_{t-1} - \cdots - A_p y_{t-p} = B\varepsilon_t$ and $y_t - A_1^* y_{t-1} - \cdots - A_p^* y_{t-p} = B^*\varepsilon_t^*$. Denoting $\boldsymbol{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})$ $(np \times 1)$, $\mathbf{A} = [A_1 : \cdots : A_p]$ $(n \times np)$, and $\mathbf{A}^* = [A_1^* : \cdots : A_p^*]$ $(n \times np)$, this implies that

$$B\varepsilon_t - B^*\varepsilon_t^* = (A_1^* - A_1)y_{t-1} + \cdots + (A_p^* - A_p)y_{t-p} = (\mathbf{A}^* - \mathbf{A})\boldsymbol{y}_{t-1}. \tag{11}$$

Multiplying this equation from the right by $\boldsymbol{y}_{t-1}'$ and taking expectations yields

$$E[(B\varepsilon_t - B^*\varepsilon_t^*)\boldsymbol{y}_{t-1}'] = (\mathbf{A}^* - \mathbf{A})E[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}'],$$

and, as both ε_t and ε_t^* are uncorrelated with $\boldsymbol{y}_{t-1}$ (due to (5) and Assumptions 1(ii) and 1(iii)), we get $(\mathbf{A}^* - \mathbf{A})E[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}'] = 0$. As there can be no exact linear dependences between the components of the vector $\boldsymbol{y}_{t-1}$, its covariance matrix $E[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}']$ is positive definite, so that $\mathbf{A}^* - \mathbf{A} = 0$ must hold. From the definitions of Ψ_j and Ψ_j^* and equation (11) it therefore follows that $\Psi_j^* = \Psi_j$, $j = 0, 1, \ldots$, and $B\varepsilon_t = B^*\varepsilon_t^*$. Using the nonsingularity of B we can solve ε_t from this equation and obtain

$$\varepsilon_t = M\varepsilon_t^*, \quad \text{where} \quad M = B^{-1}B^*. \tag{12}$$

By condition (iii) in the Proposition and Assumption 1(ii), the random variables $\varepsilon^*_{1,t}, \ldots, \varepsilon^*_{n,t}$ are mutually independent and at most one of them has a Gaussian marginal distribution. Also the random variables $\varepsilon_{1,t}, \ldots, \varepsilon_{n,t}$ are mutually independent. Therefore by Lemma A.1, at most one column of M may contain more than one nonzero element. Suppose, say, the kth column of M has at least two nonzero elements, m_{ik} and m_{jk} $(i \neq j)$. Then $\varepsilon_{i,t} = m_{ik}\varepsilon^*_{k,t} + \sum_{l=1,\ldots,n;\ l\neq k} m_{il}\varepsilon^*_{l,t}$ and $\varepsilon_{j,t} = m_{jk}\varepsilon^*_{k,t} + \sum_{l=1,\ldots,n;\ l\neq k} m_{jl}\varepsilon^*_{l,t}$ with the random variable $\varepsilon^*_{k,t}$ being Gaussian (due to Lemma A.1) with positive variance (due to Assumption 1(i) for the process ε^*_t). Moreover, for all $l = 1, \ldots, n$, $l \neq k$, it must hold that $m_{il}m_{jl} = 0$ because only the kth column of M could have more than one nonzero element. This, however, implies (because the random variables $\varepsilon^*_{1,t}, \ldots, \varepsilon^*_{n,t}$ are independent) that $E[\varepsilon_{i,t}\varepsilon_{j,t}] = m_{ik}m_{jk}E[\varepsilon^{*2}_{k,t}] \neq 0$ so that the random variables $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$ are not independent, a contradiction. Therefore each column of M has at most one nonzero element. Now, by the invertibility of M, it follows that each column of M has exactly one nonzero element, and for the same reason, also that each row of M has exactly one nonzero element. Therefore there exist a permutation matrix P and a diagonal matrix $D = diag\,(d_1, \ldots, d_n)$ with nonzero diagonal elements such that $M = DP$. This together with (12) implies that $\varepsilon^*_t = P'D^{-1}\varepsilon_t$ and $B^* = BDP$, thus completing the proof. ■

Parts (a) and (b) of Proposition 2 are rather straightforward to prove based on the Identification Scheme.

Proof of Proposition 2, parts (a) and (b). We begin with part (b). To show that $\mathcal{B}$ contains representatives from each $\sim$-equivalence class of $\mathcal{I}$, choose any $B \in \mathcal{I}$. Then by the definition of $\mathcal{B}$, the matrix $\Pi\,(B) = BD_1PD_2$ belongs to $\mathcal{B}$. Moreover, $B \sim \Pi\,(B) = BD_1PD_2$ (because necessarily $D_1PD_2 = D_3P$ for some diagonal D_3 with nonzero diagonal elements). To show that such a representative must be unique, suppose $\tilde{B}_1, \tilde{B}_2 \in \mathcal{B}$ and $\tilde{B}_1 \sim \tilde{B}_2$. Then for some $B_1 \sim B_2$ in $\mathcal{I}$, $\tilde{B}_1 = \Pi(B_1)$ and $\tilde{B}_2 = \Pi(B_2)$, so that

$$B_2 = B_1DP, \quad \tilde{B}_1 = B_1D_1(B_1)P(B_1)D_2(B_1), \quad \text{and} \quad \tilde{B}_2 = B_2D_1(B_2)P(B_2)D_2(B_2)$$

(where we have made the dependence on B_1 and B_2 explicit). Thus $\tilde{B}_2 = B_1DPD_1(B_2)P(B_2)D_2(B_2)$. In the expressions

$$\tilde{B}_1 = B_1D_1(B_1)P(B_1)D_2(B_1) \quad \text{and} \quad \tilde{B}_2 = B_1DPD_1(B_2)P(B_2)D_2(B_2)$$

the matrices $B_1D_1(B_1)$ and $B_1DPD_1(B_2)$ are matrices with the same columns but potentially in different order (this follows from the identity $B_2 = B_1DP$ and the definitions of $D_1(B_1)$

and $D_1(B_2)$). Therefore, by the definitions of the matrices $P(B_1)$ and $P(B_2)$, it necessarily holds that $B_1 D_1(B_1)P(B_1) = B_1 DPD_1(B_2)P(B_2)$. Thus, due to the definitions of $D_2(B_1)$ and $D_2(B_2)$, the result $\tilde{B}_1 = \tilde{B}_2$ also follows, implying the desired uniqueness. Finally, to show that the representatives of different equivalence classes are distinct, suppose (on the contrary) that $\Pi(B_1) = \Pi(B_2)$ but $B_1 \nsim B_2$. Then $B_1 D_1(B_1)P(B_1)D_2(B_1) = B_2 D_1(B_2)P(B_2)D_2(B_2)$, and solving this equation for B_2 implies the existence of a permutation matrix P and a diagonal matrix D such that $B_2 = B_1 DP$, a contradiction with $B_1 \nsim B_2$. Thus, the representatives must be distinct, and the proof of part (b) is complete.

Having established part (b), to prove (a), it now suffices to note that if $B, B^* \in \mathcal{B}$ are as in Proposition 1, then $B^* = BDP$ so that $B^* \sim B$. Then, by the uniqueness proved in part (b), necessarily $B^* = B$. ■

The proof of Proposition 2(c) is somewhat more intricate and we resort to using results based on basic algebraic geometry. In what follows, we first define a few concepts from algebraic geometry we need, then present three auxiliary results, and finally prove Proposition 2(c) as a (rather straightforward) consequence of these auxiliary results. A comprehensive reference for the employed concepts is, e.g., Bochnak, Coste, and Roy (1998).

Consider the m-dimensional Euclidean space $\mathbb{R}^m$. A subset $A \subseteq \mathbb{R}^m$ is called a *semi-algebraic set* (cf. Bochnak et al. (1998, Definition 2.1.4)) if it is of the form

$$A = \cup_{i=1}^{s} \cap_{j=1}^{r_i} \{x \in \mathbb{R}^m : f_{i,j}(x) *_{i,j} 0\}, \tag{13}$$

where, for each $i = 1, \ldots, s$ and $j = 1, \ldots, r_i$, $f_{i,j}(\cdot)$ is a polynomial function (of finite order) in m variables and $*_{i,j}$ is either $=$, $<$, $>$, or $\neq$. A semi-algebraic set is called an *algebraic set* if in (13) the $*_{i,j}$ is always $=$ (Bochnak et al. (1998, Definition 2.1.1)). Lacking a better term, we will call a semi-algebraic set a *semi-algebraic set with equality constraints* if in (13) for each $i = 1, \ldots, s$ at least one of the $*_{i,j}$ is $=$ with the corresponding $f_{i,j}$ not being identically equal to zero. Finally, the quantifier 'proper' is used in connection with these terms (e.g., proper algebraic set) if $A \neq \mathbb{R}^m$.

As (proper) algebraic sets are built from zeros of polynomial functions, intuition tells that in some sense they must be 'small' in $\mathbb{R}^m$ (in $\mathbb{R}$ they are finite, in $\mathbb{R}^2$ finite intersections/unions of plane curves, etc.). This is indeed the case, as the following well-known result shows (as we were unable to find a convenient reference, we include a proof in the Supplementary Appendix for completeness).

Lemma A.2. *A proper algebraic set A of $\mathbb{R}^m$ has Lebesgue measure zero in $\mathbb{R}^m$. Its complement $\mathbb{R}^m \backslash A$ in $\mathbb{R}^m$ is an open and dense subset of $\mathbb{R}^m$.*

Semi-algebraic sets are not necessarily 'small', but as the following result shows, semi-algebraic sets with equality constraints are (proof in the Supplementary Appendix).

Lemma A.3. *A proper semi-algebraic set with equality constraints A of $\mathbb{R}^m$ has Lebesgue measure zero in $\mathbb{R}^m$. Its complement $\mathbb{R}^m \backslash A$ in $\mathbb{R}^m$ contains an open and dense subset of $\mathbb{R}^m$.*

Now, consider the set of all (real) $n \times n$ matrices, which we denote with $\mathcal{M}_n^A$. As matrices belonging to $\mathcal{M}_n^A$ can be identified with vectors of $\mathbb{R}^{n^2}$ the preceding results can be applied to algebraic sets of $\mathcal{M}_n^A$ and any statement on algebraic sets of $\mathcal{M}_n^A$ can be formulated in terms of corresponding algebraic sets of $\mathbb{R}^{n^2}$ and vice versa. Recall that the set of all invertible $n \times n$ matrices is denoted with $\mathcal{M}_n$. In Proposition 2 we end up excluding the set $\mathcal{E} \overset{def}{=} \mathcal{M}_n \backslash \mathcal{I}$. This set is a proper semi-algebraic set with equality constraints as the next result shows (proof in the Supplementary Appendix).

Lemma A.4. *The set $\mathcal{E} = \mathcal{M}_n \backslash \mathcal{I}$ is a proper semi-algebraic set with equality constraints of $\mathcal{M}_n^A$.*

Part (c) of Proposition 2 now follows from the preceding lemmas in a straightforward fashion.

Proof of Proposition 2, part (c). The fact that $\mathcal{E}$ has Lebesgue measure zero in $\mathbb{R}^{n \times n}$ follows directly from Lemmas A.3 and A.4. From these Lemmas it also follows that the set $\mathcal{M}_n^A \backslash \mathcal{E}$ contains an open and dense subset of $\mathcal{M}_n^A$, say O. Note also that the set $\mathcal{M}_n^A \backslash \mathcal{M}_n$ is a proper algebraic subset of $\mathcal{M}_n^A$, and therefore $\mathcal{M}_n$ is an open and dense subset of $\mathcal{M}_n^A$ (this holds because the determinant of a matrix is a polynomial function, and a matrix is noninvertible if the determinant equals zero). Elementary calculations can now be used to show that $O \cap \mathcal{M}_n \subseteq \mathcal{I} = \mathcal{M}_n \cap (\mathcal{M}_n^A \backslash \mathcal{E})$ is an open and dense subset of $\mathcal{M}_n$. ■

Appendix B: Technical details for Section 4.2

Expression of the score. As in Sections 4.1 and 4.2, we use the notation $x_{t-1} = (1, y_{t-1}, \ldots, y_{t-p})$ and $\pi = vec\left([\nu : A_1 : \cdots : A_p]\right)$, and express $u_t\left(\pi\right) = y_t - \nu - A_1 y_{t-1} - \cdots - A_p y_{t-p}$ briefly as $u_t\left(\pi\right) = y_t - (x_{t-1}' \otimes I_n)\pi$. Regarding the matrix $B(\beta)$, for brevity we do not make its dependence on β explicit and denote $B = B(\beta)$. When $B(\beta)$ is evaluated at $\beta = \beta_0$, we denote

$B_0 = B(\beta_0)$. We also define $\varepsilon_{i,t}(\theta) = \iota_i' B^{-1} u_t(\pi)$ (in the notation we ignore the fact that $\varepsilon_{i,t}(\theta)$ does not depend on the parameter vector λ) and $\varepsilon_t(\theta) = (\varepsilon_{1,t}(\theta), \ldots, \varepsilon_{n,t}(\theta))$. Note that when evaluated at the true parameter values we have $u_t(\pi_0) = B_0 \varepsilon_t$ and $\varepsilon_{i,t}(\theta_0) = \varepsilon_{i,t}$. Furthermore, define

$$e_{i,x,t}(\theta) = \frac{f_{i,x}(\sigma_i^{-1} \iota_i' B^{-1} u_t(\pi); \lambda_i)}{f_i(\sigma_i^{-1} \iota_i' B^{-1} u_t(\pi); \lambda_i)} \quad \text{and} \quad e_{i,\lambda_i,t}(\theta) = \frac{f_{i,\lambda_i}(\sigma_i^{-1} \iota_i' B^{-1} u_t(\pi); \lambda_i)}{f_i(\sigma_i^{-1} \iota_i' B^{-1} u_t(\pi); \lambda_i)},$$

and use them to form the $n \times 1$ and $d \times 1$ vectors

$$e_{x,t}(\theta) = (e_{1,x,t}(\theta), \ldots, e_{n,x,t}(\theta)) \quad \text{and} \quad e_{\lambda,t}(\theta) = (e_{1,\lambda_1,t}(\theta), \ldots, e_{n,\lambda_n,t}(\theta)).$$

Finally, denote $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$.

As in Section 4.2, let $l_{\theta,t}(\theta) = (l_{\pi,t}(\theta), l_{\beta,t}(\theta), l_{\sigma,t}(\theta), l_{\lambda,t}(\theta))$ with $l_{\sigma,t}(\theta) = (l_{\sigma_1,t}(\theta), \ldots, l_{\sigma_n,t}(\theta))$ and $l_{\lambda,t}(\theta) = (l_{\lambda_1,t}(\theta), \ldots, l_{\lambda_n,t}(\theta))$ be the score vector of θ based on a single observation. With straightforward differentiation (details omitted but available in the Supplementary Appendix) one obtains

$$l_{\pi,t}(\theta) = -(x_{t-1} \otimes B^{-1'} \Sigma^{-1}) e_{x,t}(\theta), \tag{14a}$$

$$l_{\beta,t}(\theta) = -H'[(B^{-1} u_t(\pi) \otimes B^{-1'} \Sigma^{-1}) e_{x,t}(\theta) + vec(B^{-1'})], \tag{14b}$$

$$l_{\sigma,t}(\theta) = -\Sigma^{-2} [\varepsilon_t(\theta) \odot e_{x,t}(\theta) + \sigma], \tag{14c}$$

$$l_{\lambda,t}(\theta) = e_{\lambda,t}(\theta), \tag{14d}$$

which form $L_{\theta,T}(\theta) = T^{-1} \sum_{t=1}^{T} l_{\theta,t}(\theta)$, the score vector of θ.

An auxiliary lemma. The following lemma contains results needed in subsequent derivations. Its proof is straightforward and is given in the Supplementary Appendix.

Lemma B.1. Under Assumptions 2–4, the following hold for $i = 1, \ldots, n$: (i) $E[e_{i,x,t}] = 0$, (ii) $E[e_{i,x,t}^2] < \infty$, (iii) $E[e_{i,\lambda_i,t}] = 0$, (iv) $E[e_{i,\lambda_i,t} e_{i,\lambda_i,t}']$ is finite, (v) $E[\varepsilon_{i,t} e_{i,x,t}] = -\sigma_{i,0}$, (vi) $E[\varepsilon_{i,t}^2 e_{i,x,t}^2] < \infty$.

Martingale property of the score. Consider $L_{\theta,T}(\theta_0) = T^{-1} \sum_{t=1}^{T} l_{\theta,t}(\theta_0)$, the score vector of θ evaluated at the true parameter value. As in Section 4.2 we denote $e_{x,t}(\theta_0) = e_{x,t} = (e_{1,x,t}, \ldots, e_{n,x,t})$ and $e_{\lambda,t}(\theta_0) = e_{\lambda,t} = (e_{1,\lambda_1,t}, \ldots, e_{n,\lambda_n,t})$ and note that

$$e_{i,x,t} = \frac{f_{i,x}(\sigma_{i,0}^{-1} \varepsilon_{i,t}; \lambda_{i,0})}{f_i(\sigma_{i,0}^{-1} \varepsilon_{i,t}; \lambda_{i,0})} \quad \text{and} \quad e_{i,\lambda_i,t} = \frac{f_{i,\lambda_i}(\sigma_{i,0}^{-1} \varepsilon_{i,t}; \lambda_{i,0})}{f_i(\sigma_{i,0}^{-1} \varepsilon_{i,t}; \lambda_{i,0})}, \quad i = 1, \ldots, n.$$

Let $E_t[\cdot]$ signify the conditional expectation given the sigma-algebra $\mathcal{F}_t = \sigma(\varepsilon_{t-j}, j \geq 0)$ or, equivalently, the sigma-algebra $\sigma(y_{t-j}, j \geq 0)$ (see (4)). We need to demonstrate that $\{l_{\theta,t}(\theta_0), \mathcal{F}_t\}$ is a martingale difference sequence.

First note that $l_{\pi,t}(\theta_0) = -(x_{t-1} \otimes B_0^{-1'}\Sigma_0^{-1})e_{x,t}$ so that for this component of $l_{\theta,t}(\theta_0)$ the desired result follows from $E_{t-1}[(x_{t-1} \otimes B_0^{-1'}\Sigma_0^{-1})e_{x,t}] = 0$ which holds in view of Lemma B.1(i) and the independence of x_{t-1} and ε_t. Next consider $l_{\lambda,t}(\theta_0) = e_{\lambda,t}$ where $e_{\lambda,t}$ is an IID sequence so that it suffices to show that $E[e_{\lambda,t}] = 0$ which holds by Lemma B.1(iii). As seen from (14c), $l_{\sigma,t}(\theta_0)$ is an IID sequence and $E_{t-1}[l_{\sigma,t}(\theta_0)] = 0$ follows from the identity $E[\varepsilon_{i,t}e_{i,x,t}] = -\sigma_{i,0}$ obtained from Lemma B.1(v). Finally, consider $l_{\beta,t}(\theta_0)$. As $B_0^{-1}u_t(\pi_0) = \varepsilon_t$ and $e_{x,t}(\theta_0) = e_{x,t}$ are IID sequences, we only need to show that $E[\varepsilon_t \otimes B_0^{-1'}\Sigma_0^{-1}e_{x,t}] = -vec(B_0^{-1'})$ (see (8b)). To this end, note that $\varepsilon_{i,t}$ and $e_{j,x,t}$ are independent when $i \neq j$, so that from Lemma B.1(i) and (v) it follows that $E[\varepsilon_{i,t}e_{j,x,t}] = -\sigma_{i,0}$ when $i = j$ and zero otherwise. Thus, as $\varepsilon_t \otimes B_0^{-1'}\Sigma_0^{-1}e_{x,t} = vec\left(B_0^{-1'}\Sigma_0^{-1}e_{x,t}\varepsilon_t'\right)$ and $E[e_{x,t}\varepsilon_t'] = -\Sigma_0$ we find that

$$E[\varepsilon_t \otimes B_0^{-1'}\Sigma_0^{-1}e_{x,t}] = vec\left(E\left[B_0^{-1'}\Sigma_0^{-1}e_{x,t}\varepsilon_t'\right]\right) = -vec(B_0^{-1'}),$$

which shows the desired result.

Covariance matrix of the score – Expression. We derive the components of $Cov[l_{\theta,t}(\theta_0)]$ which equal the components of $Cov[L_{\theta,T}(\theta_0)]$ (see (8a)-(8d)). To this end, denote $V_{e_x} = Cov[e_{x,t}]$ $(n \times n)$, $V_{e_\lambda} = Cov[e_{\lambda,t}]$ $(d \times d)$, and $V_{e_x e_\lambda} = Cov[e_{x,t}, e_{\lambda,t}]$ $(n \times d)$, and note that by Assumption 2(i) and Lemma B.1(i)–(iv), V_{e_x} is a diagonal matrix with finite diagonal elements, V_{e_λ} is a block-diagonal matrix with finite diagonal blocks, and $Cov[e_{i,x,t}, e_{j,\lambda,t}] = 0$ for $i \neq j$. To derive the expression of $Cov[l_{\theta,t}(\theta_0)]$, first consider its diagonal blocks (the finiteness of the blocks of $Cov[l_{\theta,t}(\theta_0)]$ is here assumed and justified below). Straightforward computation leads to the expressions

$$\begin{aligned} Cov[l_{\pi,t}(\theta_0)] &= E\left[x_{t-1}x_{t-1}'\right] \otimes B_0^{-1'}\Sigma_0^{-1}V_{e_x}\Sigma_0^{-1}B_0^{-1}, \\ Cov[l_{\beta,t}(\theta_0)] &= H'(I_n \otimes B_0^{-1'}\Sigma_0^{-1})E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right]\left(I_n \otimes \Sigma_0^{-1}B_0^{-1}\right)H - H'vec(B_0^{-1'})vec(B_0^{-1'})'H, \\ Cov[l_{\lambda,t}(\theta_0)] &= V_{e_\lambda}, \end{aligned}$$

where in deriving the second result we have used the result $E[\varepsilon_t \otimes B_0^{-1'}\Sigma_0^{-1}e_{x,t}] = -vec(B_0^{-1'})$ obtained above. The covariance matrix of $l_{\sigma,t}(\theta_0)$ is

$$Cov[l_{\sigma,t}(\theta_0)] = \Sigma_0^{-2}E\left[(\varepsilon_t \odot e_{x,t} + \sigma_0)(\varepsilon_t \odot e_{x,t} + \sigma_0)'\right]\Sigma_0^{-2},$$

a diagonal matrix with diagonal elements

$$E[(\sigma_{i,0}^{-2}\varepsilon_{i,t}e_{i,x,t}+\sigma_{i,0}^{-1})^2]=\sigma_{i,0}^{-2}E[(\sigma_{i,0}^{-1}\varepsilon_{i,t}e_{i,x,t}+1)^2]=\sigma_{i,0}^{-4}(E[\varepsilon_{i,t}^2e_{i,x,t}^2]-\sigma_{i,0}^2),\quad i=1,\ldots,n.$$

The off-diagonal blocks of $Cov\left[l_{\theta,t}\left(\theta_0\right)\right]$ can be derived by straightforward computation by using the expressions in (8), Lemma B.1, the martingale difference property of $l_{\theta,t}\left(\theta_0\right)$, the result $E[\varepsilon_t\otimes B_0^{-1\prime}\Sigma_0^{-1}e_{x,t}]=-vec(B_0^{-1\prime})$ derived above, and the independence of x_{t-1} and $(\varepsilon_t,e_{x,t},e_{\lambda,t})$. The resulting expressions are

$$\begin{aligned}
Cov\left[l_{\pi,t}\left(\theta_0\right),l_{\beta,t}\left(\theta_0\right)\right] &= (E\left[x_{t-1}\right]\otimes B_0^{-1\prime}\Sigma_0^{-1})E\left[\varepsilon_t'\otimes e_{x,t}e_{x,t}'\right](I_n\otimes\Sigma_0^{-1}B_0^{-1})H,\\
Cov\left[l_{\pi,t}\left(\theta_0\right),l_{\sigma,t}\left(\theta_0\right)\right] &= (E\left[x_{t-1}\right]\otimes B_0^{-1\prime}\Sigma_0^{-1})E[e_{x,t}\left(\varepsilon_t\odot e_{x,t}\right)']\Sigma_0^{-2},\\
Cov\left[l_{\pi,t}\left(\theta_0\right),l_{\lambda,t}\left(\theta_0\right)\right] &= -E\left[x_{t-1}\right]\otimes B_0^{-1\prime}\Sigma_0^{-1}E[e_{x,t}e_{\lambda,t}'],\\
Cov\left[l_{\beta,t}\left(\theta_0\right),l_{\sigma,t}\left(\theta_0\right)\right] &= H'(I_n\otimes B_0^{-1\prime}\Sigma_0^{-1})E\left[\left(\varepsilon_t\otimes e_{x,t}\right)\left(\varepsilon_t\odot e_{x,t}\right)'\right]\Sigma_0^{-2}-H'vec(B_0^{-1\prime})\sigma_0'\Sigma_0^{-2},\\
Cov\left[l_{\beta,t}\left(\theta_0\right),l_{\lambda,t}\left(\theta_0\right)\right] &= -H'(I_n\otimes B_0^{-1\prime}\Sigma_0^{-1})E[\varepsilon_t\otimes e_{x,t}e_{\lambda,t}'],\\
Cov\left[l_{\sigma,t}\left(\theta_0\right),l_{\lambda,t}\left(\theta_0\right)\right] &= -\Sigma_0^{-2}E\left[\left(\varepsilon_t\odot e_{x,t}\right)e_{\lambda,t}'\right].
\end{aligned}$$

Covariance matrix of the score – Finiteness. By the Cauchy-Schwarz inequality, it suffices to show that the diagonal blocks of $Cov\left[l_{\theta,t}\left(\theta_0\right)\right]$ are finite. This, in turn, is the case if the following expectations are finite:

$$\text{(i) } E[x_{t-1}x_{t-1}'],\quad \text{(ii) } V_{e_x},\quad \text{(iii) } E[\varepsilon_t\varepsilon_t'\otimes e_{x,t}e_{x,t}'],\quad \text{(iv) } V_{e_\lambda},\ \text{ and }\ \text{(v) } E[\varepsilon_{i,t}^2e_{i,x,t}^2].$$

The elements of $E[x_{t-1}x_{t-1}']$ in (i) can be expressed in terms of the expectation of y_t and the covariance matrices $Cov\left[y_t,y_{t+k}\right]$, $k=0,\ldots,p$, and are thus finite. Finiteness of the moments in (ii) and (iv) was already noted above. A typical element of $E[\varepsilon_t\varepsilon_t'\otimes e_{x,t}e_{x,t}']$ in (iii) is $E\left[\varepsilon_{i,t}\varepsilon_{j,t}e_{k,x,t}e_{l,x,t}\right]$ which by Assumption 1(i) and Lemma B.1(i,ii,vi) is finite and zero if one of the indexes $i,j,k,$ and l is different from all others. When $i=k$ and $j=l\neq k$ we have $E\left[\varepsilon_{i,t}e_{i,x,t}\varepsilon_{j,t}e_{j,x,t}\right]=E\left[\varepsilon_{i,t}e_{i,x,t}\right]E\left[\varepsilon_{j,t}e_{j,x,t}\right]=\sigma_{i,0}^2$ because both of the last expectations are equal to $-\sigma_{i,0}$, as noted above, and similarly when $i=l$ and $j=k\neq l$. Finally, when $i=j\neq k=l$ we have $E[\varepsilon_{i,t}^2e_{k,x,t}^2]=E[\varepsilon_{i,t}^2]E[e_{k,x,t}^2]=\sigma_{i,0}^2E[e_{k,x,t}^2]$, so that altogether we have

$$E\left[\varepsilon_{i,t}\varepsilon_{j,t}e_{k,x,t}e_{l,x,t}\right]=\begin{cases}\sigma_{i,0}^2, & i=k,\ j=l\neq k\ \text{ or }\ i=l,\ j=k\neq k,\\ E[\varepsilon_{i,t}^2e_{i,x,t}^2], & i=j=k=l,\\ \sigma_{i,0}^2E[e_{k,x,t}^2], & i=j\neq k=l,\\ 0, & \text{otherwise.}\end{cases}$$

Finiteness of the moments appearing in this expression, as well as that in (v), is ensured by Assumption 1(i) and Lemma B.1(ii,vi).

Proof of Lemma 1. We have demonstrated above that $\{l_{\theta,t}(\theta_0), \mathcal{F}_t\}$ is a martingale difference sequence with a finite covariance matrix. By Assumption 4(v), this covariance matrix is positive definite. As a (measurable) function of the IID sequence ε_t, the process $l_{\theta,t}(\theta_0)$ is also stationary and ergodic, and hence the central limit theorem of Billingsley (1961) (in conjunction with the Cramér-Wold device) implies the stated asymptotic normality. ■

Appendix C: Technical details for Section 4.3

Expression for the Hessian matrix. In accordance with the partition of θ as $\theta = (\pi, \beta, \sigma, \lambda)$, we will denote the 16 blocks of the Hessian matrix with $l_{\pi\pi,t}(\theta) = \frac{\partial^2 l_t(\theta)}{\partial\pi\partial\pi'}$, $l_{\pi\beta,t}(\theta) = \frac{\partial^2 l_t(\theta)}{\partial\pi\partial\beta'}$, etc. Let us summarize what form the 16 blocks of the Hessian $l_{\theta\theta,t}(\theta)$ take. To simplify notation define, for $i = 1, \ldots, n$, the quantities

$$\begin{aligned}
e_{i,xx,t}(\theta) &= \frac{f_{i,xx}(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)} - \left(\frac{f_{i,x}(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}\right)^2, \\
e_{i,x\lambda_i,t}(\theta) &= \frac{f_{i,x\lambda_i}(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)} - \frac{f_{i,x}(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}\frac{f_{i,\lambda_i}'(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}, \\
e_{i,\lambda_i\lambda_i,t}(\theta) &= \frac{f_{i,\lambda_i\lambda_i}(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)} - \frac{f_{i,\lambda_i}(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}\frac{f_{i,\lambda_i}'(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i'B^{-1}u_t(\pi);\lambda_i)},
\end{aligned}$$

and use these to form the diagonal / block diagonal matrices

$$\begin{aligned}
e_{xx,t}(\theta) &= diag\left(e_{1,xx,t}(\theta), \ldots, e_{n,xx,t}(\theta)\right) \quad (n \times n), \\
e_{\lambda\lambda,t}(\theta) &= diag\left(e_{1,\lambda_1\lambda_1,t}(\theta), \ldots, e_{n,\lambda_n\lambda_n,t}(\theta)\right) \quad (d \times d), \\
e_{x\lambda,t}(\theta) &= diag(e_{1,x\lambda_1,t}(\theta), \ldots, e_{n,x\lambda_n,t}(\theta)) \quad (n \times d).
\end{aligned}$$

Also define the diagonal matrices

$$\begin{aligned}
E_{x,t}(\theta) &= diag\left(e_{1,x,t}(\theta), \ldots, e_{n,x,t}(\theta)\right) \quad (n \times n), \\
\mathcal{E}_t(\theta) &= diag\left(\varepsilon_{1,t}(\theta), \ldots, \varepsilon_{n,t}(\theta)\right) \quad (n \times n),
\end{aligned}$$

and let K_{nn} $(n^2 \times n^2)$ denote the commutation matrix (satisfying $K_{nn}vec(A) = vec(A')$ for any $n \times n$ matrix A). Now, straightforward but tedious differentiation (details available in the

Supplementary Appendix) yields the different blocks of $l_{\theta\theta,t}(\theta)$ as

$$
\begin{aligned}
l_{\pi\pi,t}(\theta) &= (I_n \otimes B^{-1\prime}\Sigma^{-1})(x_{t-1}x_{t-1}' \otimes e_{xx,t}(\theta))(I_n \otimes \Sigma^{-1}B^{-1}), \\
l_{\pi\beta,t}(\theta) &= x_{t-1} \otimes [(I_n \otimes e_{x,t}'(\theta))(B^{-1\prime} \otimes \Sigma^{-1}B^{-1})H] \\
&\quad + x_{t-1} \otimes [B^{-1\prime}\Sigma^{-1}(u_t'(\pi) \otimes e_{xx,t}(\theta))(B^{-1\prime} \otimes \Sigma^{-1}B^{-1})H], \\
l_{\beta\beta,t}(\theta) &= H'(B^{-1} \otimes B^{-1\prime}\Sigma^{-1})(u_t(\pi)\, u_t'(\pi) \otimes e_{xx,t}(\theta))(B^{-1\prime} \otimes \Sigma^{-1}B^{-1})H \\
&\quad + H'(B^{-1} \otimes I_n)\left(u_t(\pi)\, e_{x,t}'(\theta) \otimes I_n\right)(\Sigma^{-1}B^{-1} \otimes B^{-1\prime})K_{nn}H \\
&\quad + H'K_{nn}(B^{-1\prime}\Sigma^{-1} \otimes B^{-1})\,(e_{x,t}(\theta)\, u_t'(\pi) \otimes I_n)\,(B^{-1\prime} \otimes I_n)H \\
&\quad + H'(B^{-1} \otimes B^{-1\prime})K_{nn}H, \\
l_{\pi\sigma,t}(\theta) &= x_{t-1} \otimes B^{-1\prime}\left[\Sigma^{-2}E_{x,t}(\theta) + \Sigma^{-3}e_{xx,t}(\theta)\,\mathcal{E}_t(\theta)\right], \\
l_{\beta\sigma,t}(\theta) &= H'(B^{-1} \otimes B^{-1\prime})(u_t(\pi) \otimes \left[\Sigma^{-2}E_{x,t}(\theta) + \Sigma^{-3}e_{xx,t}(\theta)\,\mathcal{E}_t(\theta)\right]), \\
l_{\sigma\sigma,t}(\theta) &= \Sigma^{-2} + 2\Sigma^{-3}\mathcal{E}_t(\theta)E_{x,t}(\theta) + \Sigma^{-4}\mathcal{E}_t^2(\theta)e_{xx,t}(\theta), \\
l_{\pi\lambda,t}(\theta) &= -(I_{np+1} \otimes B^{-1\prime}\Sigma^{-1})(x_{t-1} \otimes e_{x\lambda,t}(\theta)), \\
l_{\beta\lambda,t}(\theta) &= -H'(B^{-1} \otimes B^{-1\prime}\Sigma^{-1})(u_t(\pi) \otimes e_{x\lambda,t}(\theta)), \\
l_{\sigma\lambda,t}(\theta) &= -\Sigma^{-2}\mathcal{E}_t(\theta)e_{x\lambda,t}(\theta), \\
l_{\lambda\lambda,t}(\theta) &= e_{\lambda\lambda,t}(\theta).
\end{aligned}
$$

Proof of Lemma 2. Regarding the uniform convergence of the Hessian, from the stationarity and ergodicity of y_t and the expressions of the components of $l_{\theta\theta,t}(\theta)$ at the beginning of this Appendix it follows that $l_{\theta\theta,t}(\theta)$ forms a stationary ergodic sequence of random variables that are continuous in θ over Θ_0. The desired result thus follows (see, e.g., Ranga Rao (1962)) if we establish that $E\left[\sup_{\theta\in\Theta_0}\|l_{\theta\theta,t}(\theta)\|\right]$ is finite or that the corresponding result holds for the (matrix) components of $l_{\theta\theta,t}(\theta)$. In light of the expression of $l_{\theta\theta,t}(\theta)$ and the definition of Θ in Assumption 3, it suffices to show that the following condition holds:

$E[\sup_{\theta\in\Theta_0}\|*\|]$ is finite whenever $*$ is replaced by any of the following expressions:
$x_{t-1}x_{t-1}' \otimes e_{xx,t}(\theta)$, $x_{t-1} \otimes I_n \otimes e_{x,t}'(\theta)$, $x_{t-1} \otimes u_t'(\pi) \otimes e_{xx,t}(\theta)$, $u_t(\pi)\, u_t'(\pi) \otimes e_{xx,t}(\theta)$,
$u_t(\pi)\, e_{x,t}'(\theta) \otimes I_n$, $x_{t-1} \otimes E_{x,t}(\theta)$, $x_{t-1} \otimes e_{xx,t}(\theta)\,\mathcal{E}_t(\theta)$, $u_t(\pi) \otimes E_{x,t}(\theta)$,
$u_t(\pi) \otimes e_{xx,t}(\theta)\,\mathcal{E}_t(\theta)$, $\mathcal{E}_t(\theta)E_{x,t}(\theta)$, $\mathcal{E}_t^2(\theta)e_{xx,t}(\theta)$, $x_{t-1} \otimes e_{x\lambda,t}(\theta)$,
$u_t(\pi) \otimes e_{x\lambda,t}(\theta)$, $\mathcal{E}_t(\theta)e_{x\lambda,t}(\theta)$, $e_{\lambda\lambda,t}(\theta)$.

By submultiplicativity and the property $\|U \otimes V\| = \|U\| \|V\|$ of the Euclidean matrix norm (for any matrices U and V), it suffices to show that the following condition holds:

$$E[\sup_{\theta\in\Theta_0} * \,] \text{ is finite whenever } * \text{ is replaced by any of the following expressions:} \quad (15)$$
$$\|x_{t-1}\|^2 \|e_{xx,t}(\theta)\|, \quad \|x_{t-1}\| \|e_{x,t}(\theta)\|, \quad \|x_{t-1}\| \|u_t(\pi)\| \|e_{xx,t}(\theta)\|, \quad \|u_t(\pi)\|^2 \|e_{xx,t}(\theta)\|,$$
$$\|u_t(\pi)\| \|e_{x,t}(\theta)\|, \quad \|x_{t-1}\| \|E_{x,t}(\theta)\|, \quad \|x_{t-1}\| \|e_{xx,t}(\theta)\| \|\mathcal{E}_t(\theta)\|, \quad \|u_t(\pi)\| \|E_{x,t}(\theta)\|,$$
$$\|u_t(\pi)\| \|e_{xx,t}(\theta)\| \|\mathcal{E}_t(\theta)\|, \quad \|\mathcal{E}_t(\theta)\| \|E_{x,t}(\theta)\|, \quad \|\mathcal{E}_t(\theta)\|^2 \|e_{xx,t}(\theta)\|, \quad \|x_{t-1}\| \|e_{x\lambda,t}(\theta)\|,$$
$$\|u_t(\pi)\| \|e_{x\lambda,t}(\theta)\|, \quad \|\mathcal{E}_t(\theta)\| \|e_{x\lambda,t}(\theta)\|, \quad \|e_{\lambda\lambda,t}(\theta)\|.$$

By the definitions of $e_{i,xx,t}(\theta)$, $e_{i,x,t}(\theta)$, $e_{i,x\lambda_i,t}(\theta)$, and $e_{i,\lambda_i\lambda_i,t}(\theta)$ and Assumption 5(iii), for some $C < \infty$ and for all $i = 1, \ldots, n$ and all $\theta \in \Theta_0$,

$$\begin{aligned}
|e_{i,x,t}(\theta)|, \quad e_{i,x,t}^2(\theta), \quad |e_{i,xx,t}(\theta)| &\le C(1 + \|u_t(\pi)\|^{a_1}), \\
\|e_{i,x\lambda_i,t}(\theta)\| &\le C(1 + \|u_t(\pi)\|^{a_2}), \\
\|e_{i,\lambda_i\lambda_i,t}(\theta)\| &\le C(1 + \|u_t(\pi)\|^{a_3}).
\end{aligned}$$

On the other hand, by the definitions of $u_t(\pi)$, $\varepsilon_{i,t}(\theta)$ $(i = 1, \ldots, n)$, and $x_{t-1} = (1, y_{t-1}, \ldots, y_{t-p})$, for some $C < \infty$ and for all $\theta \in \Theta_0$,

$$\begin{aligned}
\|u_t(\pi)\| &\le C(1 + \sum_{j=0}^{p} \|y_{t-j}\|), \qquad |\varepsilon_{i,t}(\theta)| \le \left\|\iota_i' B(\beta)^{-1}\right\| \|u_t(\pi)\| \le C(1 + \sum_{j=0}^{p} \|y_{t-j}\|), \\
\|x_{t-1}\| &\le 1 + \sum_{j=1}^{p} \|y_{t-j}\|, \qquad \text{and} \quad \|x_{t-1}\|^2 = 1 + \sum_{j=1}^{p} \|y_{t-j}\|^2.
\end{aligned}$$

Consequently by Loève's c_r-inequality, for any fixed $r > 0$ and some $C < \infty$,

$$\|u_t(\pi)\|^r \le C(1 + \sum_{j=0}^{p} \|y_{t-j}\|^r).$$

Combining the results above, it can be shown that condition (15) holds as long as $E[\|y_t\|^{2+a_1} + \|y_t\|^{1+a_2} + \|y_t\|^{a_3}] < \infty$. This, in turn, holds if $E[|\varepsilon_{i,t}|^r] < \infty$ for $r = 2 + a_1, 1 + a_2, a_3$ and all $i = 1, \ldots, n$, which is ensured by Assumption 5(iii).

Finally, using Assumptions 5(i) and (ii) (and the earlier assumptions) the identity $E[l_{\theta\theta,t}(\theta_0)] = -E\left[l_{\theta,t}(\theta_0)\, l_{\theta,t}'(\theta_0)\right]$ can be established with straightforward but quite tedious and uninteresting matrix algebra. For brevity, we omit the details, which are available in the Supplementary Appendix. ■

Appendix D: Technical details for Section 4.4

Proof of Theorem 1. Existence of a consistent root. We first show that there exists a sequence of solutions $\hat{\theta}_T$ to the likelihood equations $L_{\theta,T}(\theta) = 0$ that are strongly consistent for θ_0. To this end, choose a small fixed $\epsilon > 0$ such that the sphere $\Theta_\epsilon = \{\theta : \|\theta - \theta_0\| = \epsilon\}$ is contained in Θ_0. We will compare the values attained by $L_T(\theta)$ on this sphere with $L_T(\theta_0)$. For an arbitrary point $\theta \in \Theta_\epsilon$, using a second-order Taylor expansion around θ_0 and adding and subtracting terms yields

$$\begin{aligned} L_T(\theta) - L_T(\theta_0) &= (\theta - \theta_0)' L_{\theta,T}(\theta_0) + \frac{1}{2}(\theta - \theta_0)' [L_{\theta\theta,T}(\theta_\bullet) - E[l_{\theta\theta,t}(\theta_\bullet)]] (\theta - \theta_0) \\ &\quad + \frac{1}{2}(\theta - \theta_0)' [E[l_{\theta\theta,t}(\theta_\bullet)] - E[l_{\theta\theta,t}(\theta_0)]] (\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' E[l_{\theta\theta,t}(\theta_0)] (\theta - \theta_0) \\ &= S_1 + S_2 + S_3 + S_4, \end{aligned}$$

where $\theta_\bullet$ lies on the line segment between θ and θ_0, and the latter equality defines the terms S_i, $i = 1, \ldots, 4$. It can be shown that, for any sufficiently small fixed ϵ, $\sup_{\theta\in\Theta_\epsilon}(S_1 + S_2) \to 0$ a.s. as $T \to \infty$ (for S_1 this follows from the fact that $L_{\theta,T}(\theta_0) \to 0$ a.s. as $T \to \infty$; for S_2 the result is obtained making use of Lemma 2). The terms S_3 and S_4 do not depend on T, and it can be shown that there exists a positive δ such that for each sufficiently small ϵ, $\sup_{\theta\in\Theta_\epsilon}(S_3 + S_4) < -\delta\epsilon^2$ (for S_3 the needed arguments are obtained from Lemma 2 and the continuity of $E[l_{\theta\theta,t}(\theta)$ mentioned therein; for S_4 one can invoke the fact that $E[l_{\theta\theta,t}(\theta_0)]$ is negative definite due to Lemmas 1 and 2). Therefore, for each sufficiently small ϵ,

$$\sup_{\theta\in\Theta_\epsilon} L_T(\theta) < L_T(\theta_0) \text{ a.s. as } T \to \infty. \tag{16}$$

As a consequence, for each fixed sufficiently small ϵ, and for all T sufficiently large, $L_T(\theta)$ must have a local maximum, and hence a root of the likelihood equation $L_{\theta,T}(\theta) = 0$, in the interior of Θ_ϵ with probability one. Having established this, the existence of a sequence $\hat{\theta}_T$, independent of ϵ, such that the $\hat{\theta}_T$ are solutions of the likelihood equations $L_{\theta,T}(\theta) = 0$ for all sufficiently large T and that $\hat{\theta}_T \to \theta_0$ a.s. as $T \to \infty$ can be shown as in Serfling (1980, pp. 147–148).

Asymptotic Normality. By a standard mean value expansion of the score vector $L_{\boldsymbol{\theta},T}(\theta)$,

$$T^{1/2} L_{\theta,T}(\hat{\theta}_T) = T^{1/2} L_{\theta,T}(\theta_0) + \dot{L}_{\theta\theta,T} T^{1/2}(\hat{\theta}_T - \theta_0) \text{ a.s.}, \tag{17}$$

where $\dot{L}_{\theta\theta,T}$ signifies the matrix $L_{\theta\theta,T}(\theta)$ with each row evaluated at an intermediate point $\dot{\theta}_{i,T}$ $(i = 1, \ldots, \dim\theta)$ lying between $\hat{\theta}_T$ and θ_0. As shown above, $\hat{\theta}_T \to \theta_0$ a.s., so that

$\dot{\theta}_{i,T} \to \theta_0$ a.s. as $T \to \infty$ $(i = 1, \ldots, \dim \theta)$ which, together with the uniform convergence result for $L_{\theta\theta,T}(\theta)$ in Lemma 2, yields $\dot{L}_{\theta\theta,T} \to E[l_{\theta\theta,t}(\theta_0)]$ a.s. as $T \to \infty$. This and the invertibility of $E[l_{\theta\theta,t}(\theta_0)]$ obtained from Assumption 4(v) and the result $E[l_{\theta\theta,t}(\theta_0)] = -\mathcal{I}(\theta_0)$ established in Lemma 2 imply that, for all T sufficiently large, $\dot{L}_{\theta\theta,T}$ is also invertible (a.s.) and $\dot{L}_{\theta\theta,T}^{-1} \to E[l_{\theta\theta,t}(\theta_0)]^{-1}$ a.s. as $T \to \infty$. Multiplying the mean value expansion (17) with the Moore-Penrose inverse $\dot{L}_{\theta\theta,T}^{+}$ of $\dot{L}_{\theta\theta,T}$ (this inverse exists for all T) and rearranging we obtain

$$T^{1/2}(\hat{\theta}_T - \theta_0) = (I_{\dim\theta} - \dot{L}_{\theta\theta,T}^{+}\dot{L}_{\theta\theta,T})T^{1/2}(\hat{\theta}_T - \theta_0) + \dot{L}_{\theta\theta,T}^{+}T^{1/2}L_{\theta,T}(\hat{\theta}_T) - \dot{L}_{\theta\theta,T}^{+}T^{1/2}L_{\theta,T}(\theta_0). \quad (18)$$

The first two terms on the right hand side of (18) converge to zero a.s. (for the first term, this follows from the fact that for all T sufficiently large $\dot{L}_{\theta\theta,T}$ is invertible; for the second one, this holds because $\hat{\theta}_T$ being a maximizer of $L_T(\theta)$ and θ_0 being an interior point of Θ_0 yield $L_{\theta,T}(\hat{\theta}_T) = 0$ for all T sufficiently large). Furthermore, the eventual a.s. invertibility of $\dot{L}_{\theta\theta,T}$ also means that $\dot{L}_{\theta\theta,T}^{+} - E[l_{\theta\theta,t}(\theta_0)]^{-1} \to 0$ a.s. Hence, (18) becomes

$$T^{1/2}(\hat{\theta}_T - \theta_0) = o_1(1) - (E[l_{\theta\theta,t}(\theta_0)]^{-1} + o_2(1))T^{1/2}L_{\theta,T}(\theta_0),$$

where $o_1(1)$ and $o_2(1)$ (a vector- and a matrix-valued process, respectively) converge to zero a.s. Combining this with the result of Lemma 1 and the property $E[l_{\theta\theta,t}(\theta_0)] = -\mathcal{I}(\theta_0)$ (see Lemma 2) completes the proof. ■

Proof of Theorem 2. (i) Regarding the least squares estimator $\tilde{\pi}_T$, strong consistency and the result $T^{1/2}(\tilde{\pi}_T - \pi_0) = O_p(1)$ follow from the imposed assumptions by standard arguments.

Concerning $\tilde{\gamma}_T$, we first show that there exists a sequence of solutions $\tilde{\gamma}_T$ to the (likelihood-like) equations $\tilde{L}_{\gamma,T}(\gamma) = 0$ that are strongly consistent for γ_0. As in the proof of Theorem 1, choose a small fixed $\epsilon > 0$ such that the sphere $\Theta_\epsilon^\gamma = \{\gamma : \|\gamma - \gamma_0\| = \epsilon\}$ is contained in $\Theta_{0,\beta} \times \Theta_{0,\sigma} \times \Theta_{0,\lambda}$. For an arbitrary point $\gamma \in \Theta_\epsilon^\gamma$, by the definition of $\tilde{L}_T(\gamma)$ and by adding and subtracting terms, we have

$$\begin{aligned}
\tilde{L}_T(\gamma) - \tilde{L}_T(\gamma_0) &= [L_T(\tilde{\pi}_T, \gamma) - E[l_t(\pi_0, \gamma)]] + [E[l_t(\pi_0, \gamma)] - L_T(\pi_0, \gamma)] \\
&\quad + [L_T(\pi_0, \gamma) - L_T(\pi_0, \gamma_0)] \\
&\quad + [L_T(\pi_0, \gamma_0) - E[l_t(\pi_0, \gamma_0)]] + [E[l_t(\pi_0, \gamma_0)] - L_T(\tilde{\pi}_T, \gamma_0)] \\
&= U_1 + U_2 + U_3 + U_4 + U_5,
\end{aligned}$$

where the latter equality defines the terms $U_1, \ldots, U_5$. As $\tilde{\pi}_T$ is strongly consistent for π_0 and the $l_t(\theta)$ form a stationary ergodic sequence of random variables that are continuous in θ over

Θ_0, the terms U_1 and U_5 converge to zero a.s. as $T \to \infty$ if the uniform law of large numbers holds for $l_t(\theta)$, i.e., if $\sup_{\theta\in\Theta_0} \left| T^{-1} \sum_{t=1}^{T} l_t(\theta) - E[l_t(\theta)] \right| \to 0$ a.s. This in turn holds if (see, e.g., Ranga Rao (1962)) $E\left[\sup_{\theta\in\Theta_0} |l_t(\theta)|\right]$ is finite. By a mean value expansion of $l_t(\theta)$, the conditions $E[|l_t(\theta_0)|] < \infty$ and $E\left[\sup_{\theta\in\Theta_0} \|l_{\theta,t}(\theta)\|\right] < \infty$ suffice for this. The former condition holds due to the assumed continuity of $f_i(\cdot;\lambda_{i,0})$, $i = 1,\ldots,n$. In light of the expression of $l_{\theta,t}(\theta)$ in Appendix B and the definition of Θ in Assumption 3, it suffices to show that (cf. the proof of Lemma 2) the expressions

$$E[\|x_{t-1}\|],\ E\left[\sup_{\theta\in\Theta_0}\|e_{x,t}(\theta)\|\right],\ E\left[\sup_{\theta\in\Theta_0}\|u_t(\pi)\|\,\|e_{x,t}(\theta)\|\right],\ E\left[\sup_{\theta\in\Theta_0}\|\varepsilon_t(\theta)\|\,\|e_{x,t}(\theta)\|\right],\ E\left[\sup_{\theta\in\Theta_0}\|e_{\lambda,t}(\theta)\|\right]$$

are all finite. This follows from arguments already used in the proof of Lemma 2. The derivations above also ensure that the ergodic theorem applies to $l_t(\theta)$ for any fixed $\theta \in \Theta_0$, implying that also the terms U_2 and U_4 converge to zero a.s. as $T \to \infty$.

Concerning the term U_3, arguments similar to those in the proof of Theorem 1 can be used to show that, for each sufficiently small ϵ,

$$\sup_{\gamma\in\Theta_\epsilon^\gamma} L_T(\pi_0,\gamma) < L_T(\pi_0,\gamma_0) \text{ a.s. as } T \to \infty,$$

and the existence of a sequence of solutions $\tilde{\gamma}_T$ to the likelihood equations $\tilde{L}_{\gamma,T}(\gamma) = 0$ that are strongly consistent for γ_0 follows as in the proof of Theorem 1.

Concerning the root-T consistency of $\tilde{\gamma}_T$, mean value expansions of the functions $L_{\gamma,T}(\tilde{\pi}_T,\cdot)$ and $L_{\gamma,T}(\cdot,\gamma_0)$ yield

$$\begin{aligned} T^{1/2}L_{\gamma,T}(\tilde{\pi}_T,\tilde{\gamma}_T) &= T^{1/2}L_{\gamma,T}(\tilde{\pi}_T,\gamma_0) + \dot{L}_{\gamma\gamma,T}T^{1/2}(\tilde{\gamma}_T - \gamma_0) \text{ a.s.,} \\ T^{1/2}L_{\gamma,T}(\tilde{\pi}_T,\gamma_0) &= T^{1/2}L_{\gamma,T}(\pi_0,\gamma_0) + \dot{L}_{\gamma\pi,T}T^{1/2}(\tilde{\pi}_T - \pi_0) \text{ a.s.,} \end{aligned}$$

where $\dot{L}_{\gamma\gamma,T}$ (resp. $\dot{L}_{\gamma\pi,T}$) signifies the matrix $L_{\gamma\gamma,T}(\tilde{\pi}_T,\cdot)$ (resp. $L_{\gamma\pi,T}(\cdot,\gamma_0)$) with each row evaluated at an intermediate point $\dot{\gamma}_{i,T}$, $i = 1,\ldots,\dim\gamma$, lying between $\tilde{\gamma}_T$ and γ_0 (resp. $\dot{\pi}_{i,T}$, $i = 1,\ldots,\dim\pi$, lying between $\tilde{\pi}_T$ and π_0). Arguments similar to those used in the proof of Theorem 1 now yield

$$T^{1/2}(\tilde{\gamma}_T - \gamma_0) = -\dot{L}^+_{\gamma\gamma,T}T^{1/2}L_{\gamma,T}(\pi_0,\gamma_0) - \dot{L}^+_{\gamma\gamma,T}\dot{L}_{\gamma\pi,T}T^{1/2}(\tilde{\pi}_T - \pi_0) + o(1), \tag{19}$$

where $\dot{L}^+_{\gamma\gamma,T}$ denotes the Moore-Penrose inverse of $\dot{L}_{\gamma\gamma,T}$ and $o(1)$ ($\dim\gamma \times 1$) converges to zero a.s. By the strong consistency of $\tilde{\pi}_T$ and Lemmas 1 and 2, the first term on the right hand side of (19) converges in distribution to $N(0,\mathcal{I}_{\gamma\gamma}(\theta_0)^{-1})$. By the strong consistency of $\tilde{\pi}_T$, Lemma 2, finiteness and invertibility of $E[l_{\theta\theta,t}(\theta_0)]$, and the fact $T^{1/2}(\tilde{\pi}_T - \pi_0) = O_p(1)$, the second term on the right hand side of (19) is $O_p(1)$. This completes the proof of (i).

(ii) In light of the result (19) and the discussion following it, it suffices to show that the second term on the right hand side of (19) is $o_p(1)$. This is the case because $E[l_{\pi\gamma,t}(\theta_0)] = 0$, as we establish next. Due to the result $E[l_{\theta\theta,t}(\theta_0)] = -\mathcal{I}(\theta_0)$ in Lemma 2 and the expressions of the off-diagonal blocks of $\mathcal{I}(\theta_0) = Cov[l_{\theta,t}(\theta_0)]$ in Appendix B, it suffices to show that the moments $E[\varepsilon_t' \otimes e_{x,t}e_{x,t}']$, $E[e_{x,t}(\varepsilon_t \odot e_{x,t})']$, and $E[e_{x,t}e_{\lambda,t}']$ all equal zero. To this end, note that the elements of the matrices $E[\varepsilon_t' \otimes e_{x,t}e_{x,t}']$ and $E[e_{x,t}(\varepsilon_t \odot e_{x,t})']$ are obtained from

$$E\left[\varepsilon_{i,t}e_{j,x,t}e_{k,x,t}\right] = \begin{cases} E[\varepsilon_{i,t}e_{i,x,t}^2], & i=j=k \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad E\left[e_{i,x,t}\varepsilon_{j,t}e_{j,x,t}\right] = \begin{cases} E[\varepsilon_{i,t}e_{i,x,t}^2], & i=j \\ 0, & \text{otherwise} \end{cases},$$

respectively. The assumed symmetry and Lemma A.3 of Meitz and Saikkonen (2013) ensure that $E[\varepsilon_{i,t}e_{i,x,t}^2] = 0$, $i = 1, \ldots, n$. Regarding the moment $E[e_{x,t}e_{\lambda,t}']$, it suffices to show that $E[e_{i,x,t}e_{i,\lambda_i,t}] = 0$ for $i = 1, \ldots, n$. As

$$E\left[e_{i,x,t}e_{i,\lambda_i,t}\right] = E\left[\frac{f_{i,x}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}\frac{f_{i,\lambda_i}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}\right],$$

the desired result again follows from Lemma A.3 of Meitz and Saikkonen (2013) because if the distribution of $\varepsilon_{i,t}$ is symmetric in the sense that $f_i(x;\lambda_i) = f_i(-x;\lambda_i)$ for all $\lambda_i \in \Theta_{0,\lambda_i}$, the functions $f_i(\sigma_{i,0}^{-1}\cdot;\lambda_{i,0})$ and $f_{i,\lambda_i}(\sigma_{i,0}^{-1}\cdot;\lambda_{i,0})$ are symmetric functions (for the latter, this follows from $f_{i,\lambda_i}(\sigma_{i,0}^{-1}\cdot;\lambda_{i,0}) = \frac{\partial}{\partial\lambda_i}f_i(\sigma_{i,0}^{-1}\cdot;\lambda_{i,0})$ and the symmetry of $f_i(\sigma_{i,0}^{-1}\cdot;\lambda_i)$ for $\lambda_i \in \Theta_{0,\lambda_i}$) and the function $f_{i,x}(\sigma_{i,0}^{-1}\cdot;\lambda_{i,0})$ is an odd function.

(iii) By the definition of $\tilde{\theta}_T^{(2)}$ (and, for simplicity, assuming invertibility) we have $T^{1/2}(\tilde{\theta}_T^{(2)} - \tilde{\theta}_T) = -L_{\theta\theta,T}^{-1}(\tilde{\theta}_T)T^{1/2}L_{\theta,T}(\tilde{\theta}_T)$. A standard mean value expansion of $L_{\theta,T}(\theta)$ yields

$$T^{1/2}L_{\theta,T}(\tilde{\theta}_T) = T^{1/2}L_{\theta,T}(\theta_0) + \ddot{L}_{\theta\theta,T}T^{1/2}(\tilde{\theta}_T - \theta_0) \text{ a.s.},$$

where $\ddot{L}_{\theta\theta,T}$ signifies the matrix $L_{\theta\theta,T}(\theta)$ with each row evaluated at an intermediate point $\ddot{\theta}_{i,T}$ ($i = 1, \ldots, \dim\theta$) lying between $\tilde{\theta}_T$ and θ_0. Therefore,

$$\begin{aligned} T^{1/2}(\tilde{\theta}_T^{(2)} - \theta_0) &= T^{1/2}(\tilde{\theta}_T - \theta_0) - L_{\theta\theta}^{-1}(\tilde{\theta}_T)T^{1/2}L_{\theta,T}(\theta_0) - L_{\theta\theta}^{-1}(\tilde{\theta}_T)\ddot{L}_{\theta\theta,T}T^{1/2}(\tilde{\theta}_T - \theta_0) \\ &= (I_{\dim\theta} - L_{\theta\theta}^{-1}(\tilde{\theta}_T)\ddot{L}_{\theta\theta,T})T^{1/2}(\tilde{\theta}_T - \theta_0) - L_{\theta\theta}^{-1}(\tilde{\theta}_T)T^{1/2}L_{\theta,T}(\theta_0) \text{ a.s.} \end{aligned} \quad (20)$$

The first term on the right hand side of (20) converges to zero a.s. due to strong consistency of $\tilde{\theta}_T$, Lemma 2, and the result $T^{1/2}(\tilde{\theta}_T - \theta_0) = O_p(1)$ in part (i) of this theorem, whereas the second term converges in distribution to $N(0, \mathcal{I}(\theta_0)^{-1})$. ■

References

Anderson, B. D. O, M. Deistler, E. Felsenstein, B. Funovits, L. Koelbl, and M. Zamani (in press): "Multivariate AR systems and mixed frequency data: g-identifiability and estimation," *Econometric Theory*, forthcoming.

Andrews, B., R. A. Davis, and F. J. Breidt (2006): "Maximum likelihood estimation for all-pass time series models," *Journal of Multivariate Analysis*, 97, 1638–1659.

Billingsley, P. (1961): "The Lindeberg-Levy theorem for martingales," *Proceedings of the American Mathematical Society*, 12, 788–792.

Bjørnland, H. C., and K. Leitemo (2009): "Identifying the interdependence between US monetary policy and the stock market," *Journal of Monetary Economics*, 56, 275–282.

Bochnak, J., M. Coste, and M.-F. Roy (1998): *Real Algebraic Geometry.* Springer, Berlin.

Breidt, F. J., R. A. Davis, K.-S. Lii, and M. Rosenblatt (1991): "Maximum likelihood estimation for noncausal autoregressive processes," *Journal of Multivariate Analysis*, 36, 175–198.

Castelnuovo, E. (2013): "Monetary policy shocks and financial conditions: A Monte Carlo experiment," *Journal of International Money and Finance*, 32, 282–303.

Castelnuovo, E., and S. Nisticò (2010): "Stock market conditions and monetary policy in a DSGE model for the U.S.," *Journal of Economic Dynamics and Control*, 34, 1700–1731.

Chan, K.-S., and L. H. Ho (2004): "On the unique representation of non-Gaussian multivariate linear processes," Technical Report #341, University of Iowa, http://www.stat.uiowa.edu/files/stat/techrep/tr341.pdf.

Chan, K.-S., L. H. Ho, and H. Tong (2006): "A note on time-reversibility of multivariate linear processes," *Biometrika*, 93, 221–227.

Chen, A., and P. Bickel (2005): "Consistent independent component analysis and prewhitening," *IEEE Transactions on Signal Processing*, 53, 3625–3632.

Cheng, L., and Y. Jin (2013): "Asset prices, monetary policy, and aggregate fluctuations: An empirical investigation," *Economics Letters*, 119, 24–27.

Chib, S., and S. Ramamurthy (2014): "DSGE models with Student-t errors," *Econometric Reviews*, 33, 152–177.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary policy shocks: What have we learned and to what end?," in Taylor, J. B., and M. Woodford (eds.), Handbook of Macroeconomics, vol 1A. Elsevier, New York, pp. 65–148.

Comon, P. (1994): "Independent component analysis, A new concept?," *Signal Processing*, 36, 287–314.

Cúrdia, V., M. Del Negro, and D. L. Greenwald (2014): "Rare shocks, great recessions," *Journal of Applied Econometrics*, 29, 1031–1052.

Fry, R., and A. Pagan (2011): "Sign restrictions in structural vector autoregressions: A critical review," *Journal of Economic Literature*, 49, 938–960.

Gouriéroux, C., and A. Monfort (2014): "Revisiting identification and estimation in structural VARMA models," CREST Discussion Paper 2014-30.

Gouriéroux, C., and J.-M. Zakoïan (2014): "On uniqueness of moving average representations of heavy-tailed stationary processes," MPRA Paper No. 54907.

Hakkio, C. S., and W. R. Keeton (2009): "Financial stress: What it is, how can it be measured, and why does it matter?," Federal Reserve Bank of Kansas City Economic Review, Second Quarter, 5–50.

Hallin, M., and C. Mehta (in press): "R-estimation for asymmetric independent component analysis," *Journal of the American Statistical Association*, forthcoming.

Hannan, E. J., and M. Deistler (1988): *The Statistical Theory of Linear Systems.* Wiley, New York.

Hyvärinen, A., K. Zhang, S. Shimizu, and P. O. Hoyer (2010): "Estimation of a structural vector autoregression model using non-Gaussianity," *Journal of Machine Learning Research*, 11, 1709–1731.

Ilmonen, P., and D. Paindaveine (2011): "Semiparametrically efficient inference based on signed ranks in symmetric independent component models," *Annals of Statistics*, 39, 2448–2476.

Johansen, S. (1995): "Identifying restrictions of linear equations with applications to simultaneous equations and cointegration," *Journal of Econometrics*, 69, 111–132.

Kagan, A. M., Y. V. Linnik, and C. R. Rao (1973): *Characterization Problems in Mathematical Statistics.* Wiley, New York.

Kilian, L. (2013): "Structural vector autoregressions," in Hashimzade, N., and M. A. Thornton (eds.), *Handbook of Research Methods and Applications in Empirical Macroeconomics*, Cheltenham, U.K.: Edward Elgar, pp. 515–554.

Kohn, R. (1979): "Identification results for ARMAX structures," *Econometrica*, 47, 1295–1304.

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

Lanne, M., and H. Lütkepohl (2010): "Structural vector autoregressions with nonnormal residuals," *Journal of Business and Economic Statistics*, 28, 159–168.

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

Lanne, M., and P. Saikkonen (2011): "Noncausal autoregressions for economic time series," *Journal of Time Series Econometrics*, 3, Issue 3, Article 2.

Lastrapes, W. D. (1998): "International evidence on equity prices, interest rates and money," *Journal of International Money and Finance*, 17, 377–406.

Li, Y. D., T. B. Iscan, and K. Xu (2010): "The impact of monetary shocks on stock prices: Evidence from Canada and the United States," *Journal of International Money and Finance*, 29, 876–896.

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

Lütkepohl, H., and A. Netšunajev (2014a): "Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs," *Journal of Applied Econometrics*, 29, 479–496.

Lütkepohl, H., and A. Netšunajev (2014b): "Structural vector autoregressions with smooth transition in variances: The interaction between U.S. monetary policy and the stock market," DIW Discussion Paper 1388.

Meitz, M., and P. Saikkonen (2013): "Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity," *Journal of Multivariate Analysis*, 114, 227–255.

Moneta, A., D. Entner, P. O. Hoyer, and A. Coad (2013): "Causal inference by independent component analysis: theory and applications," *Oxford Bulletin of Economics and Statistics*, 75, 705–730.

Normandin, M., and L. Phaneuf (2004): "Monetary policy shocks: Testing identification conditions under time-varying conditional volatility," *Journal of Monetary Economics*, 51, 1217–1243.

Patelis, A. D. (1997): "Stock return predictability and the role of monetary policy," *Journal of Finance*, 52, 1951–1972.

Pham, D. T., and P. Garat (1997): "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Transactions on Signal Processing*, 45, 1712–1725.

Ranga Rao, R. (1962): "Relations between weak and uniform convergence of measures with applications," *Annals of Mathematical Statistics*, 33, 659–680.

Rapach, D. E. (2001): "Macro shocks and real stock prices," *Journal of Economics and Business*, 53, 5–26.

Rigobon, R. (2003): "Identification through heteroskedasticity," *Review of Economics and Statistics*, 85, 777–792.

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

Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Thorbecke, W. (1997): "On stock market returns and monetary policy," *Journal of Finance*, 52, 635–654.

Supplementary Appendix to 'Identification and estimation of non-Gaussian structural vector autoregressions' by Lanne, Meitz, and Saikkonen. (Not to be published.)

A1 Proofs of some results in Appendix A

Proof of Lemma A.2. The algebraic set

$$A = \cup_{i=1}^{s} \cap_{j=1}^{r_i} \{x \in \mathbb{R}^m : f_{i,j}(x) = 0\}$$

can always be written as

$$A = \cup_{i=1}^{s} \{x \in \mathbb{R}^m : f_i(x) = 0\} \overset{def}{=} \cup_{i=1}^{s} A_i$$

if one defines the (finite-order polynomial) functions $f_i(x)$ as $f_i(x) = \sum_{j=1}^{r_i} f_{i,j}^2(x)$ $(i = 1, \ldots, s)$, and where the latter equality defines the (algebraic) sets A_i. The assumption that A is a proper subset of $\mathbb{R}^m$ ensures that none of the $f_i(\cdot)$ is identically equal to zero. Now, as $A = \cup_{i=1}^{s} A_i$ and $\mathbb{R}^m \backslash A = \cap_{i=1}^{s} (\mathbb{R}^m \backslash A_i)$, it suffices to show that, for each $i = 1, \ldots, s$, A_i has Lebesgue measure zero in $\mathbb{R}^m$ and its complement $\mathbb{R}^m \backslash A_i$ in $\mathbb{R}^m$ is an open and dense subset of $\mathbb{R}^m$. Without loss of generality, consider A_1, the set of zeros of the finite-order polynomial $f_1(\cdot)$ in $\mathbb{R}^m$.

We next prove the following subresult: For an arbitrary finite-order polynomial $f(\cdot)$ in $\mathbb{R}^m$, not identically equal to zero, the set of zeros $\{x \in \mathbb{R}^m : f(x) = 0\}$ has Lebesgue measure zero in $\mathbb{R}^m$. If $m = 1$, the set of zeros of an arbitrary $f(\cdot)$ is a finite set of points, and hence of Lebesgue measure zero in $\mathbb{R}$. Proceed inductively and assume the statement holds in dimension $m - 1$. Now in dimension m, consider an arbitrary nonzero polynomial $f(x_1, \ldots, x_m)$ and, for ease of notation, let $\boldsymbol{x}_{m-1} = (x_1, \ldots, x_{m-1})$ denote a vector in $\mathbb{R}^{m-1}$. For any $\boldsymbol{x}_{m-1}$, define the sets

$$S_m(\boldsymbol{x}_{m-1}) = \{x_m \in \mathbb{R} : f(\boldsymbol{x}_{m-1}, x_m) = 0\} \quad \text{and} \quad S_m^c(\boldsymbol{x}_{m-1}) = \{x_m \in \mathbb{R} : f(\boldsymbol{x}_{m-1}, x_m) \neq 0\},$$

and, furthermore, define the sets

$$S_m = \cap_{\boldsymbol{x}_{m-1} \in \mathbb{R}^{m-1}} S_m(\boldsymbol{x}_{m-1}) \quad \text{and} \quad S_m^c = \cup_{\boldsymbol{x}_{m-1} \in \mathbb{R}^{m-1}} S_m^c(\boldsymbol{x}_{m-1}).$$

Then the sets S_m and S_m^c form a partition of $\mathbb{R}$. For each fixed $x \in S_m^c$, define the set

$$S_{m-1}(x) = \{\boldsymbol{x}_{m-1} \in \mathbb{R}^{m-1} : f(\boldsymbol{x}_{m-1}, x) = 0\}.$$

Note that, for each fixed $\boldsymbol{x}_{m-1} \in \mathbb{R}^{m-1}$, the set $S_m(\boldsymbol{x}_{m-1})$ has Lebesgue measure zero in $\mathbb{R}$ (as it is finite), and therefore also S_m has Lebesgue measure zero in $\mathbb{R}$. Moreover, for each fixed $x \in S_m^c$, the set $S_{m-1}(x)$ has Lebesgue measure zero in $\mathbb{R}^{m-1}$ (by the inductive hypothesis and noting that $f(\boldsymbol{x}_{m-1}, x)$ cannot be identically equal to zero because $x \in S_m^c$).

Next note that the set of zeros of $f(x_1, \ldots, x_{m-1}, x_m)$ is

$$\begin{aligned} Z_m &= \{(x_1, \ldots, x_{m-1}, x_m) \in \mathbb{R}^m : f(x_1, \ldots, x_{m-1}, x_m) = 0\} \\ &= (\mathbb{R}^{m-1} \times S_m) \cup (\cup_{x \in S_m^c} (S_{m-1}(x) \times \{x\})), \end{aligned}$$

where the two sets of the union are disjoint. Let $\mu_m(\cdot)$ signify the Lebesgue measure in dimension m and $1_S(\cdot)$ the indicator function that takes the value one when the argument belongs to the set S. Then,

$$\begin{aligned} \mu_m(Z_m) &= \mu_m(\mathbb{R}^{m-1} \times S_m) + \mu_m(\cup_{x \in S_m^c} (S_{m-1}(x) \times \{x\})) \\ &= \int_{\mathbb{R}^{m-1}} \left[\int_{\mathbb{R}} 1_{S_m}(x_m)\, d\mu_1(x_m) \right] d\mu_{m-1}(\boldsymbol{x}_{m-1}) \\ &\quad + \int_{\mathbb{R}} 1_{S_m^c}(x_m) \left[\int_{\mathbb{R}^{m-1}} 1_{S_{m-1}(x_m)}(\boldsymbol{x}_{m-1}) d\mu_{m-1}(\boldsymbol{x}_{m-1}) \right] d\mu_1(x_m), \end{aligned}$$

where the second equality is justified by the Tonelli-Fubini Theorem (see, e.g., Dudley (2002, pp. 137, 139)). In the last expression, the former integral is zero because the set S_m has Lebesgue measure zero in $\mathbb{R}$ and the same is true for the latter integral because, for each fixed $x \in S_m^c$, the set $S_{m-1}(x)$ has Lebesgue measure zero in $\mathbb{R}^{m-1}$. This proves the subresult.

The subresult shows that the set A_1 has Lebesgue measure zero in $\mathbb{R}^m$. Finally, regarding the complement $\mathbb{R}^m \backslash A_1$, it is open as the preimage of the open set $\mathbb{R}\backslash\{0\}$ in the continuous transformation $f_1(\cdot)$ and dense because the interior of A_1 is empty. ■

Proof of Lemma A.3. Let A be a proper semi-algebraic set with equality constraints and, without loss of generality, assume that $*_{i,1}$ is $=$ for all $i = 1, \ldots, s$. Then the set

$$\tilde{A} = \cup_{i=1}^{s} \tilde{A}_i = \cup_{i=1}^{s} \{x \in \mathbb{R}^m : f_{i,1}(x) = 0\},$$

where the sets $\tilde{A}_i$ are defined via the latter equality, is a proper algebraic set with the properties $A \subseteq \tilde{A}$ and $\mathbb{R}^m \backslash A \supseteq \mathbb{R}^m \backslash \tilde{A}$. The result now follows from Lemma A.2. ■

Proof of Lemma A.4. First some preliminaries. Note that the determinant of a matrix is a polynomial function of the elements of the matrix, and a matrix is noninvertible if the determinant equals zero. Next consider the transformation $B \to BD_1$ (for $B \in \mathcal{M}_n$) and note in this transformation the elements b_{ij} of B are mapped as follows:

$$b_{ij} \to \frac{b_{ij}}{(\sum b_{\cdot j}^2)^{1/2}} \overset{def}{=} \tilde{b}_{ij}$$

where for brevity we denote $\sum b_{\cdot j}^2 = \sum_{k=1}^n b_{kj}^2$ (and similarly below). All indices (i, j, etc.) are assumed to belong to $\{1, \ldots, n\}$ and this is tacitly assumed in what follows. In step (ii) of the identification scheme, the absolute values $|\tilde{b}_{ij}|$ and $|\tilde{b}_{ik}|$ are compared for some i and $j \neq k$. Equivalently, one may compare $\tilde{b}_{ij}^2$ and $\tilde{b}_{ik}^2$, that is,

$$\frac{b_{ij}^2}{\sum b_{\cdot j}^2} \quad \text{and} \quad \frac{b_{ik}^2}{\sum b_{\cdot k}^2}.$$

Note that any statement of the form '$\tilde{b}_{ij}^2 * \tilde{b}_{ik}^2$', where $*$ is either $=$, $<$, $>$, or $\neq$, can equivalently be expressed as '$b_{ij}^2 \sum b_{\cdot k}^2 - b_{ik}^2 \sum b_{\cdot j}^2 * 0$' (note that the sums appearing here are always positive as $B \in \mathcal{M}_n$). Therefore, recalling the note about the determinant at the beginning of the proof, statements like '$\det(B) * 0$' or '$\tilde{b}_{ij}^2 * \tilde{b}_{ik}^2$' (with $*$ interpreted as before) can equivalently be seen as statements about a polynomial in n^2 variables, the variables being the elements of the matrix B. As a last preparatory comment, in what follows, $(\tilde{1}, \ldots, \tilde{n})$ will always denote some permutation of $(1, \ldots, n)$.

Now, let us describe the set $\mathcal{E} = \mathcal{M}_n \backslash \mathcal{I}$ as a subset of $\mathcal{M}_n^A$, starting with dimension $n = 2$. The set

$$E_{2,1} = \{B \in \mathcal{M}_2 : \tilde{b}_{11}^2 = \tilde{b}_{12}^2\} = \{B \in \mathcal{M}_2^A : \det(B) \neq 0, \quad \tilde{b}_{11}^2 = \tilde{b}_{12}^2\}$$

consists of those matrices $B \in \mathcal{M}_2$ for which a tie occurs (on the first row) in step (ii) of the Identification Scheme. Furthermore, although (in the two-dimensional case) the following set is empty, to clarify the argument define the set

$$E_{2,2} = \{B \in \mathcal{M}_2 : \exists(\tilde{1}, \tilde{2}) \text{ such that } \tilde{b}_{1\tilde{1}}^2 > \tilde{b}_{1\tilde{2}}^2; \text{ and } \tilde{b}_{2\tilde{2}} = 0\},$$

that is, the set of those $B \in \mathcal{M}_2$ for which step (ii) can be done but in step (iii) a zero would occur at the lower right-hand corner of the matrix C (in the two-dimensional case this is impossible). Note that $E_{2,2}$ could alternatively be expressed as the union of the sets

$$\begin{aligned} E_{2,2,1} &= \{B \in \mathcal{M}_2^A : \det(B) \neq 0, \quad \tilde{b}_{11}^2 > \tilde{b}_{12}^2 \text{ and } \tilde{b}_{22} = 0\}, \\ E_{2,2,2} &= \{B \in \mathcal{M}_2^A : \det(B) \neq 0, \quad \tilde{b}_{12}^2 > \tilde{b}_{11}^2 \text{ and } \tilde{b}_{21} = 0\}. \end{aligned}$$

Therefore, in the case $n = 2$,

$$\mathcal{E} = \mathcal{M}_2 \backslash \mathcal{I} = E_{2,1} \cup E_{2,2} = E_{2,1} \cup E_{2,2,1} \cup E_{2,2,2},$$

a proper semi-algebraic set with equality constraints of $\mathcal{M}_2^A$.

Now consider the case $n = 3$. The set of those matrices $B \in \mathcal{M}_3$ for which a tie occurs in step (ii) on the first row can be expressed as

$$E_{3,1} = \{B \in \mathcal{M}_3 : \exists(\tilde{1}, \tilde{2}, \tilde{3}) \text{ such that } \tilde{b}^2_{1\tilde{1}} = \tilde{b}^2_{1\tilde{2}} \geq \tilde{b}^2_{1\tilde{3}}\}.$$

(This could equivalently be expressed as

$$E_{3,1} = \{B \in \mathcal{M}_3^A : \det(B) \neq 0, \quad \exists(\tilde{1}, \tilde{2}, \tilde{3}) \text{ such that } \tilde{b}^2_{1\tilde{1}} = \tilde{b}^2_{1\tilde{2}} \geq \tilde{b}^2_{1\tilde{3}}\}$$

but as this latter way is longer we won't bother writing this explicitly anymore in what follows.) Furthermore, the set $E_{3,1}$ can be expressed as the union of the (disjoint) sets

$$\begin{aligned} E_{3,1,1} &= \{B \in \mathcal{M}_3 : \tilde{b}^2_{11} = \tilde{b}^2_{12} = \tilde{b}^2_{13}\}, \\ E_{3,1,2} &= \{B \in \mathcal{M}_3 : \tilde{b}^2_{11} = \tilde{b}^2_{12} > \tilde{b}^2_{13}\}, \\ E_{3,1,3} &= \{B \in \mathcal{M}_3 : \tilde{b}^2_{11} = \tilde{b}^2_{13} > \tilde{b}^2_{12}\}, \\ E_{3,1,4} &= \{B \in \mathcal{M}_3 : \tilde{b}^2_{12} = \tilde{b}^2_{13} > \tilde{b}^2_{11}\}. \end{aligned}$$

Next, the set of those matrices $B \in \mathcal{M}_3$ for which a tie occurs in step (ii) on the second row (that is, no tie on the first row) can be expressed as

$$E_{3,2} = \{B \in \mathcal{M}_3 : \exists(\tilde{1}, \tilde{2}, \tilde{3}) \text{ such that } \tilde{b}^2_{1\tilde{1}} > \max\{\tilde{b}^2_{1\tilde{2}}, \tilde{b}^2_{1\tilde{3}}\}; \tilde{b}^2_{2\tilde{2}} = \tilde{b}^2_{2\tilde{3}}\}.$$

Again, $E_{3,2}$ can be expressed as the union of the (disjoint) sets

$$\begin{aligned} E_{3,2,1} &= \{B \in \mathcal{M}_3 : \tilde{b}^2_{11} > \tilde{b}^2_{12},\ \tilde{b}^2_{11} > \tilde{b}^2_{13},\ \tilde{b}^2_{22} = \tilde{b}^2_{23}\} \\ E_{3,2,2} &= \{B \in \mathcal{M}_3 : \tilde{b}^2_{12} > \tilde{b}^2_{11},\ \tilde{b}^2_{12} > \tilde{b}^2_{13},\ \tilde{b}^2_{21} = \tilde{b}^2_{23}\} \\ E_{3,2,3} &= \{B \in \mathcal{M}_3 : \tilde{b}^2_{13} > \tilde{b}^2_{11},\ \tilde{b}^2_{13} > \tilde{b}^2_{12},\ \tilde{b}^2_{21} = \tilde{b}^2_{22}\} \end{aligned}$$

Finally, the set of those matrices $B \in \mathcal{M}_3$ for which no tie occurs in step (ii) but for which a zero occurs at the lower right-hand corner of the matrix C can be expressed as

$$E_{3,3} = \{B \in \mathcal{M}_3 : \exists(\tilde{1}, \tilde{2}, \tilde{3}) \text{ such that } \tilde{b}^2_{1\tilde{1}} > \max\{\tilde{b}^2_{1\tilde{2}}, \tilde{b}^2_{1\tilde{3}}\};\ \tilde{b}^2_{2\tilde{2}} > \tilde{b}^2_{2\tilde{3}};\ \tilde{b}^2_{3\tilde{3}} = 0\}.$$

Again, $E_{3,3}$ can be expressed as the union of the (disjoint) sets

$$\begin{aligned}
E_{3,3,1} &= \{B \in \mathcal{M}_3 : \tilde{b}_{11}^2 > \tilde{b}_{12}^2,\ \tilde{b}_{11}^2 > \tilde{b}_{13}^2,\ \tilde{b}_{22}^2 > \tilde{b}_{23}^2,\ \tilde{b}_{33}^2 = 0\}, \\
E_{3,3,2} &= \{B \in \mathcal{M}_3 : \tilde{b}_{11}^2 > \tilde{b}_{12}^2,\ \tilde{b}_{11}^2 > \tilde{b}_{13}^2,\ \tilde{b}_{23}^2 > \tilde{b}_{22}^2,\ \tilde{b}_{32}^2 = 0\}, \\
E_{3,3,3} &= \{B \in \mathcal{M}_3 : \tilde{b}_{12}^2 > \tilde{b}_{11}^2,\ \tilde{b}_{12}^2 > \tilde{b}_{13}^2,\ \tilde{b}_{21}^2 > \tilde{b}_{23}^2,\ \tilde{b}_{33}^2 = 0\}, \\
E_{3,3,4} &= \{B \in \mathcal{M}_3 : \tilde{b}_{12}^2 > \tilde{b}_{11}^2,\ \tilde{b}_{12}^2 > \tilde{b}_{13}^2,\ \tilde{b}_{23}^2 > \tilde{b}_{21}^2,\ \tilde{b}_{31}^2 = 0\}, \\
E_{3,3,5} &= \{B \in \mathcal{M}_3 : \tilde{b}_{13}^2 > \tilde{b}_{11}^2,\ \tilde{b}_{13}^2 > \tilde{b}_{12}^2,\ \tilde{b}_{21}^2 > \tilde{b}_{22}^2,\ \tilde{b}_{32}^2 = 0\}, \\
E_{3,3,6} &= \{B \in \mathcal{M}_3 : \tilde{b}_{13}^2 > \tilde{b}_{11}^2,\ \tilde{b}_{13}^2 > \tilde{b}_{12}^2,\ \tilde{b}_{22}^2 > \tilde{b}_{21}^2,\ \tilde{b}_{31}^2 = 0\}.
\end{aligned}$$

Therefore, in the case $n = 3$,

$$\mathcal{E} = \mathcal{M}_3 \backslash \mathcal{I} = E_{3,1} \cup E_{3,2} \cup E_{3,3},$$

a finite union of (disjoint) sets each of which is a proper semi-algebraic set with equality constraints of $\mathcal{M}_3^A$ (because each of $E_{3,1}, E_{3,2}, E_{3,3}$ can be expressed as a finite union of (disjoint) sets, each of which is a proper semi-algebraic set with equality constraints of $\mathcal{M}_3^A$).

The preceding calculations for the cases $n = 2$ and $n = 3$ could be repeated for any $n \geq 2$. Without going into details, we just note that the set of those matrices $B \in \mathcal{M}_n$ for which ties occur in step (ii) of the Identification Scheme or a zero occurs in the lower right hand corner of C can be described by equalities and inequalities involving the row-wise elements of permuted versions of the matrix BD_1 with at least one equality occurring on each row. Hence, for any $n \geq 2$, the set $\mathcal{E} = \mathcal{M}_n \backslash \mathcal{I}$ can be expressed as a finite union of (disjoint) sets each of which is a proper semi-algebraic set with equality constraints of $\mathcal{M}_n^A$. ■

A2 Technical details for Section 4.1

Justification for the openness of the set Θ_β in Assumption 3. First note that the openness of Θ_β cannot be directly deduced from Proposition 2(c) or Lemma A.3 because the statement 'contains an open ... subset' therein cannot be strengthened to 'is an open ... subset'. Now, in order to show that Θ_β is an open subset of $\mathbb{R}^{n(n-1)}$, pick an arbitrary $\beta_\bullet \in \Theta_\beta$. We need to show that there exists a neighborhood of $\beta_\bullet$ in $\mathbb{R}^{n(n-1)}$ that is contained in Θ_β. The point $\beta_\bullet$ corresponds to a $B_\bullet \in \mathcal{B}$. Let $\mathcal{M}_n^R$ denote the set of all $n \times n$ matrices whose diagonal elements are all restricted to unity. We may equally well show that there exists a neighborhood of $B_\bullet$ in $\mathcal{M}_n^R$ that is contained in $\mathcal{B}$.

As a first step, we show that there exists a neighborhood of $B_\bullet$ in $\mathcal{M}_n^R$ that is contained in $\mathcal{I}$. First note that (due to the continuity of the determinant function) we can find a neighborhood of $B_\bullet$ in $\mathcal{M}_n^R$ in which all matrices are invertible. Next note that, as $B_\bullet \in \mathcal{B}$, it is necessarily the case that $\Pi(B_\bullet) = B_\bullet$ (because $\Pi(B_\bullet) \in \mathcal{B}$ and $\Pi(B_\bullet) \sim B_\bullet$, $\Pi(B_\bullet) \neq B_\bullet$ would violate Proposition 2(b)). Hence, the matrices D_1, P, and D_2 in the Identification Scheme (for the matrix $B_\bullet$) must satisfy $P = I_n$ and $D_1 D_2 = I_n$ ($\Pi(B_\bullet) = B_\bullet$ implies $D_1 P D_2 = I_n$, so that $P = D_1^{-1} D_2^{-1}$ necessarily equals I_n). To avoid confusion below, make the dependence of the matrices D_1, P, and D_2 on $B_\bullet$ explicit and express these results as $P(B_\bullet) = I_n$ and $D_1(B_\bullet) D_2(B_\bullet) = I_n$.

Next note that in the Identification Scheme for an arbitrary $B \in \mathcal{M}_n$, the matrix D_1, and hence the matrix BD_1, is a continuous function of the elements of B and that the strict inequalities used to define the permutation matrix P are in terms of the elements of BD_1. As $P(B_\bullet) = I_n$, the elements of $B_\bullet D_1(B_\bullet)$ satisfy the strict inequality constraints required in step (ii) of the Identification Scheme and, due to continuity, we can find a neighborhood $N(B_\bullet)$ of $B_\bullet$ in $\mathcal{M}_n^R$ such that the matrix $P(B)$ in the Identification Scheme equals I_n for all $B \in N(B_\bullet)$. With similar reasoning and redefining the neighborhood $N(B_\bullet)$ if necessary, we can conclude that step (iii) of the Identification Scheme is possible for all $B \in N(B_\bullet)$. To summarize, we have so far shown that there exists a neighborhood $N(B_\bullet)$ of $B_\bullet$ in $\mathcal{M}_n^R$ such that, for all $B \in N(B_\bullet)$, $B \in \mathcal{I}$ and that the matrix $P(B)$ in the Identification Scheme equals I_n.

To express the preceding conclusion differently, it holds that for all $B \in N(B_\bullet) \subseteq \mathcal{M}_n^R$, $\Pi(B) = BD_1D_2$ where, by the definition of D_2, also $\Pi(B) \in \mathcal{M}_n^R$. But if both B and $\Pi(B)$ belong to $\mathcal{M}_n^R$, it must be the case that $D_1 D_2 = I_n$. Therefore, for all $B \in N(B_\bullet)$, $\Pi(B) = B$. By the definition of $\mathcal{B}$ (and the fact that $B \in \mathcal{I}$), it thus holds that, for all $B \in N(B_\bullet)$, $B \in \mathcal{B}$. The proof is complete. ■

A3 Derivation of score and Hessian

Here we derive the expressions given for the score and Hessian in Appendices B and C, respectively. First, we give two differentiation rules that are frequently used below. The first rule (PR1) is obtained from result 17.30(h) in Seber (2008), whereas the second one (PR2) can be deduced from the first one.

Rule PR1. If X is $m \times n$, U is a $p \times q$ matrix function of X, and V is a $q \times r$ matrix function of X, then

$$\frac{\partial \operatorname{vec}(UV)}{\partial \operatorname{vec}(X)'} = (V \otimes I_p)' \frac{\partial \operatorname{vec}(U)}{\partial \operatorname{vec}(X)'} + (I_r \otimes U) \frac{\partial \operatorname{vec}(V)}{\partial \operatorname{vec}(X)'}$$

or, equivalently,

$$\frac{\partial \operatorname{vec}(UV)'}{\partial \operatorname{vec}(X)} = \frac{\partial \operatorname{vec}(U)'}{\partial \operatorname{vec}(X)}(V \otimes I_p) + \frac{\partial \operatorname{vec}(V)'}{\partial \operatorname{vec}(X)}(I_r \otimes U)'.$$

Rule PR2. If X is $m \times n$, U is a $p \times q$ matrix function of X, and V is a $q \times r$ matrix function of X, and W is a $r \times s$ matrix function of X, then

$$\frac{\partial \operatorname{vec}(UVW)}{\partial \operatorname{vec}(X)'} = (VW \otimes I_p)' \frac{\partial \operatorname{vec}(U)}{\partial \operatorname{vec}(X)'} + (W' \otimes U) \frac{\partial \operatorname{vec}(V)}{\partial \operatorname{vec}(X)'} + (I_s \otimes UV) \frac{\partial \operatorname{vec}(W)}{\partial \operatorname{vec}(X)'}$$

or, equivalently,

$$\frac{\partial \operatorname{vec}(UVW)'}{\partial \operatorname{vec}(X)} = \frac{\partial \operatorname{vec}(U)'}{\partial \operatorname{vec}(X)}(VW \otimes I_p) + \frac{\partial \operatorname{vec}(V)'}{\partial \operatorname{vec}(X)}(W \otimes U') + \frac{\partial \operatorname{vec}(W)'}{\partial \operatorname{vec}(X)}(I_s \otimes UV)'.$$

In what follows, several other results from Seber (2008) will also be used and referred to with abbreviations such as S 17.30(h). Similar abbreviations are used when references to results from Lütkepohl (1996) are made (e.g., L 9.2.2(5)(a)).

Derivation of the score.

$\underline{l_{\pi,t}(\theta)}$: First note that $\frac{\partial u_t(\pi)'}{\partial \pi} = -(x_{t-1} \otimes I_n)$ and

$$l_{\pi,t}(\theta) \overset{\text{PR1}}{=} \sum_{i=1}^{n} e_{i,x,t}(\theta) \frac{\partial u_t(\pi)'}{\partial \pi} (\sigma_i^{-1} \iota_i' B(\beta)^{-1})' = -(x_{t-1} \otimes I_n) B(\beta)^{-1'} \sum_{i=1}^{n} \sigma_i^{-1} e_{i,x,t}(\theta) \iota_i.$$

The expression of $l_{\pi,t}(\theta)$ in (14a) can be obtained by noting that $(x_{t-1} \otimes I_n) B^{-1'} = \left(x_{t-1} \otimes B^{-1'}\right)$ and $\sum_{i=1}^{n} \sigma_i^{-1} e_{i,x,t}(\theta) \iota_i = \Sigma^{-1} e_{x,t}(\theta)$, and using simple matrix algebra.

$\underline{l_{\beta,t}(\theta)}$: As $\operatorname{vec}(B(\beta)) = H\beta + \operatorname{vec}(I_n)$, we have $\frac{\partial \operatorname{vec}(B(\beta))'}{\partial \beta} = H'$ and hence

$$\begin{aligned}\frac{\partial \log \det(B(\beta))}{\partial \beta} &\overset{\text{S 17.7(b)(ii)}}{=} \frac{1}{\det(B(\beta))} \frac{\partial \det(B(\beta))}{\partial \beta} \\ &\overset{\text{S 17.26(c)}}{=} \frac{1}{\det(B(\beta))} H' \operatorname{vec}(\operatorname{adj}(B(\beta))') = H' \operatorname{vec}(B(\beta)^{-1'}).\end{aligned}$$

Thus,

$$\begin{aligned}l_{\beta,t}(\theta) &\overset{\text{PR2}}{=} \sum_{i=1}^{n} e_{i,x,t}(\theta) \frac{\partial \operatorname{vec}(B(\beta)^{-1})'}{\partial \beta}(u_t(\pi) \otimes \sigma_i^{-1} \iota_i) - H' \operatorname{vec}(B(\beta)^{-1'}) \\ &\overset{\text{S 17.33(b)}}{=} -H'(B(\beta)^{-1} \otimes B(\beta)^{-1'}) \sum_{i=1}^{n} (u_t(\pi) \otimes \iota_i) \sigma_i^{-1} e_{i,x,t}(\theta) - H' \operatorname{vec}(B(\beta)^{-1'}).\end{aligned}$$

The expression of $l_{\beta,t}(\theta)$ in (14b) can be obtained from this by noting that

$$\sum_{i=1}^{n}(u_t(\pi)\otimes\iota_i)\,\sigma_i^{-1}e_{i,x,t}(\theta)=u_t(\pi)\otimes\sum_{i=1}^{n}\sigma_i^{-1}e_{i,x,t}(\theta)\,\iota_i=\left(u_t(\pi)\otimes\Sigma^{-1}\right)e_{x,t}(\theta).$$

$\underline{l_{\sigma,t}(\theta)\text{ and }l_{\lambda,t}(\theta)}$: Differentiation directly gives

$$\begin{aligned}
l_{\sigma_i,t}(\theta) &= -\sigma_i^{-1}-\sigma_i^{-2}\iota_i' B(\beta)^{-1}u_t(\pi)\,e_{i,x,t}(\theta), \quad i=1,\ldots,n,\\
l_{\lambda_i,t}(\theta) &= e_{i,\lambda_i,t}(\theta), \quad i=1,\ldots,n,
\end{aligned}$$

which form the components of the expressions of $l_{\sigma,t}(\theta)$ and $l_{\lambda,t}(\theta)$ in (14c) and (14d).

Derivation of the Hessian. We begin with the diagonal blocks, and then derive the off-diagonal blocks.

$\underline{l_{\pi\pi,t}(\theta)}$: First note that

$$\frac{\partial e_{i,x,t}(\theta)}{\partial\pi'}=-e_{i,xx,t}(\theta)\,\sigma_i^{-1}\iota_i' B(\beta)^{-1}(x_{t-1}'\otimes I_n),$$

so that

$$\begin{aligned}
l_{\pi\pi,t}(\theta) &= -\frac{\partial}{\partial\pi'}\left[(x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1})e_{x,t}(\theta)\right]\\
&= -(x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1})\frac{\partial e_{x,t}(\theta)}{\partial\pi'}\\
&= (x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1})e_{xx,t}(\theta)\,\Sigma^{-1}B(\beta)^{-1}(x_{t-1}'\otimes I_n)\\
&= x_{t-1}x_{t-1}'\otimes B(\beta)^{-1'}\Sigma^{-1}e_{xx,t}(\theta)\,\Sigma^{-1}B(\beta)^{-1}\\
&= (I_n\otimes B(\beta)^{-1'}\Sigma^{-1})(x_{t-1}x_{t-1}'\otimes e_{xx,t}(\theta))(I_n\otimes\Sigma^{-1}B(\beta)^{-1}).
\end{aligned}$$

$\underline{l_{\beta\beta,t}(\theta)}$: We first note that $\frac{\partial\,\mathrm{vec}(B(\beta))}{\partial\beta'}=H$ and derive the following intermediate results:

$$\frac{\partial vec(B(\beta)^{-1})}{\partial\beta'}\overset{\text{S 17.33(b)}}{=}-(B(\beta)^{-1'}\otimes B(\beta)^{-1})\frac{\partial vec(B(\beta))}{\partial\beta'}=-(B(\beta)^{-1'}\otimes B(\beta)^{-1})H$$

$$\frac{\partial vec(B(\beta)^{-1'})}{\partial\beta'}=K_{nn}\frac{\partial vec(B(\beta)^{-1})}{\partial\beta'}=-K_{nn}(B(\beta)^{-1'}\otimes B(\beta)^{-1})H\overset{\text{L 9.2.2(5)(a)}}{=}-(B(\beta)^{-1}\otimes B(\beta)^{-1'})K_{nn}H$$

$$\begin{aligned}
\frac{\partial e_{i,x,t}(\theta)}{\partial\beta'} &= \frac{\partial}{\partial\beta'}\left[\frac{f_{i,x}(\sigma_i^{-1}\iota_i' B(\beta)^{-1}u_t(\pi);\lambda_i)}{f_i(\sigma_i^{-1}\iota_i' B(\beta)^{-1}u_t(\pi);\lambda_i)}\right]\overset{\text{PR2}}{=}e_{i,xx,t}(\theta)\left[(u_t'(\pi)\otimes\sigma_i^{-1}\iota_i')\frac{\partial vec(B(\beta)^{-1})}{\partial\beta'}\right]\\
&= -e_{i,xx,t}(\theta)\left[(u_t'(\pi)\otimes\sigma_i^{-1}\iota_i')(B(\beta)^{-1'}\otimes B(\beta)^{-1})H\right]\\
&= -(u_t'(\pi)\,B(\beta)^{-1'}\otimes e_{i,xx,t}(\theta)\,\sigma_i^{-1}\iota_i' B(\beta)^{-1})H
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial e_{x,t}(\theta)}{\partial \beta'} &= -(u_t'(\pi)\, B(\beta)^{-1'} \otimes e_{xx,t}(\theta)\, \Sigma^{-1} B(\beta)^{-1}) H \\
&= -(u_t'(\pi) \otimes e_{xx,t}(\theta))(B(\beta)^{-1'} \otimes \Sigma^{-1} B(\beta)^{-1}) H
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial}{\partial \beta'} \left[B(\beta)^{-1} u_t(\pi) \otimes B(\beta)^{-1'} \Sigma^{-1} e_{x,t}(\theta) \right] \\
&\overset{\text{S 17.34(a)(i)}}{=} B(\beta)^{-1} u_t(\pi) \otimes \frac{\partial B(\beta)^{-1'} \Sigma^{-1} e_{x,t}(\theta)}{\partial \beta'} + \frac{\partial B(\beta)^{-1} u_t(\pi)}{\partial \beta'} \otimes B(\beta)^{-1'} \Sigma^{-1} e_{x,t}(\theta) \\
&\overset{\text{PR1,PR2}}{=} B(\beta)^{-1} u_t(\pi) \otimes \left((\Sigma^{-1} e_{x,t}(\theta) \otimes I_n)' \frac{\partial vec(B(\beta)^{-1'})}{\partial \beta'} + B(\beta)^{-1'} \Sigma^{-1} \frac{\partial e_{x,t}(\theta)}{\partial \beta'} \right) \\
&+ \left((u_t(\pi) \otimes I_n)' \frac{\partial vec(B(\beta)^{-1})}{\partial \beta'} \right) \otimes B(\beta)^{-1'} \Sigma^{-1} e_{x,t}(\theta) \\
&= -B(\beta)^{-1} u_t(\pi) \otimes \left((\Sigma^{-1} e_{x,t}(\theta) \otimes I_n)' (B(\beta)^{-1} \otimes B(\beta)^{-1'}) K_{nn} H \right) \\
&- B(\beta)^{-1} u_t(\pi) \otimes \left(B(\beta)^{-1'} \Sigma^{-1} (u_t'(\pi) \otimes e_{xx,t}(\theta))(B(\beta)^{-1'} \otimes \Sigma^{-1} B(\beta)^{-1}) H \right) \\
&- \left((u_t(\pi) \otimes I_n)' (B(\beta)^{-1'} \otimes B(\beta)^{-1}) H \right) \otimes B(\beta)^{-1'} \Sigma^{-1} e_{x,t}(\theta) \\
&= -\left(B(\beta)^{-1} u_t(\pi) \otimes e_{x,t}'(\theta)\, \Sigma^{-1} B(\beta)^{-1} \otimes B(\beta)^{-1'} \right) K_{nn} H \\
&- B(\beta)^{-1} u_t(\pi) \otimes (u_t'(\pi)\, B(\beta)^{-1'} \otimes B(\beta)^{-1'} \Sigma^{-1} e_{xx,t}(\theta)\, \Sigma^{-1} B(\beta)^{-1}) H \\
&- \left(u_t'(\pi)\, B(\beta)^{-1'} \otimes B(\beta)^{-1} \right) H \otimes B(\beta)^{-1'} \Sigma^{-1} e_{x,t}(\theta) \\
&\overset{\text{L 2.4(12), L 9.2.2(3)ii}}{=} -\left(B(\beta)^{-1} u_t(\pi) \otimes e_{x,t}'(\theta)\, \Sigma^{-1} B(\beta)^{-1} \otimes B(\beta)^{-1'} \right) K_{nn} H \\
&- (B(\beta)^{-1} \otimes B(\beta)^{-1'} \Sigma^{-1})(u_t(\pi)\, u_t'(\pi) \otimes e_{xx,t}(\theta))(B(\beta)^{-1'} \otimes \Sigma^{-1} B(\beta)^{-1}) H \\
&- K_{nn} \left(B(\beta)^{-1'} \Sigma^{-1} e_{x,t}(\theta) \otimes \left(u_t'(\pi)\, B(\beta)^{-1'} \otimes B(\beta)^{-1} \right) \right) H,
\end{aligned}
$$

where in the last two equalities we also used the rule $a \otimes BC = a1 \otimes BC = (a \otimes B)C$ (a column vector, B and C conformable).

Hence, $l_{\beta\beta,t}(\theta)$ becomes

$$\begin{aligned}
l_{\beta\beta,t}(\theta) &= -H'\left[\frac{\partial}{\partial\beta'}\left[(B(\beta)^{-1}u_t(\pi)\otimes B(\beta)^{-1'}\Sigma^{-1})e_{x,t}(\theta)\right]+\frac{\partial}{\partial\beta'}vec(B(\beta)^{-1'})\right] \\
&= H'\left(B(\beta)^{-1}u_t(\pi)\otimes e'_{x,t}(\theta)\Sigma^{-1}B(\beta)^{-1}\otimes B(\beta)^{-1'}\right)K_{nn}H \\
&\quad +H'(B(\beta)^{-1}\otimes B(\beta)^{-1'}\Sigma^{-1})(u_t(\pi)u'_t(\pi)\otimes e_{xx,t}(\theta))(B(\beta)^{-1'}\otimes\Sigma^{-1}B(\beta)^{-1})H \\
&\quad +H'K_{nn}\left(B(\beta)^{-1'}\Sigma^{-1}e_{x,t}(\theta)\otimes u'_t(\pi)B(\beta)^{-1'}\otimes B(\beta)^{-1}\right)H \\
&\quad +H'(B(\beta)^{-1}\otimes B(\beta)^{-1'})K_{nn}H \\
&\overset{\text{L 2.4(12)}}{=} H'\left(B(\beta)^{-1}u_t(\pi)e'_{x,t}(\theta)\otimes I_n\right)\left(\Sigma^{-1}B(\beta)^{-1}\otimes B(\beta)^{-1'}\right)K_{nn}H \\
&\quad +H'(B(\beta)^{-1}\otimes B(\beta)^{-1'}\Sigma^{-1})(u_t(\pi)u'_t(\pi)\otimes e_{xx,t}(\theta))(B(\beta)^{-1'}\otimes\Sigma^{-1}B(\beta)^{-1})H \\
&\quad +H'K_{nn}\left(B(\beta)^{-1'}\Sigma^{-1}\otimes B(\beta)^{-1}\right)\left(e_{x,t}(\theta)u'_t(\pi)B(\beta)^{-1'}\otimes I_n\right)H \\
&\quad +H'(B(\beta)^{-1}\otimes B(\beta)^{-1'})K_{nn}H \\
&= H'(B(\beta)^{-1}\otimes B(\beta)^{-1'}\Sigma^{-1})(u_t(\pi)u'_t(\pi)\otimes e_{xx,t}(\theta))(B(\beta)^{-1'}\otimes\Sigma^{-1}B(\beta)^{-1})H \\
&\quad +H'(B(\beta)^{-1}\otimes I_n)\left(u_t(\pi)e'_{x,t}(\theta)\otimes I_n\right)\left(\Sigma^{-1}B(\beta)^{-1}\otimes B(\beta)^{-1'}\right)K_{nn}H \\
&\quad H'K_{nn}(B(\beta)^{-1'}\Sigma^{-1}\otimes B(\beta)^{-1})\left(e_{x,t}(\theta)u'_t(\pi)\otimes I_n\right)(B(\beta)^{-1'}\otimes I_n)H \\
&\quad +H'(B(\beta)^{-1}\otimes B(\beta)^{-1'})K_{nn}H.
\end{aligned}$$

$\underline{l_{\sigma\sigma,t}(\theta)}$: Recall that

$$\begin{aligned}
l_{\sigma_i,t}(\theta) &= -\sigma_i^{-1}-\sigma_i^{-2}\iota'_iB(\beta)^{-1}u_t(\pi)e_{i,x,t}(\theta), \quad i=1,...,n, \\
l_{\sigma,t}(\theta) &= -\Sigma^{-2}\left[\varepsilon_t(\theta)\odot e_{x,t}(\theta)+\sigma\right].
\end{aligned}$$

Thus $l_{\sigma\sigma,t}(\theta)$ is a diagonal matrix with diagonal elements

$$\begin{aligned}
l_{\sigma_i\sigma_i,t}(\theta) &= \frac{\partial}{\partial\sigma_i}\left[-\sigma_i^{-1}-\sigma_i^{-2}\iota'_iB(\beta)^{-1}u_t(\pi)e_{i,x,t}(\theta)\right] \\
&= \sigma_i^{-2}+2\sigma_i^{-3}\iota'_iB(\beta)^{-1}u_t(\pi)e_{i,x,t}(\theta)-\sigma_i^{-2}\iota'_iB(\beta)^{-1}u_t(\pi)e_{i,xx,t}(\theta)\left[-\sigma_i^{-2}\iota'_iB(\beta)^{-1}u_t(\pi)\right] \\
&= \sigma_i^{-2}+2\sigma_i^{-3}\varepsilon_{i,t}(\theta)e_{i,x,t}(\theta)+\sigma_i^{-4}\varepsilon^2_{i,t}(\theta)e_{i,xx,t}(\theta),
\end{aligned}$$

so that $l_{\sigma\sigma,t}(\theta)$ can be expressed as

$$l_{\sigma\sigma,t}(\theta)=\Sigma^{-2}+2\Sigma^{-3}\mathcal{E}_t(\theta)E_{x,t}(\theta)+\Sigma^{-4}\mathcal{E}^2_t(\theta)e_{xx,t}(\theta).$$

$\underline{l_{\lambda\lambda,t}(\theta)}$: Directly,

$$l_{\lambda\lambda,t}(\theta)=e_{\lambda\lambda,t}(\theta).$$

$\underline{l_{\pi\beta,t}(\theta)}$: As a preliminary result we first obtain (see the derivation of $l_{\beta\beta,t}(\theta)$),

$$\begin{aligned}
&\frac{\partial}{\partial\beta'}\left[B(\beta)^{-1'}\Sigma^{-1}e_{x,t}(\theta)\right] \\
&\overset{\text{PR2}}{=} (\Sigma^{-1}e_{x,t}(\theta)\otimes I_n)'\frac{\partial vec(B(\beta)^{-1'})}{\partial\beta'} + B(\beta)^{-1'}\Sigma^{-1}\frac{\partial e_{x,t}(\theta)}{\partial\beta'} \\
= \;& -(\Sigma^{-1}e_{x,t}(\theta)\otimes I_n)'(B(\beta)^{-1}\otimes B(\beta)^{-1'})K_{nn}H \\
& -B(\beta)^{-1'}\Sigma^{-1}(u_t'(\pi)\otimes e_{xx,t}(\theta))(B(\beta)^{-1'}\otimes\Sigma^{-1}B(\beta)^{-1})H \\
= \;& -(e_{x,t}'(\theta)\Sigma^{-1}B(\beta)^{-1}\otimes B(\beta)^{-1'})K_{nn}H - B(\beta)^{-1'}\Sigma^{-1}(u_t'(\pi)B(\beta)^{-1'}\otimes e_{xx,t}(\theta)\Sigma^{-1}B(\beta)^{-1})H.
\end{aligned}$$

Hence,

$$\begin{aligned}
l_{\pi\beta,t}(\theta) &= \frac{\partial}{\partial\beta'}\left[-(x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1})e_{x,t}(\theta)\right] \\
&= -\frac{\partial}{\partial\beta'}\left[x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1}e_{x,t}(\theta)\right] \\
&\overset{\text{S 17.34(a)(i)}}{=} -x_{t-1}\otimes\frac{\partial}{\partial\beta'}\left[B(\beta)^{-1'}\Sigma^{-1}e_{x,t}(\theta)\right] \\
&= x_{t-1}\otimes\left[(e_{x,t}'(\theta)\Sigma^{-1}B(\beta)^{-1}\otimes B(\beta)^{-1'})K_{nn}H\right] \\
&\quad +x_{t-1}\otimes\left[B(\beta)^{-1'}\Sigma^{-1}(u_t'(\pi)B(\beta)^{-1'}\otimes e_{xx,t}(\theta)\Sigma^{-1}B(\beta)^{-1})H\right] \\
&= x_{t-1}\otimes\left[(I_n\otimes e_{x,t}'(\theta))(B(\beta)^{-1'}\otimes\Sigma^{-1}B(\beta)^{-1})H\right] \\
&\quad +x_{t-1}\otimes\left[B(\beta)^{-1'}\Sigma^{-1}(u_t'(\pi)\otimes e_{xx,t}(\theta))(B(\beta)^{-1'}\otimes\Sigma^{-1}B(\beta)^{-1})H\right].
\end{aligned}$$

$\underline{l_{\pi\sigma,t}(\theta)}$: Note that $\Sigma^{-1}e_{x,t}(\theta)$ is a vector with the ith component given by $\sigma_i^{-1}e_{i,x,t}(\theta)$, so that

$$\begin{aligned}
\frac{\partial}{\partial\sigma_i}\left[\sigma_i^{-1}e_{i,x,t}(\theta)\right] &= -\sigma_i^{-2}e_{i,x,t}(\theta) - \sigma_i^{-1}e_{i,xx,t}(\theta)\sigma_i^{-2}\iota_i'B(\beta)^{-1}u_t(\pi) \\
&= -\sigma_i^{-2}e_{i,x,t}(\theta) - \sigma_i^{-1}e_{i,xx,t}(\theta)\sigma_i^{-2}\varepsilon_{i,t}(\theta)
\end{aligned}$$

and $\frac{\partial}{\partial\sigma_j}\left[\sigma_i^{-1}e_{i,x,t}(\theta)\right] = 0$ for $j \neq i$. Therefore,

$$\frac{\partial}{\partial\sigma'}\left[\Sigma^{-1}e_{x,t}(\theta)\right] = -\Sigma^{-2}E_{x,t}(\theta) - \Sigma^{-3}e_{xx,t}(\theta)\mathcal{E}_t(\theta)$$

and

$$\begin{aligned}
l_{\pi\sigma,t}(\theta) &= \frac{\partial}{\partial\sigma'}\left[-(x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1})e_{x,t}(\theta)\right] \\
&\overset{\text{S 17.34(a)(i)}}{=} -x_{t-1}\otimes B(\beta)^{-1'}\frac{\partial}{\partial\sigma'}\left[\Sigma^{-1}e_{x,t}(\theta)\right] \\
&= x_{t-1}\otimes B(\beta)^{-1'}\left[\Sigma^{-2}E_{x,t}(\theta) + \Sigma^{-3}e_{xx,t}(\theta)\mathcal{E}_t(\theta)\right].
\end{aligned}$$

$\underline{l_{\pi\lambda,t}(\theta)}$: With direct computation,

$$\begin{aligned} l_{\pi\lambda,t}(\theta) &= \frac{\partial}{\partial \lambda'}\left[-(x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1})e_{x,t}(\theta)\right] \\ &= -x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1}\frac{\partial e_{x,t}(\theta)}{\partial\lambda'} \\ &= -x_{t-1}\otimes B(\beta)^{-1'}\Sigma^{-1}e_{x\lambda,t}(\theta) \\ &= -(I_{np+1}\otimes B(\beta)^{-1'}\Sigma^{-1})(x_{t-1}\otimes e_{x\lambda,t}(\theta)). \end{aligned}$$

$\underline{l_{\beta\sigma,t}(\theta)}$: With direct computation,

$$\begin{aligned} l_{\beta\sigma,t}(\theta) &= \frac{\partial}{\partial\sigma'}\left[-H'\left[(B(\beta)^{-1}u_t(\pi)\otimes B(\beta)^{-1'}\Sigma^{-1})e_{x,t}(\theta)+vec(B(\beta)^{-1'})\right]\right] \\ &= -H'(B(\beta)^{-1}u_t(\pi)\otimes B(\beta)^{-1'}\frac{\partial}{\partial\sigma'}\left[\Sigma^{-1}e_{x,t}(\theta)\right]) \\ &= H'(B(\beta)^{-1}u_t(\pi)\otimes B(\beta)^{-1'}\left[\Sigma^{-2}E_{x,t}(\theta)+\Sigma^{-3}e_{xx,t}(\theta)\mathcal{E}_t(\theta)\right]) \\ &= H'(B(\beta)^{-1}\otimes B(\beta)^{-1'})(u_t(\pi)\otimes\left[\Sigma^{-2}E_{x,t}(\theta)+\Sigma^{-3}e_{xx,t}(\theta)\mathcal{E}_t(\theta)\right]). \end{aligned}$$

$\underline{l_{\beta\lambda,t}(\theta)}$: With direct computation,

$$\begin{aligned} l_{\beta\lambda,t}(\theta) &= \frac{\partial}{\partial\lambda'}\left[-H'\left[(B(\beta)^{-1}u_t(\pi)\otimes B(\beta)^{-1'}\Sigma^{-1})e_{x,t}(\theta)+vec(B(\beta)^{-1'})\right]\right] \\ &= -H'(B(\beta)^{-1}u_t(\pi)\otimes B(\beta)^{-1'}\Sigma^{-1}\frac{\partial e_{x,t}(\theta)}{\partial\lambda'}) \\ &= -H'(B(\beta)^{-1}u_t(\pi)\otimes B(\beta)^{-1'}\Sigma^{-1}e_{x\lambda,t}(\theta)) \\ &= -H'(B(\beta)^{-1}\otimes B(\beta)^{-1'}\Sigma^{-1})(u_t(\pi)\otimes e_{x\lambda,t}(\theta)). \end{aligned}$$

$\underline{l_{\sigma\lambda,t}(\theta)}$: As $l_{\sigma_i,t}(\theta)=-\sigma_i^{-1}-\sigma_i^{-2}\varepsilon_{i,t}(\theta)e_{i,x,t}(\theta)$, we have

$$\begin{aligned} l_{\sigma_i\lambda_i,t}(\theta) &= \frac{\partial}{\partial\lambda_i'}l_{\sigma_i,t}(\theta)=-\sigma_i^{-2}\varepsilon_{i,t}(\theta)e_{i,x\lambda_i,t}(\theta) \\ l_{\sigma_i\lambda_j,t}(\theta) &= 0 \quad \text{for } j\neq i, \end{aligned}$$

so that

$$l_{\sigma\lambda,t}(\theta)=-\Sigma^{-2}\mathcal{E}_t(\theta)e_{x\lambda,t}(\theta).$$

A4 Further details omitted from Appendix B

Proof of Lemma B.1. Part (i) holds because, due to integrability of $f_{i,x}(x;\lambda_{i,0})$ (Assumption 4(ii)) and the fundamental theorem of calculus,

$$E[e_{i,x,t}]=\int f_{i,x}(\sigma_{i,0}^{-1}x;\lambda_{i,0})\sigma_{i,0}^{-1}dx=\int f_{i,x}(x;\lambda_{i,0})dx=|_{-\infty}^{\infty}f_i(x;\lambda_{i,0})=0,$$

where the last equality holds due to $\int f_i(x;\lambda_{i,0})dx = 1$. Parts (ii), (iv), and (vi) are immediate consequences of the dominance conditions in Assumption 4(iii). Part (iii) holds because

$$\begin{aligned} E\left[e_{i,\lambda_i,t}\right] &= \int f_{i,\lambda_i}(\sigma_{i,0}^{-1}x;\lambda_{i,0})\sigma_{i,0}^{-1}dx = \int f_{i,\lambda_i}(x;\lambda_{i,0})dx \\ &= \int \left[\frac{\partial f_i(x;\lambda_i)}{\partial \lambda_i}\right]_{\lambda_i=\lambda_{i,0}} dx = \left[\frac{d}{d\lambda_i}\int f_i(x;\lambda_i)dx\right]_{\lambda_i=\lambda_{i,0}} = 0, \end{aligned}$$

where the penultimate equality is justified by Assumption 4(iv) and a theorem on the differentiation of an integral (see, e.g., Theorem 24.5 and the discussion following it in Aliprantis and Burkinshaw (1998)). Part (v) follows because, in view of the integrability of $f_{i,x}(x;\lambda_{i,0})$,

$$\begin{aligned} E\left[\varepsilon_{i,t}e_{i,x,t}\right] &= \int x f_{i,x}(\sigma_{i,0}^{-1}x;\lambda_{i,0})\sigma_{i,0}^{-1}dx = \sigma_{i,0}\int x f_{i,x}(x;\lambda_{i,0})dx \\ &= \sigma_{i,0}|_{-\infty}^{\infty} x f_i(x;\lambda_{i,0}) - \sigma_{i,0}\int f_i(x;\lambda_{i,0})dx = -\sigma_{i,0}, \end{aligned}$$

where the last equality holds due to the facts $E[\varepsilon_{i,t}] = 0$ and $\int f_i(x;\lambda_{i,0})dx = 1$. ■

A5 Further details omitted from Appendix C

First define, for $i = 1,\ldots,n$,

$$\begin{aligned} e_{i,xx,t} &= e_{i,xx,t}(\theta_0) = \frac{f_{i,xx}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})} - \left(\frac{f_{i,x}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}\right)^2 \\ e_{i,x\lambda_i,t} &= e_{i,x\lambda_i,t}(\theta_0) = \frac{f_{i,x\lambda_i}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})} - \frac{f_{i,x}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}\frac{f'_{i,\lambda_i}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})} \\ e_{i,\lambda_i\lambda_i,t} &= e_{i,\lambda_i\lambda_i,t}(\theta_0) = \frac{f_{i,\lambda_i\lambda_i}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})} - \frac{f_{i,\lambda_i}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}\frac{f'_{i,\lambda_i}(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})}{f_i(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0})} \end{aligned}$$

and use these to form the diagonal / block diagonal matrices

$$\begin{aligned} e_{xx,t} &= e_{xx,t}(\theta_0) = diag\left(e_{1,xx,t},\ldots,e_{n,xx,t}\right) \quad (n\times n) \\ e_{\lambda\lambda,t} &= e_{\lambda\lambda,t}(\theta_0) = diag\left(e_{1,\lambda_1\lambda_1,t},\ldots,e_{n,\lambda_n\lambda_n,t}\right) \quad (d\times d) \\ e_{x\lambda,t} &= e_{x\lambda,t}(\theta_0) = diag(e_{1,x\lambda_1,t},\ldots,e_{n,x\lambda_n,t}) \quad (n\times d) \end{aligned}$$

The following auxiliary lemmas contains results needed in subsequent derivations.

Lemma C.1. Under Assumptions 2–5, the following hold (for $i = 1,\ldots,n$ when apropriate): (i) $E[e_{xx,t}] = -E[e_{x,t}e'_{x,t}]$, (ii) $\int x^2 f_{i,xx}\left(x;\lambda_{i,0}\right)dx = 2$, (iii) $E[e_{\lambda\lambda,t}] = -E[e_{\lambda,t}e'_{\lambda,t}]$, (iv)

$E[e_{x\lambda,t}] = -E[e_{x,t}e'_{\lambda,t}]$, (v) $E[\varepsilon_{i,t}e_{x\lambda_i,t}] = -E[\varepsilon_{i,t}e_{i,x,t}e'_{i,\lambda_i,t}]$, (vi) $E[\varepsilon_{i,t}e_{i,xx,t}] = -E\left[\varepsilon_{i,t}e^2_{i,x,t}\right]$, (vii) $E[\varepsilon^2_{i,t}e_{i,xx,t}] = 2\sigma^2_{i,0} - E\left[\varepsilon^2_{i,t}e^2_{i,x,t}\right]$.

Proof of Lemma C.1. (i) We need to show that $E\left[e_{i,xx,t}\right] = -E[e^2_{i,x,t}]$ for $i = 1, \ldots, n$, which, recalling the definition of $e_{i,xx,t}$, holds if $\int f_{i,xx}(x; \lambda_{i,0})\, dx = 0$. As $f_{i,xx}(x; \lambda_{i,0})$ is integrable by Assumption 5(i), fundamental theorem of calculus yields $\int f_{i,xx}(x; \lambda_{i,0})\, dx = |^{\infty}_{-\infty} f_{i,x}(x; \lambda_{i,0})$, and the expression on the right equals zero because $E\left[e_{i,x,t}\right] = 0$ (Lemma B.1(i)). (ii) Integration by parts yields

$$\int x^2 f_{i,xx}(x; \lambda_{i,0})\, dx = |^{\infty}_{-\infty} x^2 f_{i,x}(x; \lambda_{i,0}) - 2\int x f_{i,x}(x; \lambda_{i,0})\, dx = 2,$$

where the last equality holds because $E\left[\varepsilon^2_{i,t}\left|e_{i,x,t}\right|\right]$ is finite (due to Lemma B.1(vi)) and because the integral in the penultimate expression equals -1 (see the proof of Lemma B.1(v)). (iii) By the definition of $e_{\lambda\lambda,t}$ it suffices to show that $E\left[f_{i,\lambda_i\lambda_i}\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)/f_i\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)\right] = 0$ for $i = 1, \ldots, n$. This holds because

$$\begin{aligned} E\left[\frac{f_{i,\lambda_i\lambda_i}\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)}{f_i\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)}\right] &= \int f_{i,\lambda_i\lambda_i}\left(\sigma^{-1}_{i,0}x; \lambda_{i,0}\right)\sigma^{-1}_{i,0}dx = \int f_{i,\lambda_i\lambda_i}(x; \lambda_{i,0})\, dx \\ &= \int \left[\frac{\partial^2 f_i(x; \lambda_i)}{\partial\lambda_i\partial\lambda'_i}\right]_{\lambda_i=\lambda_{i,0}} dx = \left[\frac{d^2}{d\lambda_i d\lambda'_i}\int f_i(x; \lambda_i)dx\right]_{\lambda_i=\lambda_{i,0}} = 0, \end{aligned}$$

where the penultimate equality is justified by Assumption 5(ii), cf. proof of Lemma B.1(iii). (iv) By the definition of $e_{x\lambda,t}$ it suffices to show that $E\left[f_{i,x\lambda_i}\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)/f_i\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)\right] = 0$ for $i = 1, \ldots, n$. This holds because, in view of the integrability of $f_{i,x\lambda_i}(x; \lambda_{i,0})$ (Assumption 5(i)),

$$E\left[\frac{f_{i,x\lambda_i}\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)}{f_i\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)}\right] = \int f_{i,x\lambda_i}\left(\sigma^{-1}_{i,0}x; \lambda_{i,0}\right)\sigma^{-1}_{i,0}dx = \int f_{i,x\lambda_i}(x; \lambda_{i,0})\, dx = |^{\infty}_{-\infty} f_{i,\lambda_i}(x; \lambda_{i,0}) = 0,$$

where the last equality holds because $E\left[e_{i,\lambda_i,t}\right] = 0$ by Lemma B.1(iii). (v) By the definition of $e_{x\lambda,t}$ it suffices to show that $E\left[\varepsilon_{i,t}f_{i,x\lambda_i}\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)/f_i\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)\right] = \int x f_{i,x\lambda_i}(x; \lambda_{i,0})\, dx = 0$ for all $i = 1, \ldots, n$. This holds because, by change of variables and integration by parts,

$$\begin{aligned} E\left[\varepsilon_{i,t}\frac{f_{i,x\lambda_i}\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)}{f_i\left(\sigma^{-1}_{i,0}\varepsilon_{i,t}; \lambda_{i,0}\right)}\right] &= \int x f_{i,x\lambda_i}\left(\sigma^{-1}_{i,0}x; \lambda_{i,0}\right)\sigma^{-1}_{i,0}dx = \sigma_{i,0}\int x f_{i,x\lambda_i}(x; \lambda_{i,0})\, dx \\ &= \sigma_{i,0}|^{\infty}_{-\infty} x f_{i,\lambda_i}(x; \lambda_{i,0}) - \sigma_{i,0}\int f_{i,\lambda_i}(x; \lambda_{i,0})\, dx. \end{aligned}$$

In the last expression derived, the integral on the right is zero because $E\left[e_{i,\lambda_i,t}\right] = 0$ (due to Lemma B.1(iii)), whereas the first term therein is zero because $E\left[\left|\varepsilon_{i,t}e_{i,\lambda,t}\right|\right] < \infty$ (due to

Assumption 1 and Lemma B.1(iv)). (vi) By the definition of $e_{i,xx,t}$ it suffices to establish that $E[\varepsilon_{i,t} f_{i,xx}\left(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0}\right)/f_i\left(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0}\right)] = 0$. To this end, by change of variables and integration by parts,

$$\begin{aligned} E\left[\varepsilon_{i,t}\frac{f_{i,xx}\left(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0}\right)}{f_i\left(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0}\right)}\right] &= \int \sigma_{i,0}^{-1}x f_{i,xx}\left(\sigma_{i,0}^{-1}x;\lambda_{i,0}\right)dx = \sigma_{i,0}\int \sigma_{i,0}^{-1}x f_{i,xx}\left(\sigma_{i,0}^{-1}x;\lambda_{i,0}\right)\sigma_{i,0}^{-1}dx \\ &= \sigma_{i,0}\int x f_{i,xx}\left(x;\lambda_{i,0}\right)dx = \sigma_{i,0}|_{-\infty}^{\infty} x f_{i,x}\left(x;\lambda_{i,0}\right) - \sigma_{i,0}\int f_{i,x}\left(x;\lambda_{i,0}\right)dx. \end{aligned}$$

The integral in the last expression is zero because $E\left[e_{i,x,t}\right] = 0$ (Lemma B.1(i)), whereas the first term therein is zero because $E\left[\varepsilon_{i,t}e_{i,x,t}\right] = -\sigma_{i,0}$ (Lemma B.1(v)). (vii) By the definition of $e_{i,xx,t}$

$$E\left[\varepsilon_{i,t}^2 e_{i,xx,t}\right] = E\left[\varepsilon_{i,t}^2\frac{f_{i,xx}\left(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0}\right)}{f_i\left(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0}\right)}\right] - E\left[\varepsilon_{i,t}^2 e_{i,x,t}^2\right] = 2\sigma_{i,0}^2 - E\left[\varepsilon_{i,t}^2 e_{i,x,t}^2\right].$$

Here the second equality follows because

$$\begin{aligned} E\left[\varepsilon_{i,t}^2\frac{f_{i,xx}\left(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0}\right)}{f_i\left(\sigma_{i,0}^{-1}\varepsilon_{i,t};\lambda_{i,0}\right)}\right] &= \int \sigma_{i,0}^{-1}x^2 f_{i,xx}\left(\sigma_{i,0}^{-1}x;\lambda_{i,0}\right)dx = \sigma_{i,0}^2\int \sigma_{i,0}^{-2}x^2 f_{i,xx}\left(\sigma_{i,0}^{-1}x;\lambda_{i,0}\right)\sigma_{i,0}^{-1}dx \\ &= \sigma_{i,0}^2\int x^2 f_{i,xx}\left(x;\lambda_{i,0}\right)dx = 2\sigma_{i,0}^2, \end{aligned}$$

where the last equality follows from the proof of part (ii). ■

The following lemma is used below when proving the result $E\left[l_{\theta\theta,t}\left(\theta_0\right)\right] = -E[l_{\theta,t}\left(\theta_0\right)l_{\theta,t}'\left(\theta_0\right)]$. We introduce the notation

$$\mathbb{I}_n = \begin{bmatrix} \iota_1\iota_1' \\ \vdots \\ \iota_n\iota_n' \end{bmatrix} \qquad (n^2\times n).$$

Lemma C.2. Under Assumptions 2–5, the following hold:

(i) $E[\varepsilon_t'\otimes e_{xx,t}] = -E[\varepsilon_t'\otimes e_{x,t}e_{x,t}']$

(ii) $E[\varepsilon_t\otimes E_{x,t}] = -(\Sigma_0\otimes I_n)\mathbb{I}_n$

(iii) $E[\varepsilon_t\otimes e_{xx,t}\mathcal{E}_t] = 2(\Sigma_0^2\otimes I_n)\mathbb{I}_n - (E[\mathcal{E}_t^2 E_{x,t}^2]\otimes I_n)\mathbb{I}_n$

(iv) $E[\varepsilon_t\varepsilon_t'\otimes e_{xx,t}] = 2(\Sigma_0^2\otimes I_n)\mathbb{I}_n - E[\mathcal{E}_t^2\otimes E_{x,t}^2]$

(v) $E\left[\varepsilon_t\varepsilon_t'\otimes e_{x,t}e_{x,t}'\right] = E[\mathcal{E}_t^2\otimes E_{x,t}^2] - 2(\Sigma_0^2\otimes I_n)\mathbb{I}_n + (I_n\otimes\Sigma_0)\left[(vec(I_n)vec(I_n)') + K_{nn}\right](I_n\otimes\Sigma_0)$

Proof of Lemma C.2. (i) Using Lemma C.1(vi),

$$\begin{aligned} E[\varepsilon_t' \otimes e_{xx,t}] &= E\begin{bmatrix}\varepsilon_{1,t}e_{xx,t} & \cdots & \varepsilon_{n,t}e_{xx,t}\end{bmatrix} \\ &= \begin{bmatrix}E[\varepsilon_{1,t}e_{1,xx,t}]\iota_1\iota_1' & \cdots & E[\varepsilon_{n,t}e_{n,xx,t}]\iota_n\iota_n'\end{bmatrix} \\ &= -\begin{bmatrix}E[\varepsilon_{1,t}e_{1,x,t}^2]\iota_1\iota_1' & \cdots & E[\varepsilon_{n,t}e_{n,x,t}^2]\iota_n\iota_n'\end{bmatrix} \\ &= -E[\varepsilon_t' \otimes e_{x,t}e_{x,t}']. \end{aligned}$$

(ii) Using Lemma B.1(v),

$$E[\varepsilon_t \otimes E_{x,t}] = \begin{bmatrix}E[\varepsilon_{1,t}E_{x,t}] \\ \vdots \\ E[\varepsilon_{n,t}E_{x,t}]\end{bmatrix} = \begin{bmatrix}-\sigma_{1,0}\iota_1\iota_1' \\ \vdots \\ -\sigma_{n,0}\iota_n\iota_n'\end{bmatrix} = -(\Sigma_0 \otimes I_n)\mathbb{I}_n.$$

(iii) Using Lemma C.1(vii),

$$\begin{aligned} E[\varepsilon_t \otimes e_{xx,t}\mathcal{E}_t] &= \begin{bmatrix}E[\varepsilon_{1,t}e_{xx,t}\mathcal{E}_t] \\ \vdots \\ E[\varepsilon_{n,t}e_{xx,t}\mathcal{E}_t]\end{bmatrix} = \begin{bmatrix}E[\varepsilon_{1,t}^2e_{1,xx,t}]\iota_1\iota_1' \\ \vdots \\ E[\varepsilon_{n,t}^2e_{n,xx,t}]\iota_n\iota_n'\end{bmatrix} \\ &= \begin{bmatrix}(2\sigma_{1,0}^2 - E[\varepsilon_{1,t}^2e_{1,x,t}^2])\iota_1\iota_1' \\ \vdots \\ (2\sigma_{n,0}^2 - E[\varepsilon_{n,t}^2e_{n,x,t}^2])\iota_n\iota_n'\end{bmatrix} = 2(\Sigma_0^2 \otimes I_n)\mathbb{I}_n - (E[\mathcal{E}_t^2E_{x,t}^2] \otimes I_n)\mathbb{I}_n. \end{aligned}$$

(iv) First note that $E[\varepsilon_t\varepsilon_t' \otimes e_{xx,t}]$ can be written as

$$E[\varepsilon_t\varepsilon_t' \otimes e_{xx,t}] = E\begin{bmatrix}\varepsilon_{1,t}\varepsilon_{1,t}e_{xx,t} & \cdots & \varepsilon_{1,t}\varepsilon_{n,t}e_{xx,t} \\ \vdots & \ddots & \vdots \\ \varepsilon_{n,t}\varepsilon_{1,t}e_{xx,t} & \cdots & \varepsilon_{n,t}\varepsilon_{n,t}e_{xx,t}\end{bmatrix} = \begin{bmatrix}E[\varepsilon_{1,t}\varepsilon_{1,t}e_{xx,t}] & & \\ & \ddots & \\ & & E[\varepsilon_{n,t}\varepsilon_{n,t}e_{xx,t}]\end{bmatrix},$$

as the non-diagonal blocks are zeros. Now note that the first diagonal block can be expressed as (using Lemma C.1(vii))

$$\begin{aligned} E[\varepsilon_{1,t}\varepsilon_{1,t}e_{xx,t}] &= \begin{bmatrix}E[\varepsilon_{1,t}\varepsilon_{1,t}e_{1,xx,t}] & & \\ & \ddots & \\ & & E[\varepsilon_{1,t}\varepsilon_{1,t}e_{n,xx,t}]\end{bmatrix} \\ &= \begin{bmatrix}2\sigma_{1,0}^2 - E[\varepsilon_{1,t}^2e_{1,x,t}^2] & & & \\ & -\sigma_{1,0}^2E[e_{2,x,t}^2] & & \\ & & \ddots & \\ & & & -\sigma_{1,0}^2E[e_{n,x,t}^2]\end{bmatrix} \\ &= 2\sigma_{1,0}^2\iota_1\iota_1' - E[\varepsilon_{1,t}^2E_{x,t}^2], \end{aligned}$$

and so on, and the last diagonal block can be expressed as

$$\begin{aligned} E[\varepsilon_{n,t}\varepsilon_{n,t}e_{xx,t}] &= \begin{bmatrix} E[\varepsilon_{n,t}\varepsilon_{n,t}e_{1,xx,t}] & & \\ & \ddots & \\ & & E[\varepsilon_{n,t}\varepsilon_{n,t}e_{n,xx,t}] \end{bmatrix} \\ &= \begin{bmatrix} -\sigma_{n,0}^2 E[e_{1,x,t}^2] & & & \\ & \ddots & & \\ & & -\sigma_{n,0}^2 E[e_{n-1,x,t}^2] & \\ & & & 2\sigma_{n,0}^2 - E[\varepsilon_{n,t}^2 e_{n,x,t}^2] \end{bmatrix} \\ &= 2\sigma_{n,0}^2 \iota_n \iota_n' - E[\varepsilon_{n,t}^2 E_{x,t}^2]. \end{aligned}$$

Therefore,

$$\begin{aligned} E[\varepsilon_t\varepsilon_t' \otimes e_{xx,t}] &= \begin{bmatrix} 2\sigma_{1,0}^2\iota_1\iota_1' & & \\ & \ddots & \\ & & 2\sigma_{n,0}^2\iota_n\iota_n' \end{bmatrix} - \begin{bmatrix} E[\varepsilon_{1,t}^2 E_{x,t}^2] & & \\ & \ddots & \\ & & E[\varepsilon_{n,t}^2 E_{x,t}^2] \end{bmatrix} \\ &= 2(\Sigma_0^2 \otimes I_n)\mathbb{I}_n - E[\mathcal{E}_t^2 \otimes E_{x,t}^2]. \end{aligned}$$

(v) First note that $E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right]$ can be written as

$$E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right] = E\begin{bmatrix} E[\varepsilon_{1,t}\varepsilon_{1,t}e_{x,t}e_{x,t}'] & \cdots & E[\varepsilon_{1,t}\varepsilon_{n,t}e_{x,t}e_{x,t}'] \\ \vdots & \ddots & \vdots \\ E[\varepsilon_{n,t}\varepsilon_{1,t}e_{x,t}e_{x,t}'] & \cdots & E[\varepsilon_{n,t}\varepsilon_{n,t}e_{x,t}e_{x,t}'] \end{bmatrix},$$

where every block is nonzero. First consider the n diagonal blocks which can be written as

$$\begin{bmatrix} E[\varepsilon_{1,t}^2 e_{1,x,t}^2] & & & \\ & \sigma_{1,0}^2 E[e_{2,x,t}^2] & & \\ & & \ddots & \\ & & & \sigma_{1,0}^2 E[e_{n,x,t}^2] \end{bmatrix}, \quad \begin{bmatrix} \sigma_{2,0}^2 E[e_{1,x,t}^2] & & & \\ & E[\varepsilon_{2,t}^2 e_{2,x,t}^2] & & \\ & & \ddots & \\ & & & \sigma_{2,0}^2 E[e_{n,x,t}^2] \end{bmatrix},$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad , \quad \begin{bmatrix} \sigma_{n,0}^2 E[e_{1,x,t}^2] & & & \\ & \ddots & & \\ & & \sigma_{n,0}^2 E[e_{n-1,x,t}^2] & \\ & & & E[\varepsilon_{n,t}^2 e_{n,x,t}^2] \end{bmatrix}.$$

The contribution of these diagonal blocks to the entire matrix $E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right]$ can be expressed as

$$E[\mathcal{E}_t^2 \otimes E_{x,t}^2].$$

Now consider the off-diagonal blocks. The $n-1$ remaining blocks on the first row of blocks are

$$\begin{bmatrix} 0 & \sigma_{1,0}\sigma_{2,0} & 0 & \cdots & 0 \\ \sigma_{2,0}\sigma_{1,0} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix}, \quad \cdots, \quad \begin{bmatrix} 0 & 0 & \cdots & 0 & \sigma_{1,0}\sigma_{n,0} \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ \sigma_{n,0}\sigma_{1,0} & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

These are also the $n-1$ remaining blocks on the first column of blocks. The remaining off-diagonal blocks can be expressed in a similar fashion. Overall, the contribution of these off-diagonal blocks to the entire matrix $E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right]$ can, after some algebra, be expressed as

$$-2(\Sigma_0^2 \otimes I_n)\mathbb{I}_n + (I_n \otimes \Sigma_0)\left[(vec(I_n)vec(I_n)') + K_{nn}\right](I_n \otimes \Sigma_0).$$

Putting the contributions of the diagonal and off-diagonal blocks together, the entire matrix $E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right]$ can be expressed as

$$E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right] = E[\mathcal{E}_t^2 \otimes E_{x,t}^2] - 2(\Sigma_0^2 \otimes I_n)\mathbb{I}_n + (I_n \otimes \Sigma_0)\left[(vec(I_n)vec(I_n)') + K_{nn}\right](I_n \otimes \Sigma_0).$$

This completes the proof. ■

Proof of the result $E\left[l_{\theta\theta,t}(\theta_0)\right] = -E[l_{\theta,t}(\theta_0)\, l_{\theta,t}'(\theta_0)]$. In order to show that $E\left[l_{\theta\theta,t}(\theta_0)\right] = -E[l_{\theta,t}(\theta_0)\, l_{\theta,t}'(\theta_0)]$, we'll go through the blocks one by one. Recall the property $B_0^{-1}u_t(\pi_0) = \varepsilon_t$, and also that x_{t-1} is independent of ε_t (and functions of ε_t).

$\underline{E[l_{\pi\pi,t}(\theta_0)]}$: Because $E[l_{\pi\pi,t}(\theta_0)]$ can be written as

$$E[l_{\pi\pi,t}(\theta_0)] = E[x_{t-1}x_{t-1}'] \otimes B_0^{-1\prime}\Sigma_0^{-1}E[e_{xx,t}]\Sigma_0^{-1}B_0^{-1},$$

it suffices to show that $E[e_{xx,t}] = -V_{e_x} = -Cov[e_{x,t}]$. This holds by Lemma B.1(i) and Lemma C.1(i).

$\underline{E[l_{\pi\beta,t}(\theta_0)]}$: Because $E[e_{x,t}] = 0$ by Lemma B.1(i) and using matrix algebra we obtain

$$\begin{aligned} E[l_{\pi\beta,t}(\theta_0)] &= E[x_{t-1}] \otimes \left[(I_n \otimes E[e_{x,t}'])(B_0^{-1\prime} \otimes \Sigma_0^{-1}B_0^{-1})H\right] \\ &\quad + E\left[x_{t-1} \otimes \left(B_0^{-1\prime}\Sigma_0^{-1}(\varepsilon_t' \otimes e_{xx,t})(I_n \otimes \Sigma_0^{-1}B_0^{-1})H\right)\right] \\ &= (E[x_{t-1}] \otimes B_0^{-1\prime}\Sigma_0^{-1})E[\varepsilon_t' \otimes e_{xx,t}](I_n \otimes \Sigma_0^{-1}B_0^{-1})H. \end{aligned}$$

The desired result follows because $E[\varepsilon_t' \otimes e_{xx,t}] = -E[\varepsilon_t' \otimes e_{x,t}e_{x,t}']$ by Lemma C.2(i).

$\underline{E[l_{\beta\beta,t}(\theta_0)]}$: Regarding $E[l_{\beta\beta,t}(\theta_0)]$, note that $l_{\beta\beta,t}(\theta_0)$ can be written as

$$\begin{aligned} l_{\beta\beta,t}(\theta_0) &= H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})(\varepsilon_t\varepsilon_t' \otimes e_{xx,t})(I_n \otimes \Sigma_0^{-1}B_0^{-1})H \\ &\quad + H'(I_n \otimes I_n)\left(\varepsilon_t e_{x,t}' \otimes I_n\right)\left(\Sigma_0^{-1}B_0^{-1} \otimes B_0^{-1\prime}\right)K_{nn}H \\ &\quad + H'K_{nn}(B_0^{-1\prime}\Sigma_0^{-1} \otimes B_0^{-1})\left(e_{x,t}\varepsilon_t' \otimes I_n\right)(I_n \otimes I_n)H \\ &\quad + H'(B_0^{-1} \otimes B_0^{-1\prime})K_{nn}H. \end{aligned}$$

Note that by Lemma B.1(v) $E[\varepsilon_t e_{x,t}'] = E[e_{x,t}\varepsilon_t'] = -\Sigma_0$, so that

$$\begin{aligned} E[l_{\beta\beta,t}(\theta_0)] &= H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})E[\varepsilon_t\varepsilon_t' \otimes e_{xx,t}](I_n \otimes \Sigma_0^{-1}B_0^{-1})H \\ &\quad - H'(B_0^{-1} \otimes B_0^{-1\prime})K_{nn}H - H'K_{nn}(B_0^{-1\prime} \otimes B_0^{-1})H + H'(B_0^{-1} \otimes B_0^{-1\prime})K_{nn}H \\ &= H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})E[\varepsilon_t\varepsilon_t' \otimes e_{xx,t}](I_n \otimes \Sigma_0^{-1}B_0^{-1})H - H'K_{nn}(B_0^{-1\prime} \otimes B_0^{-1})H. \end{aligned}$$

According to Lemma C.2(iv,v),

$$E[\varepsilon_t\varepsilon_t' \otimes e_{xx,t}] = -E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right] + (I_n \otimes \Sigma_0)\left[(vec(I_n)vec(I_n)') + K_{nn}\right](I_n \otimes \Sigma_0).$$

Therefore

$$\begin{aligned} E[l_{\beta\beta,t}(\theta_0)] &= -H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right](I_n \otimes \Sigma_0^{-1}B_0^{-1})H \\ &\quad + H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})(I_n \otimes \Sigma_0)\left[(vec(I_n)vec(I_n)') + K_{nn}\right](I_n \otimes \Sigma_0)(I_n \otimes \Sigma_0^{-1}B_0^{-1})H \\ &\quad - H'K_{nn}(B_0^{-1\prime} \otimes B_0^{-1})H. \end{aligned}$$

Noting that

$$H'(I_n \otimes B_0^{-1\prime})vec(I_n)vec(I_n)'(I_n \otimes B_0^{-1})H = H'vec(B_0^{-1\prime})vec(B_0^{-1\prime})'H$$

and

$$H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})(I_n \otimes \Sigma_0)K_{nn}(I_n \otimes \Sigma_0)(I_n \otimes \Sigma_0^{-1}B_0^{-1})H \overset{\text{L 9.2.2(5)(a)}}{=} H'K_{nn}(B_0^{-1\prime} \otimes B_0^{-1})H$$

gives

$$\begin{aligned} E[l_{\beta\beta,t}(\theta_0)] &= -H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})E\left[\varepsilon_t\varepsilon_t' \otimes e_{x,t}e_{x,t}'\right](I_n \otimes \Sigma_0^{-1}B_0^{-1})H + H'vec(B_0^{-1\prime})vec(B_0^{-1\prime})'H \\ &= -Cov\left[l_{\beta,t}(\theta_0)\right]. \end{aligned}$$

$\underline{E[l_{\pi\sigma,t}(\theta_0)]}$: By Lemma B.1(i), $E[E_{x,t}] = 0$, and by Lemma C.1(vi), $E[\varepsilon_{i,t}e_{i,xx,t}] = -E\left[\varepsilon_{i,t}e^2_{i,x,t}\right]$, so that

$$\begin{aligned} E[l_{\pi\sigma,t}(\theta_0)] &= E[x_{t-1}] \otimes B_0^{-1\prime}\left[\Sigma_0^{-2}E[E_{x,t}] + \Sigma_0^{-3}E[e_{xx,t}\mathcal{E}_t]\right] \\ &= -E[x_{t-1}] \otimes B_0^{-1\prime}\Sigma_0^{-3}E[e_{x,t}(\varepsilon_t \odot e_{x,t})'] \\ &= -E[l_{\pi,t}(\theta_0)\, l'_{\sigma,t}(\theta_0)]. \end{aligned}$$

$\underline{E[l_{\beta\sigma,t}(\theta_0)]}$: Recall that Σ_0, $e_{xx,t}$, and $\mathcal{E}_t$ are diagonal matrices and write $l_{\beta\sigma,t}(\theta_0)$ as

$$\begin{aligned} l_{\beta\sigma,t}(\theta_0) &= H'(B_0^{-1} \otimes B_0^{-1\prime})(u_t(\pi_0) \otimes \left[\Sigma_0^{-2}E_{x,t} + \Sigma_0^{-3}e_{xx,t}\mathcal{E}_t\right]) \\ &= H'(I_n \otimes B_0^{-1\prime})(\varepsilon_t \otimes E_{x,t})\Sigma_0^{-2} + H'(I_n \otimes B_0^{-1\prime})(\varepsilon_t \otimes e_{xx,t}\mathcal{E}_t)\Sigma_0^{-3} \\ &= H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})(\varepsilon_t \otimes E_{x,t})\Sigma_0^{-1} + H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})(\varepsilon_t \otimes e_{xx,t}\mathcal{E}_t)\Sigma_0^{-2} \\ &= H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})\left[(\varepsilon_t \otimes E_{x,t})\Sigma_0^{-1} + (\varepsilon_t \otimes e_{xx,t}\mathcal{E}_t)\Sigma_0^{-2}\right]. \end{aligned}$$

Thus, using Lemma C.2(ii,iii),

$$\begin{aligned} E[l_{\beta\sigma,t}(\theta_0)] &= H'(I_n \otimes B_0^{-1\prime})E[(\varepsilon_t \otimes E_{x,t})]\Sigma_0^{-2} + H'(I_n \otimes B_0^{-1\prime})E[(\varepsilon_t \otimes e_{xx,t}\mathcal{E}_t)]\Sigma_0^{-3} \\ &= -H'(I_n \otimes B_0^{-1\prime})(\Sigma_0 \otimes I_n)\mathbb{I}_n\Sigma_0^{-2} + 2H'(I_n \otimes B_0^{-1\prime})(\Sigma_0^2 \otimes I_n)\mathbb{I}_n\Sigma_0^{-3} \\ &\quad -H'(I_n \otimes B_0^{-1\prime})(E[\mathcal{E}_t^2E_{x,t}^2] \otimes I_n)\mathbb{I}_n\Sigma_0^{-3} \\ &= H'(I_n \otimes B_0^{-1\prime})\left\{-(\Sigma_0 \otimes I_n)\mathbb{I}_n + 2(\Sigma_0^2 \otimes I_n)\mathbb{I}_n\Sigma_0^{-1} - (E[\mathcal{E}_t^2E_{x,t}^2] \otimes I_n)\mathbb{I}_n\Sigma_0^{-1}\right\}\Sigma_0^{-2}, \end{aligned}$$

where $\mathbb{I}_n\Sigma_0^{-1} = (\Sigma_0^{-1} \otimes I_n)\mathbb{I}_n$ and hence the term in curly brackets can be expressed as

$$\begin{aligned} & -(\Sigma_0 \otimes I_n)\mathbb{I}_n + 2(\Sigma_0^2 \otimes I_n)\mathbb{I}_n\Sigma_0^{-1} - (E[\mathcal{E}_t^2E_{x,t}^2] \otimes I_n)\mathbb{I}_n\Sigma_0^{-1} \\ =\; & -(\Sigma_0 \otimes I_n)\mathbb{I}_n + 2(\Sigma_0^1 \otimes I_n)\mathbb{I}_n - (\Sigma_0^{-1}E[\mathcal{E}_t^2E_{x,t}^2] \otimes I_n)\mathbb{I}_n \\ =\; & (\Sigma_0 \otimes I_n)\mathbb{I}_n - (\Sigma_0^{-1}E[\mathcal{E}_t^2E_{x,t}^2] \otimes I_n)\mathbb{I}_n. \end{aligned}$$

On the other hand, note that

$$\begin{aligned} Cov\left[l_{\beta,t}(\theta_0), l_{\sigma,t}(\theta_0)\right] &= H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})E\left[(\varepsilon_t \otimes e_{x,t})(\varepsilon_t \odot e_{x,t})'\right]\Sigma_0^{-2} - H'vec(B_0^{-1\prime})\sigma_0'\Sigma_0^{-2} \\ &= H'(I_n \otimes B_0^{-1\prime}\Sigma_0^{-1})E\left[(\varepsilon_t \otimes e_{x,t})(\varepsilon_t \odot e_{x,t})'\right]\Sigma_0^{-2} - H'(I_n \otimes B_0^{-1\prime})vec(I_n)\sigma_0'\Sigma_0^{-2} \\ &= H'(I_n \otimes B_0^{-1\prime})\left\{(I_n \otimes \Sigma_0^{-1})E\left[(\varepsilon_t \otimes e_{x,t})(\varepsilon_t \odot e_{x,t})'\right] - vec(I_n)\sigma_0'\right\}\Sigma_0^{-2}. \end{aligned}$$

Let's examine the expression in curly brackets piece by piece:

$$E\left[(\varepsilon_t \otimes e_{x,t})(\varepsilon_t \odot e_{x,t})'\right] = E\begin{bmatrix} \iota_1\varepsilon_{1,t}e_{1,x,t}(\varepsilon_t \odot e_{x,t})' \\ \vdots \\ \iota_n\varepsilon_{n,t}e_{n,x,t}(\varepsilon_t \odot e_{x,t})' \end{bmatrix}$$

$$= \begin{bmatrix} \iota_1(E[\varepsilon_{1,t}^2 e_{1,x,t}^2] & \sigma_{1,0}\sigma_{2,0} & \cdots & \sigma_{1,0}\sigma_{n,0}) \\ \vdots & \vdots & \vdots & \vdots \\ \iota_n(\sigma_{1,0}\sigma_{n,0} & \cdots & \sigma_{n-1,0}\sigma_{n,0} & E[\varepsilon_{n,t}^2 e_{n,x,t}^2]) \end{bmatrix},$$

$$(I_n \otimes \Sigma_0^{-1})E\left[(\varepsilon_t \otimes e_{x,t})(\varepsilon_t \odot e_{x,t})'\right]$$

$$= \begin{bmatrix} \Sigma_0^{-1}\iota_1(E[\varepsilon_{1,t}^2 e_{1,x,t}^2] & \sigma_{1,0}\sigma_{2,0} & \cdots & \sigma_{1,0}\sigma_{n,0}) \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_0^{-1}\iota_n(\sigma_{1,0}\sigma_{n,0} & \cdots & \sigma_{n-1,0}\sigma_{n,0} & E[\varepsilon_{n,t}^2 e_{n,x,t}^2]) \end{bmatrix}$$

$$= \begin{bmatrix} \iota_1(\sigma_{1,0}^{-1}E[\varepsilon_{1,t}^2 e_{1,x,t}^2] & \sigma_{2,0} & \cdots & \sigma_{n,0}) \\ \vdots & \vdots & \vdots & \vdots \\ \iota_n(\sigma_{1,0} & \cdots & \sigma_{n-1,0} & \sigma_{n,0}^{-1}E[\varepsilon_{n,t}^2 e_{n,x,t}^2]) \end{bmatrix},$$

$$(I_n \otimes \Sigma_0^{-1})E\left[(\varepsilon_t \otimes e_{x,t})(\varepsilon_t \odot e_{x,t})'\right] - vec(I_n)\sigma_0'$$

$$= \begin{bmatrix} \iota_1(\sigma_{1,0}^{-1}E[\varepsilon_{1,t}^2 e_{1,x,t}^2] & \sigma_{2,0} & \cdots & \sigma_{n,0}) \\ \vdots & \vdots & \vdots & \vdots \\ \iota_n(\sigma_{1,0} & \cdots & \sigma_{n-1,0} & \sigma_{n,0}^{-1}E[\varepsilon_{n,t}^2 e_{n,x,t}^2]) \end{bmatrix} - \begin{bmatrix} \iota_1[\sigma_{1,0} : \sigma_{2,0} : \cdots : \sigma_{n,0}] \\ \vdots \\ \iota_n[\sigma_{1,0} : \sigma_{2,0} : \cdots : \sigma_{n,0}] \end{bmatrix}$$

$$= \begin{bmatrix} \iota_1(\sigma_{1,0}^{-1}(E[\varepsilon_{1,t}^2 e_{1,x,t}^2] - \sigma_{1,0}^2)) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \iota_n(\sigma_{n,0}^{-1}(E[\varepsilon_{n,t}^2 e_{n,x,t}^2] - \sigma_{n,0}^2)) \end{bmatrix}$$

$$= (\Sigma_0^{-1}(E[\mathcal{E}_t^2 E_{x,t}^2] - \Sigma_0^2) \otimes I_n)\mathbb{I}_n.$$

These calculations show that $E[l_{\beta\sigma,t}(\theta_0)] = -Cov[l_{\beta,t}(\theta_0), l_{\sigma,t}(\theta_0)]$.

$E[l_{\sigma\sigma,t}(\theta_0)]$: Note that $l_{\sigma\sigma,t}(\theta_0)$ is a diagonal matrix with diagonal elements

$$\sigma_{i,0}^{-2} + 2\sigma_{i,0}^{-3}\varepsilon_{i,t}e_{i,x,t} + \sigma_{i,0}^{-4}\varepsilon_{i,t}^2 e_{i,xx,t}.$$

By Lemma B.1(v), $E[\varepsilon_{i,t}e_{i,x,t}] = -\sigma_{i,0}$. By Lemma C.1(vii), $E[\varepsilon_{i,t}^2 e_{i,xx,t}] = 2\sigma_{i,0}^2 - E\left[\varepsilon_{i,t}^2 e_{i,x,t}^2\right]$. Thus, the diagonal elements of $E[l_{\sigma\sigma,t}(\theta_0)]$ are

$$\sigma_{i,0}^{-2} - 2\sigma_{i,0}^{-2} + \sigma_{i,0}^{-4}(2\sigma_{i,0}^2 - E\left[\varepsilon_{i,t}^2 e_{i,x,t}^2\right])$$
$$= -\sigma_{i,0}^{-4}\left(E\left[\varepsilon_{i,t}^2 e_{i,x,t}^2\right] - \sigma_{i,0}^2\right),$$

which yields the desired result.

$\underline{E[l_{\pi\lambda,t}(\theta_0)]}$: By Lemma C.1(iv), $E[e_{x\lambda,t}] = -E[e_{x,t}e'_{\lambda,t}]$, and thus

$$\begin{aligned} E[l_{\pi\lambda,t}(\theta_0)] &= -E[x_{t-1}] \otimes B_0^{-1'}\Sigma_0^{-1}E[e_{x\lambda,t}] \\ &= E[x_{t-1}] \otimes B_0^{-1'}\Sigma_0^{-1}E[e_{x,t}e'_{\lambda,t}], \end{aligned}$$

establishing the desired result.

$\underline{E[l_{\beta\lambda,t}(\theta_0)]}$: Note that

$$l_{\beta\lambda,t}(\theta_0) = -H'(I_n \otimes B_0^{-1'}\Sigma_0^{-1})(\varepsilon_t \otimes e_{x\lambda,t}).$$

By Lemma C.1(v), $E[\varepsilon_t \otimes e_{x\lambda,t}] = -E[\varepsilon_t \otimes e_{x,t}e'_{\lambda,t}]$, yielding the desired result.

$\underline{E[l_{\sigma\lambda,t}(\theta_0)]}$: Note that $l_{\sigma\lambda,t}(\theta_0) = -\Sigma_0^{-2}\mathcal{E}_t e_{x\lambda,t}$, where $\mathcal{E}_t e_{x\lambda,t}$ is a block diagonal matrix with diagonal blocks $\varepsilon_{i,t}e_{x\lambda_i,t}$. By Lemma C.1(v), $E[\varepsilon_{i,t}e_{x\lambda_i,t}] = -E[\varepsilon_{i,t}e_{i,x,t}e'_{i,\lambda_i,t}]$, and thus (also using Lemma B.1(iii))

$$E[l_{\sigma\lambda,t}(\theta_0)] = -\Sigma_0^{-2}E[\mathcal{E}_t e_{x\lambda,t}] = \Sigma_0^{-2}E[(\varepsilon_t \odot e_{x,t})e'_{\lambda,t}] = -E[l_{\sigma,t}(\theta_0)\, l'_{\lambda,t}(\theta_0)].$$

$\underline{E[l_{\lambda\lambda,t}(\theta_0)]}$: By Lemma C.1(iii) and Lemma B.1(iii)

$$E[l_{\lambda\lambda,t}(\theta_0)] = E[e_{\lambda\lambda,t}] = -E[e_{\lambda,t}e'_{\lambda,t}] = -V_{e_\lambda},$$

as desired. ■

Additional references

Aliprantis, C. D., and O. Burkinshaw (1998): *Principles of Real Analysis*, 3rd ed., Academic Press, San Diego.

Dudley, R. M. (2002): *Real Analysis and Probability.* Cambridge University Press, Cambridge.

Lütkepohl, H. (1996): *Handbook of Matrices.* Wiley, Chichester.

Seber, G. A. F. (2008): *A Matrix Handbook for Statisticians.* Wiley, Hoboken.

Research Papers 2013

CREATES
Center for Research in Econometric Analysis of Time Series

2014-57: Eduardo Rossi and Paolo Santucci de Magistris: Indirect inference with time series observed with error

2014-58: Anders Bredahl Kock and Haihan Tang: Inference in High-dimensional Dynamic Panel Data Models

2015-01 Tom Engsted, Simon J. Hviid and Thomas Q. Pedersen: Explosive bubbles in house prices? Evidence from the OECD countries

2015-02: Tim Bollerslev, Andrew J. Patton and Wenjing Wang: Daily House Price Indices: Construction, Modeling, and Longer-Run Predictions

2015-03: Christian M. Hafner, Sebastien Laurent and Francesco Violante: Weak diffusion limits of dynamic conditional correlation models

2015-04: Maria Eugenia Sanin, Maria Mansanet-Bataller and Francesco Violante: Understanding volatility dynamics in the EU-ETS market

2015-05: Peter Christoffersen and Xuhui (Nick) Pan: Equity Portfolio Management Using Option Price Information

2015-06: Peter Christoffersen and Xuhui (Nick) Pan: Oil Volatility Risk and Expected Stock Returns

2015-07: Peter Christoffersen, Bruno Feunou and Yoontae Jeon: Option Valuation with Observable Volatility and Jump Dynamics

2015-08: Alfonso Irarrazabal and Juan Carlos Parra-Alvarez: Time-varying disaster risk models: An empirical assessment of the Rietz-Barro hypothesis

2015-09: Daniela Osterrieder, Daniel Ventosa-Santaulària and Eduardo Vera-Valdés: Unbalanced Regressions and the Predictive Equation

2015-10: Laurent Callot, Mehmet Caner, Anders Bredahl Kock and Juan Andres Riquelme: Sharp Threshold Detection Based on Sup-norm Error rates in High-dimensional Models

2015-11: Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen and Anders Rahbek: Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)

2015-12: Tommaso Proietti, Martyna Marczak and Gianluigi Mazzi: EuroMInd-D: A Density Estimate of Monthly Gross Domestic Product for the Euro Area

2015-13: Michel van der Wel, Sait R. Ozturk and Dick van Dijk: Dynamic Factor Models for the Volatility Surface

2015-14: Tim Bollerslev, Andrew J. Patton and Rogier Quaedvlieg: Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting

2015-15: Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Effects of Macroeconomic Uncertainty upon the Stock and Bond Markets

2015-16: Markku Lanne, Mika Meitz and Pentti Saikkonen: Identification and estimation of non-Gaussian structural vector autoregressions