Econometric Theory, **40**, 2024, 705–751.
doi:10.1017/S0266466622000536

INSTRUMENTAL VARIABLES INFERENCE IN A SMALL-DIMENSIONAL VAR MODEL WITH DYNAMIC LATENT FACTORS

FEDERICO CARLINI
LUISS University Rome

PATRICK GAGLIARDINI
Università della Svizzera italiana

We study semiparametric inference in a small-dimensional vector autoregressive (VAR) model of order p augmented by unobservable common factors with a dynamic described by a VAR process of order q. This state-space specification is useful to measure separately the direct causality effects and the responses to dynamic common factors. We show that the state-space parameters are identifiable from the autocovariance function of the observed process. We estimate the model by means of a multistep procedure in closed-form, which combines an eigenvalue–eigenvector matrix decomposition and a linear instrumental variable estimation allowing for Hansen–Sargan specification tests. We study the asymptotic and finite-sample properties of the parameter estimators and of rank tests for selecting the number of unobservable factors and VAR orders. In an empirical illustration, we investigate the dynamic common factors and the spillover effects that explain the co-movements among the log daily realized volatilities of four European stock market indices.

1. INTRODUCTION

State-space systems are the backbone of many macroeconomic and financial models because they enable to incorporate the time evolution of latent variables (see Hamilton, 1994, for a review). The econometrics and statistics literatures have devoted great efforts in studying the statistical properties of these dynamical systems in terms of parameter identification and estimation. Hannan (1971) establishes identifiability conditions for the parameters of a vector autoregressive

We are grateful to the Editor, a Co-Editor, three anonymous referees, M. Deistler, C. Gourieroux, S. Johansen, L. Mancini, P. Santucci de Magistris, O. Scaillet, and participants at the 5th Empirical Finance Workshop in Cergy, the ESEM 2018 Conference in Cologne, the SFI 2019 Research Day in Gerzensee, the First Rome Workshop of Time Series and Financial Econometrics, and seminars at LUISS University, Durham University, and Geneva University for useful comments. We thank F. Diebold and K. Yilmaz for kindly providing us with their datasets used in a previous version of the paper circulated under the title of "Vector Autoregressive Model with Dynamic Factors." The first version of this paper has been written while F. Carlini was a postdoctoral fellow at the Faculty of Economics of the Università della Svizzera italiana, Lugano, Switzerland. We gratefully acknowledge the Swiss National Science Foundation for Grant 105218-162633. Address correspondence to Patrick Gagliardini, Faculty of Economics, Università della Svizzera italiana, Lugano, Switzerland; e-mail: patrick.gagliardini@usi.ch.

moving average model with exogenous variables (VARMAX). This result is strictly linked with the identifiability of the parameters of a state-space model, since there exists a correspondence between state-space and VARMAX models (see Akaike, 1974). Glover and Willems (1974) are among the first to propose a (global) identifiability condition for the parameters of a linear state-space system. They find that a state-space system is identifiable if it is observable and reachable (i.e., minimal) and a rank condition on a matrix involving transformations of the data generating process (DGP) parameters is satisfied. More recently, Komunjer and Ng (2011) find local identifiability conditions for structural state-space parameters in a class of dynamic stochastic general equilibrium models. Furthermore, Hannan and Deistler (1988), among others, discuss the asymptotic properties of the quasi-maximum likelihood estimator (QMLE) in linear state-space models.

Despite their popularity, inference in linear state-space systems is confronted with theoretical and practical difficulties. Indeed, identifiability conditions are often hard to make explicit in terms of the DGP parameters, the QMLE is difficult to compute when the dimension of the parameter space is large, and formal inferential procedures for model selection (e.g., inference on the number of latent states and the memory length) and specification testing are scarce. A recent literature addresses related issues in a Bayesian framework (see, e.g., Chan, Eisenstat, and Koop, 2016, and the references therein). The goal of this paper is to develop a novel fully fledged frequentist methodology for identification, estimation, model selection, and specification testing in a general class of linear state-space systems. The key novelty is the use of orthogonality restrictions from instrumental variables internal to the model yielding closed-form estimators.

We consider a multivariate linear state-space model with n observed variables, $K < n$ latent variables (factors), and p predetermined variables that are the lags of the endogenous observed process. From the vantage point of factor models, our specification resembles the reduced form of the factor structural VAR model of Stock and Watson (2005), but in our model, the latent factors are dynamic. Moreover, our state-space specification corresponds to a factor-augmented vector autoregressive (FAVAR) model (Bernanke, Boivin, and Eliasz, 2005). However, the latent variables in our model are not introduced to solve the "limited information problem" of standard small-dimensional vector autoregressive (VAR) models but rather to capture dynamic common shocks. Moreover, while in the FAVAR literature it is typically assumed that an external large dataset is available from which to extract consistent estimates of the factor values using a static factor representation (e.g., Bai, Li, and Lu, 2016), in our paper, this external source of information is not necessary for statistical inference. An interesting feature of our specification for applications in Economics and Finance, such as financial stability analysis, is the possibility to empirically disentangle two conceptually distinct sources of co-movements across the series when measuring impulse responses (Darolles, Dubecq, and Gouriéroux, 2014, Darolles and Gourieroux (2015)) and defining measures of network interconnectedness. These sources are (i) the exogenous latent process, that we interpret as a systematic unobserved risk factor with a pervasive effect on the individual series, and (ii) the lags of the endogenous observed process, that we interpret as direct causality effects—i.e., lagged contagion—across the series. Identification of the model parameters, notably the matrices of factor loadings and autoregressive coefficients, is made difficult by the overlapping effects of contagion and common latent factors (Manski, 1993, reflection problem).

We point out that this paper presents results for a finite-dimensional dynamical system. When the number of series n is large, the VAR model augmented by a latent factor can be approximated by generalized dynamic factor models (see, e.g., Forni et al., 2000; Hallin and Lippi, 2013), leading to different identification strategies.

The methodological contributions of this paper are manifold. First, we show under which conditions the parameters of the state-space specification are globally identifiable from the autocovariance function of the observed process. The identification strategy is constructive and relies on the orthogonality restrictions for instrumental variable (IV) estimation. A key step consists in proving that an $(n-K)$-dimensional White Noise component obtained from the observable vector and its lags is a valid instrument for estimating the autoregressive coefficients in the measurement equation. This identification strategy is novel compared with the previous literature on state-space models, notably the pioneering work of Darolles et al. (2014). Second, we provide estimators for factor loadings, VAR coefficients, and error variances that are in closed form (up to eigenvalues–eigenvectors decomposition of matrices of small dimension) and easy to implement. Third, we show that the estimators are consistent and asymptotically normal when the dimension of the vector of observables is fixed and the time series dimension of the sample increases. We provide a detailed comparison of our estimators with the QMLE in terms of asymptotic standard errors and show that efficiency loss with our approach is moderate, except in a neighborhood of DGP parameter values which are not identified. Fourth, we provide statistical tests for conducting inference on the number of unobservable factors and the memory length of VAR models. Fifth, we define Hansen–Sargan statistics for testing correct model specification.

We illustrate our methodology with an empirical application on a system of daily realized volatilities of four European stock market indices: CAC 40, OMX Stockholm 30, IBEX 35, and DAX 30. Our procedure selects a state-space model with four lags in the VAR specification that capture spillover effects in realized volatilities across markets, and two dynamic latent common factors with six lags in their VAR dynamics. In this state-space specification with 81 parameters, our multistep estimator in closed form is a useful alternative to QMLE.

Our empirical illustration relates to the vast literature on contagion.[1] Billio et al. (2012) estimate measures of connectedness among financial institutions using,

[1]There is not a neat consensus in the literature on the exact meaning of the word "contagion." For instance, in a literature that contrasts "contagion" versus "interdependence," Pesaran and Pick (2007) refer to Masson (1999) and Forbes and Rigobon (2002) and distinguish among "monsoonal effects," "spillovers," and "pure contagion effects." In our paper, the meaning of "contagion" is more similar to spillovers. The theoretical literature on financial contagion models this phenomenon as a widespread transmission of shocks among institutions through various market or

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

e.g., pairwise Granger causality. Diebold and Yılmaz (2014) use the generalized impulse response functions (IRFs) in Koop, Pesaran, and Potter (1996) and Pesaran and Shin (1998) to understand in a VAR model the amount of connectedness of different U.S. financial institutions. More recently, Adrian and Brunnermeier (2016), Acharya et al. (2017), and Brownlees and Engle (2017) estimate measures of association in market tail risks. Ait-Sahalia, Cacho-Diaz, and Laeven (2015) model contagion via mutually exciting jump processes in which past jumps in one series increase the jump intensity of all series. Related to our volatility spillovers application, in an early important contribution on foreign exchange markets, Engle, Ito, and Lin (1990) examine the impact of news in one market on volatility in other markets. The distinctive feature of our analysis compared to the aforementioned literature is the empirical disentangling of dynamic common factors from contagion when measuring interconnectedness in a similar vein as Darolles et al. (2014).[2]

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 proves parameter identification from various representations of the model. Section 4 presents the multistep estimation procedure and establishes the large sample properties of the estimators and the model specification testing procedures. Section 5 studies tests for the selection of the number of dynamic latent factors and of the number of VAR lags. Section 6 presents the results of our Monte Carlo experiments. Section 7 provides an empirical illustration with the series of daily realized volatilities of four European stock market indices. Section 8 concludes. We denote our assumptions with letters M, ID, IR, LS, and SEL to distinguish the conditions to define the model (M), prove identification (ID), restrict the parameter space and normalize the latent factor (IR, for identification restriction), derive the large sample properties (LS), and select model orders (SEL), respectively. Proofs of theorems and propositions are provided in Appendixes A and B. In the Supplementary Material available at Cambridge Core (www.cambridge.org/core/journals/econometric-theory), we provide the proofs of technical lemmas and additional theoretical and empirical results, and report the results of the Monte Carlo experiments.

2. VAR MODEL WITH UNOBSERVABLE DYNAMIC FACTORS

We consider a VAR(p) model augmented by unobservable dynamic factors. The state-space representation is defined in the next two assumptions.

information mechanisms (see, e.g., King and Wadhwani, 1990, Allen and Gale, 2000; Eisenberg and Noe, 2001; Elliott, Golub, and Jackson, 2014; Acemoglu, Ozdaglar, and Tahbaz-Salehi, 2015; Trevino, 2020).

[2]Besides financial stability analysis, other strands of the economics literature have devoted interest to disentangling linkages from common factors. In the multisector Real Business Cycle model of Long and Plosser (1983), in equilibrium the sectoral log-output processes follow a VAR(1) dynamics. The autoregressive matrix is given by the elasticities of commodity outputs with respect to commodity inputs and is responsible for the transmission of productivity shocks across sectors. Foerster, Sarte, and Watson (2011) write the productivity shocks as a sum of an idiosyncratic shock and a common factor driven by systematic shocks, leading to a model similar to the one studied statistically in this paper.

Assumption M.1. The measurement equation is

$$C(L)Y_t = Bf_t + u_t, \tag{2.1}$$

where Y_t is an n-dimensional vector of endogenous observable variables and f_t is a K-dimensional vector of exogenous unobservable factors. (i) The errors $u_t \sim WN(0, \Sigma_u)$ are a (weak) White Noise process with positive definite variance matrix Σ_u[3]. (ii) Process (f_t) is covariance stationary and uncorrelated at all leads and lags with (u_t). (iii) The matrix lag polynomial $C(L) = I_n - \sum_{j=1}^{p} C_j L^j$, where $p \geq 1$ and L is the lag operator and C_j, for $j = 1, \ldots, p$, are $n \times n$ matrices, is such that $\det C(z)$ has roots outside the complex unit circle. (iv) The number of unobservable factors is strictly smaller than the number of observable variables, i.e., $K < n$, and the $n \times K$ matrix B of factor loadings has full column rank K.

Assumption M.2. The transition equation is

$$\Phi(L)f_t = v_t, \tag{2.2}$$

where (i) the error terms $v_t \sim WN(0, \Sigma_v)$, with Σ_v positive definite, are uncorrelated at all leads and lags with process (u_t), and (ii) the matrix lag polynomial $\Phi(L) = I_K - \sum_{j=1}^{q} \Phi_j L^j$, with $q \geq 1$, is such that $\det \Phi(z)$ has roots outside the complex unit circle.

Assumptions M.1 and M.2 define a general linear state-space model with lagged endogenous variables (e.g., Hamilton, 1994). We do not assume specific distributions for the innovations u_t and v_t, which are not necessarily i.i.d. across time. The framework allows for nonlinear and conditionally heteroskedastic processes. We state Assumptions M.1 and M.2 separately since we show some identification results in Section 3 in the semiparametric framework of Assumption M.1 alone without necessitating to specify a VAR(q) dynamics for the latent factor. In the full model, we denote by $\theta \in \Theta_{p,q,K}$ the vector of unknown parameters, which consists of the unique elements in matrices B, $C_j, j = 1, \ldots, p$, $\Phi_j, j = 1, \ldots, q$, Σ_u and Σ_v subject to the normalization restrictions introduced in Section 3, and denote by $\theta_0 \in \Theta_{p,q,K}$ the true value in the DGP.

In our empirical illustration to stock market realized volatility series, we interpret the dynamic latent factor f_t as a vector of systematic shocks affecting the entire economy. The autoregressive matrices C_j, for $j = 1, \ldots, p$, yield lagged contagion effects across markets akin to volatility spillovers.

Remark 1. For the variance matrix Σ_u, we can consider three (constrained) specifications corresponding to different structural interpretations: (i) Σ_u is diagonal; (ii) $\Sigma_u = \Lambda\Lambda' + D$, where Λ is an $n \times s$ matrix, with $s < n$, and D is a diagonal matrix; (iii) $\Sigma_u = Q_0^{-1} D (Q_0^{-1})'$, where Q_0 is nonsingular and D is diagonal. In case (i), the error term u_t corresponds to a vector of idiosyncratic shocks. The covariance between $Y_{i,t}$ and $Y_{j,t}$, for $i \neq j$, conditional on $Y_{t-1}, Y_{t-2}, \ldots$,

[3]That is, it holds $\boldsymbol{E}(u_t) = 0$, $V(u_t) = \Sigma_u$, and $\boldsymbol{E}(u_t u'_{t-j}) = 0$, for $j \neq 0$.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

stems from the common factor f_t only. In case (ii), the error term u_t contains s static factors. In case (iii), our model is the reduced form of a structural dynamic simultaneous equations system with latent factors. Here, matrix D is the variance–covariance matrix of the structural shocks and Q_0 is the matrix of the contemporaneous contagion effects.

Darolles et al. (2014) and Darolles and Gourieroux (2015) consider the model in Assumptions M.1 and M.2 with $p = q = 1$ for the study of the impulse response functions (IRFs) in a framework similar to our empirical illustration. They define an estimator for the loadings matrix B based on an eigenvalue–eigenvector decomposition. In their empirical application, they estimate matrix C_1 by the QMLE with a single latent factor following linear autoregressive dynamics. In this paper, we construct internal instruments and use a linear IV approach to estimate matrices $C_1, \ldots, C_p$ semiparametrically and in closed form. Besides the computational appeal, the interest in the semiparametric approach is to conduct inference on the contagion matrices beyond linear autoregressive or moving average specifications for the latent factor process.

Our model with finite dimension n cannot be written as a generalized dynamic factor model. Indeed, the process in (2.1) and (2.2) can be decomposed as $Y_t = C(L)^{-1}u_t + C(L)^{-1}B\Phi(L)^{-1}v_t$, namely as a sum of a VAR process with innovations u_t and a dynamic factor model with innovations v_t. In general, the components of VAR process $C(L)^{-1}u_t$ are mutually correlated (unless $C(L)$ is diagonal) and the process defined in Assumptions M.1 and M.2 has not a generalized dynamic factor model representation (see, e.g., Forni et al., 2000; Forni and Lippi, 2001). For large n, the generalized dynamic factor representation yields powerful implications for identification and estimation.

The state-space model in equations (2.1) and (2.2) admits a Markovian representation in which vector $(\boldsymbol{Y}_t', \boldsymbol{f}_t)'$ follows a VAR(1) model in companion form, where $\boldsymbol{Y}_t := [Y_t', \ldots, Y_{t-p+1}']'$ and $\boldsymbol{f}_t := [f_t', \ldots, f_{t-q+1}']'$ (see Appendix A.1). Moreover, Assumption M.1(i)–(iii) implies that process $(\boldsymbol{Y}_t', \boldsymbol{f}_t')'$ is covariance stationary and causal. Finally, we can rewrite equations (2.1) and (2.2) as $Y_t = \boldsymbol{C}\boldsymbol{Y}_{t-1} + Bf_t + u_t$ and $f_t = \boldsymbol{\Phi}\boldsymbol{f}_{t-1} + v_t$, where $\boldsymbol{C} := [C_1 \, : \, \ldots \, : \, C_p]$ and $\boldsymbol{\Phi} = [\Phi_1 \, : \, \ldots \, : \, \Phi_q]$. We use those compact notations for the analysis below.

3. IDENTIFICATION FROM THE AUTOCOVARIANCE FUNCTION

In this section, we study the global identification of the parameters in the state-space model in equations (2.1) and (2.2). Some parameter transformations are not identifiable simply because of the rotational invariance of the model induced by the unobservability of the factors. Similarly as in the factor models literature, we deal with this indeterminacy by introducing an identification restriction on the factor loadings.

Assumption IR.1. Matrix B is such that $B = [B_1' : I_K]'$, where B_1 is the upper $(n-K) \times K$ block.

The condition in Assumption IR.1 is important to find asymptotic results in Section 4. However, this assumption is not invariant to the ordering of the observable variables. A general discussion of this issue and a solution for tractable inference on the parameter B in a Bayesian framework is provided by Chan, Leon-Gonzalez, and Strachan (2018).[4]

Propositions 1–3 and Theorem 1 in this section provide conditions for the global identification of the model parameters from a finite number of autocovariances of the observable process $\{Y_t\}$. More precisely, under the conditions detailed below, we show that there exists an integer h^* such that

$$\Gamma(h;\theta) = \Gamma(h;\theta_0), \quad \forall h \in \mathbb{N} \text{ with } |h| \leq h^* \text{ and for } \theta \in \Theta_{p,q,K} \Rightarrow \theta = \theta_0,$$

where $\Gamma(\cdot;\theta)$ denotes the autocovariance function of process $\{Y_t\}$ with parameter θ. Our identification strategy is constructive in the sense that we characterize explicitly the mapping

$$\theta_0 = \tau(\gamma_0), \tag{3.1}$$

which links the true parameter vector θ_0 to the vector γ_0 of the different elements in the autocovariance matrices of the observable process $\{Y_t\}$ up to order h^*. The mapping τ only involves standard matrix operations (including spectral decomposition). This mapping straightforwardly implies the estimators by the analogy principle (see the next section). Sections 3.1–3.3 deal with the identification of matrix parameters B, $\boldsymbol{C}$, and $\boldsymbol{\Phi}$, respectively. We clarify which subvectors of parameters are semiparametrically identifiable with unrestricted factor dynamics. In Section 3.4, we illustrate the identification assumptions in an example. In Section 3.5, we compare our findings with the implications of available results in the literature on identification in linear state-space models.

3.1. Semiparametric Identification of the Factor Loadings Matrix B from a Pseudo-Model

A key insight in Darolles et al. (2014) is the use of a pseudo-model for the identification of the factor loadings in state-space specifications (2.1) and (2.2) with $p = q = 1$. A similar argument applies for generic VAR orders p and q. Let us consider the following VAR($p+1$) pseudo-model:

$$Y_t = A_1^* Y_{t-1} + \cdots + A_{p+1}^* Y_{t-p-1} + u_t^*, \tag{3.2}$$

where error term u_t^* is orthogonal to lags Y_{t-1}, Y_{t-2},…,Y_{t-p-1}. The matrix pseudo-true parameters $A_1^*, \ldots, A_{p+1}^*$ are defined by the linear projection of Y_t onto $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p-1}$, i.e., we have $EL(Y_t|Y_{t-1}, \ldots, Y_{t-p-1}) = \sum_{j=1}^{p+1} A_j^* Y_{t-j}$.

[4]The condition in Assumption IR.1 can be imposed by a one-to-one transformation of the latent factor vector and possibly a reordering of the series, as long as matrix B is full column rank (Assumption M.1(iv)). Other identification restrictions could be used to fix the rotational invariance. After normalization, the number of free parameters in $\Theta_{p,q,K}$ is $K(n-K) + n^2p + K^2q + n(n+1)/2 + K(K+1)/2$.

To characterize the A_j^*, let $EL(f_t|Y_{t-1},\dots,Y_{t-p-1}) = \sum_{j=1}^{p+1} F_j Y_{t-j}$ be the linear projection of factor f_t onto the lags $Y_{t-1},\dots,Y_{t-p-1}$, where the F_j are $K\times n$ matrices of projection coefficients. Then, from (2.1) and (3.2), we have $A_{p+1}^* = B_0 F_{p+1}$ (see equation (A.4)). This implies that $\mathscr{R}(A_{p+1}^*) \subseteq \mathscr{R}(B_0)$, where $\mathscr{R}(\cdot)$ denotes the range (i.e., the column space) of a matrix. This provides a kind of lower bound for the range of matrix B_0. The latter is identified under the following assumption.

Assumption ID.1. The rank of matrix A_{p+1}^* is equal to K.

Under Assumption ID.1, we have $\mathscr{R}(B_0) = \mathscr{R}(A_{p+1}^*)$ because both linear spaces have dimension K. Hence, the column space of B_0 coincides with the eigenspace of matrix $A_{p+1}^*(A_{p+1}^*)'$ associated with the K nonzero eigenvalues. Under the normalization restriction IR.1, we get the next result.

PROPOSITION 1. *Under Assumptions M.1, IR.1, and ID.1, matrix B_0 is semiparametrically identifiable from the autocovariances of process $\{Y_t\}$ up to order $p+1$.*

Factor dynamics are irrelevant for the identification of matrix B_0 as long as the full-rank property in Assumption ID.1 holds. If the factor dynamics are given by the VAR(q) model in state equation (2.2), we have $A_{p+1}^* = B_0 \mathbf{\Phi}_0$ $Cov(\boldsymbol{f}_{t-1}, \tilde{Y}_{t-p-1}) V(\tilde{Y}_{t-p-1})^{-1}$, where $\tilde{Y}_{t-p-1} = Y_{t-p-1} - EL(Y_{t-p-1}|Y_{t-1},\dots,Y_{t-p})$. Hence, Assumption ID.1 is equivalent to the $n\times K$ matrix $Cov(\tilde{Y}_{t-p-1}, \boldsymbol{f}_{t-1})\mathbf{\Phi}_0'$ being full column rank. In particular, a necessary condition for Assumption ID.1 is the full-rank condition of $\mathbf{\Phi}_0$. In the case with $K=p=q=1$, in Appendix C.5 of the Supplementary Material, we show that

$$A_2^* = \phi_0 \sigma_{f,0}^2 B_0 B_0' (\lambda_0 I_n - \Lambda_0)(I_n - \phi_0 C_0')^{-1} \left(\Gamma_0(0) - \Gamma_0(1)'\Gamma_0(0)^{-1}\Gamma_0(1)\right)^{-1}, \tag{3.3}$$

for $\Lambda_0 = (I_n - \phi_0 C_0')^{-1}\Gamma_0(0)^{-1} C_0 \Gamma_0(0)(I_n - \phi_0 C_0')$ and $\lambda_0 = \phi_0 \Big[1 - \sigma_{f,0}^2 B_0' \Gamma_0(0)^{-1} (I_n - \phi_0 C_0)^{-1} B_0\Big]$, where $\Gamma_0(0) = V(Y_t)$, $\Gamma_0(1) = Cov(Y_t, Y_{t-1}) = C_0\Gamma_0(0) + \phi_0\sigma_{f,0}^2 B_0 B_0'(I_n - \phi_0 C_0')^{-1}$ and $\sigma_{f,0}^2 = V(f_t)$. Thus, Assumption ID.1 holds if $\phi_0 \neq 0$ and vector B_0 is *not* an eigenvector of matrix Λ_0 to the eigenvalue λ_0. We have not been able to find an explicit characterization of matrix A_{p+1}^* in terms of the state-space parameters in the general case.

3.2. Semiparametric Identification of Matrix C from Instrumental Variables Conditions

Let us define the identifiable matrices $\bar{B}_0 = B_0(B_0'B_0)^{-1}$ and $B_{0\perp} = [I_{n-K} : -B_{1,0}]'$. The columns of the $n\times(n-K)$ matrix $B_{0\perp}$ span the linear space $\mathscr{R}(B_0)^{\perp}$, i.e., the orthogonal complement of $\mathscr{R}(B_0)$. The condition $B_0' B_{0\perp} = 0$ identifies matrix $B_{0\perp}$ up to a nonsingular $(n-K)\times(n-K)$ transformation, and we normalize its upper block to I_{n-K} in analogy with Assumption IR.1.[5] Define the processes

[5]Our results are independent of the selected normalization restrictions for matrix $B_{0\perp}$.

$$\eta_t := B'_{0\perp}(Y_t - \boldsymbol{C}_0\boldsymbol{Y}_{t-1}) = B'_{0\perp}u_t \text{ and } \xi_t := \bar{B}'_0(Y_t - \boldsymbol{C}_0\boldsymbol{Y}_{t-1}) = f_t + \bar{B}'_0 u_t. \quad \textbf{(3.4)}$$

Vector $(\eta'_t, \xi'_t)'$ is a one-to-one linear transformation of $Y_t - \boldsymbol{C}_0\boldsymbol{Y}_{t-1}$, such that the first $n-K$ components η_t are a White Noise process and the last K components ξ_t are a serially persistent process corresponding to a noisy measurement of the factor. Since $B'_{0\perp}Y_t = \boldsymbol{\Delta}_0\boldsymbol{Y}_{t-1} + \eta_t$ with $\boldsymbol{\Delta}_0 := B'_{0\perp}\boldsymbol{C}_0$, and η_t is uncorrelated with the lagged values $\boldsymbol{Y}_{t-1}$, we identify matrix $\boldsymbol{\Delta}_0$ by the population regression of $B'_{0\perp}Y_t$ onto $\boldsymbol{Y}_{t-1}$, i.e.,

$$\boldsymbol{\Delta}_0 = B'_{0\perp}\boldsymbol{E}(Y_t\boldsymbol{Y}'_{t-1})\boldsymbol{E}(\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1})^{-1}. \quad \textbf{(3.5)}$$

Hence, $\eta_t = B'_{0\perp}Y_t - \boldsymbol{\Delta}_0\boldsymbol{Y}_{t-1}$ is a function of observed variables and identifiable parameters.

Let us write the measurement equation as $Y_t = \boldsymbol{C}_0\boldsymbol{Y}_{t-1} + \varepsilon_t$ where $\varepsilon_t = B_0 f_t + u_t$. From Assumption M.1(i) and (ii), for any integer $j \neq 0$, variable η_{t-j} is uncorrelated with the error term ε_t:

$$\boldsymbol{E}(\varepsilon_t\eta'_{t-j}) = 0, \qquad \forall j \neq 0. \quad \textbf{(3.6)}$$

These orthogonality conditions are equivalent to the White Noise property of η_t and its noncorrelation with all nonzero leads and lags of ξ_t, i.e., $\boldsymbol{E}(\eta_t\eta'_{t-j}) = 0$ and $\boldsymbol{E}(\xi_t\eta'_{t-j}) = 0$, for all $j \neq 0$. We use variables η_{t-j} as instruments to identify matrix $\boldsymbol{C}$. Indeed, equation (3.6) yields the IV condition

$$Q_{Y\boldsymbol{Z}} = \boldsymbol{C}Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}, \quad \textbf{(3.7)}$$

for $\boldsymbol{C} = \boldsymbol{C}_0$, where $Q_{Y\boldsymbol{Z}} := \boldsymbol{E}\left[Y_t\boldsymbol{Z}'_t\right]$ and $Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} := \boldsymbol{E}\left[\boldsymbol{Y}_{t-1}\boldsymbol{Z}'_t\right]$ and the instrument vector is $\boldsymbol{Z}_t = [\eta'_{t-1} : \cdots : \eta'_{t-M}]'$, for $M \geq 1$.[6] Equation (3.7) has a unique solution for $\boldsymbol{C}$ if, and only if, matrix $Q'_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}$ has full column rank. In Appendix A.4, we write matrix $Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}$ in terms of the state-space parameters B_0, $\boldsymbol{C}_0$, $\Sigma_{u,0}$, which allows us to get an identification condition in terms of the latter parameters.

Assumption ID.2. Let $M \geq p+1$ and define the $n \times n$ matrices $G_{m,i} = \sum_j C_{i+j-1}\mathfrak{C}_{m-j}$, for $i = 1,\dots,p$ and $m = 1,2,\dots,M-p$, where the sum extends over all positive integers j, with $C_j = 0$, for $j > p$, and $\mathfrak{C}_j$, for $j \geq 0$, is defined by $C(z)^{-1} = \sum_{j=0}^{\infty}\mathfrak{C}_j z^j$ on disk $|z| \leq 1$, and $\mathfrak{C}_j = 0$, for $j < 0$, with parameters evaluated at their true values. The $(M-p)(n-K) \times pK$ matrix with blocks $\left[B'_{0,\perp}\Sigma_{u,0}G'_{m,i}\Sigma_{u,0}^{-1}B_0\right]_{m=1,\dots,M-p,\ i=1,\dots,p}$ has full column rank.

The necessary order condition is $(M-p)(n-K) \geq pK$, i.e., $M(n-K) \geq pn$, which requires at least as many instruments as contagion parameters to estimate for each series. We have the next proposition.

PROPOSITION 2. *(a) Matrix $Q'_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}$ has full column rank if, and only if, Assumption ID.2 holds. (b) Under Assumptions M.1, IR.1, ID.1, and ID.2, matrix*

[6] We can also use leads η_{t+j} with $j \geq 1$ as instruments.

$\boldsymbol{C}_0 = [C_{1,0} : \cdots : C_{p,0}]$ *is semiparametrically identifiable from the autocovariances of process* $\{Y_t\}$ *up to order* $M+p$.

We can elaborate on Assumption ID.2 to provide interpretations for this identification condition. First, the sum defining the matrices $G_{m,i}$ extends over a finite number of terms, $G_{1,i} = C_{i,0}$, for $i = 1,\dots,p$, and $G_{m,1} = \mathfrak{C}_m$, for $m \geq 1$. Second, in the special case $p = 1$, we have $G_m \equiv G_{m,1} = C_{1,0}^m$, and Assumption ID.2 is equivalent to the following property: if $y \in \mathbb{R}^n$ is such that $(\Sigma_{u,0} C'_{1,0} \Sigma_{u,0}^{-1})^m y \in \mathscr{R}(B_0)$ for all integers $m = 0, 1, \dots, M-1$, then $y = 0$. In other words, the property of a nonnull vector to be in the column space of matrix B_0 cannot be maintained under $(M-1)$-fold application of matrix $\Sigma_{u,0} C'_{1,0} \Sigma_{u,0}^{-1}$. In particular, a necessary condition is that the column space of matrix B_0 is *not* an eigenspace of matrix $\Sigma_{u,0} C'_{1,0} \Sigma_{u,0}^{-1}$. Furthermore, if $p = K = 1$, Assumption ID.2 is equivalent to B_0 not being an eigenvector of $\Sigma_{u,0} C'_{1,0} \Sigma_{u,0}^{-1}$.

The instruments η_{t-j} are internal to the model. They differ from external instruments used to identify dynamic causal effects in structural VAR models (Stock and Watson, 2018). If some observed variables are available, which are orthogonal to both the latent factor process $\{f_t\}$ and the idiosyncratic shocks $\{u_t\}$, then they can be used to construct additional orthogonality restrictions. These variables, however, are not necessary for identification under our assumptions.

Previous literature considered semiparametric identification in state-space models. Shiu and Hu (2013), Schennach (2014), Gallant, Giacomini, and Ragusa (2017), and Gagliardini and Gouriéroux (2019) study semiparametric identification in models with unobserved components that are more general than ours. In particular, Gagliardini and Gouriéroux (2019) consider nonlinear panel models with unobserved dynamic components, such that the measurement equation is exponentially affine in the latent factor and lagged endogenous variables. The specification in this paper (when $p = 1$, process u_t is i.i.d. and factor (f_t) is Markovian) is a special case of those in Gagliardini and Gouriéroux (2019). Their identification strategy relies on well-chosen moment restrictions that are implied by the conditional moment generating function after "partialling out" the infinite-dimensional parameters corresponding to the unknown density of innovation u_t and transition density of the latent factor. Their approach is general and covers a broad class of nonlinear state-space models. In the general framework considered by Gagliardini and Gouriéroux (2019), the identification strategies rely on high-level assumptions that are more difficult to check from primitive conditions on the DGP parameters, and are less simple to implement for estimation purposes compared to the identification approach pursued in this paper.

3.3. Parametric Identification of Matrix Φ in Factor Dynamics and Other Parameters

For the identification of the stacked autoregressive matrix $\boldsymbol{\Phi}$ in the VAR(q) factor dynamics, we use process ξ_t defined in equation (3.4). From Propositions 2 and 3,

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

process ξ_t is a function of observable data and identifiable parameters. By plugging $f_t = \xi_t - \bar{B}_0' u_t$ in the state equation (2.2), we get $\xi_t = \sum_{j=1}^q \Phi_j \xi_{t-j} + e_t$, where $e_t = \bar{B}_0' u_t - \sum_{j=1}^q \Phi_j \bar{B}_0' u_{t-j} + v_t$. Since $e_t \sim MA(q)$, process ξ_t follows a VARMA(q,q) dynamics. Define the instrument vector $\boldsymbol{W}_t = [\xi_{t-q-1}', \xi_{t-q-2}', \dots, \xi_{t-q-L}']'$, for integer $L \geq q$.[7] Then, we get the IV condition

$$Q_{\xi \boldsymbol{W}} = \boldsymbol{\Phi} Q_{\boldsymbol{\xi}_{-1} \boldsymbol{W}}, \tag{3.8}$$

where $Q_{\xi \boldsymbol{W}} = \boldsymbol{E}[\xi_t \boldsymbol{W}_t']$ and $Q_{\boldsymbol{\xi}_{-1} \boldsymbol{W}} = \boldsymbol{E}[\boldsymbol{\xi}_{t-1} \boldsymbol{W}_t']$ with $\boldsymbol{\xi}_{t-1} := [\xi_{t-1}', \dots, \xi_{t-q}']'$. Equation (3.8) has the unique solution $\boldsymbol{\Phi} = \boldsymbol{\Phi}_0$ if, and only if, the $LK \times qK$ matrix $Q_{\boldsymbol{\xi}_{-1} \boldsymbol{W}}'$ has full column rank.

Assumption ID.3. The matrix $\Phi_{q,0}$ is nonsingular.

PROPOSITION 3. *Let $L \geq q$. (a) Matrix $Q_{\boldsymbol{\xi}_{-1} \boldsymbol{W}}'$ has full column rank if, and only if, matrix $\Phi_{q,0}$ is nonsingular. (b) Under Assumptions M.1, M.2, IR.1, and ID.1–ID.3, matrix $\boldsymbol{\Phi}_0$ is identifiable from the autocovariances of process $\{Y_t\}$ up to order* $\max\{p+M, q+L\}$.

Our identification strategy for $\boldsymbol{\Phi}$ builds on the literature using IV to obtain orthogonality restrictions in models with lagged endogenous variables and serially dependent errors (see, e.g., Hansen and Singleton, 1991). Other identification approaches are proposed in the dynamic factor model literature for large n, which do not require Φ_q to be full rank (see Forni et al., 2009). In our case, with n finite, we cannot follow the same identification strategy as Forni et al. (2009).

The condition in Assumption ID.3 is admittedly quite restrictive and is not met by some simple DGP. For instance, let $K = q = 2$ and suppose that the two latent factors are mutually uncorrelated AR(1) and AR(2) processes: $f_{1t} = \phi_{1,11} f_{1t-1} + v_{1t}$ and $f_{2t} = \phi_{1,22} f_{2t-1} + \phi_{2,22} f_{2t-2} + v_{2t}$. Then, we have $\Phi_{2,0} = \begin{bmatrix} 0 & 0 \\ 0 & \phi_{2,22}^0 \end{bmatrix}$, and Assumption ID.3 does not hold.[8] To overcome the restrictive condition in Assumption ID.3, we could consider an approach that combines our identification strategy for B and $\boldsymbol{C}$ together with QMLE. In fact, once the values of process ξ_t are identified by Assumptions ID.1 and ID.2, we can identify matrix $\boldsymbol{\Phi}$ from the linear state-space models

$$\xi_t = f_t + \bar{B}_0' u_t, \qquad f_t = \boldsymbol{\Phi} \boldsymbol{f}_{t-1} + v_t. \tag{3.9}$$

[7] We could use lags of the observable process as instruments. However, variables Y_{t-q-j}, $j \geq 1$, are one-to-one transformations of variables $\xi_{t-q-j}, \eta_{t-q-j}, j \geq 1$, and the latter variables $\eta_{t-q-j}, j \geq 1$, are uncorrelated with regressor $\boldsymbol{\xi}_{t-1}$ and hence irrelevant for instrumentation purposes.

[8] In fact, there is a linear combination of the elements in $(\xi_{t-1}', \xi_{t-2}')'$, namely $\xi_{1t-1} - \phi_{1,11}^0 \xi_{1t-2}$ that is a linear combination of $v_{1,t-1}$, u_{t-1}, u_{t-2}, and hence is uncorrelated with the lags $\xi_{t-3}, \xi_{t-4}, \dots$. Thus, matrix $Q_{\boldsymbol{\xi}_{-1} \boldsymbol{W}}'$ has not full column rank. Note that, if $\phi_{1,11}^0 \neq 0$, matrix $\boldsymbol{\Phi}_0 = [\Phi_{1,0} : \Phi_{2,0}]$ is full rank, a necessary condition for Assumption ID.1.

We briefly consider the implied estimation approach in Section 8, but leave a detailed analysis for future research. We conjecture that it provides a consistent and asymptotically normal estimator of $\mathbf{\Phi}$, and we can apply a rank testing methodology similar to that deployed in Section 5 to test the full-rank condition for $\Phi_{q,0}$. Furthermore, we stress that, for $q = 1$, our model with singular matrix Φ_0 of rank $r < K$, say, corresponds to a model with r dynamic factors and $K - r$ static factors. The identification in this case is discussed in Appendix E.4 of the Supplementary Material. Finally, we remark that, even for a DGP for which $\Phi_{q,0}$ has full rank, the use of lagged values of ξ_t prior to date $t - q$ might induce a weak instrument problem when (some linear transformations of) the factor process is weakly serially correlated.

In Appendix E.2 of the Supplementary Material, we show that the variance–covariance matrices Σ_u and Σ_v are identified under Assumptions ID.1–ID.3. This result, together with Propositions 1–3, yields the next theorem.

THEOREM 1. *Under Assumptions M.1, M.2, IR.1, and ID.1–ID.3, the parameter vector $\theta \in \Theta_{p,q,K}$ is globally identifiable from a finite number of autocovariances of the observable process $\{Y_t\}$.*

The arguments in Sections 3.1–3.3 show that mapping τ in equation (3.1) corresponds to computing the eigenvectors of matrix $A^*_{p+1}(A^*_{p+1})'$, and solving the linear IV systems (3.7) and (3.8).

3.4. An Example

As an illustration of the identification assumptions presented in the previous subsections, we consider next a model with $n = 2$ observables, a single latent factor ($K = 1$), and VAR orders $p = q = 1$. The DGP is such that

$$C_{1,0} = \begin{pmatrix} c_{11,0} & 0 \\ 0 & c_{22,0} \end{pmatrix}, \quad B_0 = \begin{pmatrix} b_{1,0} \\ 1 \end{pmatrix}, \quad \Sigma_{u,0} = \begin{pmatrix} \sigma^2_{u,1} & 0 \\ 0 & \sigma^2_{u,2} \end{pmatrix}, \qquad \textbf{(3.10)}$$

and scalar ϕ_0 is the autoregressive coefficient of the factor dynamics. In the Supplementary Material, we show the next result.

PROPOSITION 4. *Let $n = 2$ and $K = p = q = 1$, and let the DGP parameters be as in (3.10). Then, Assumptions ID.1–ID.3 hold if, and only if, $\phi_0 \neq 0$, $b_{1,0} \neq 0$, and $c_{11,0} \neq c_{22,0}$.*

From Sections 3.1 and 3.3, condition $\phi_0 \neq 0$ is necessary for Assumptions ID.1 and ID.3 to hold. Condition $b_{1,0} \neq 0$ implies that the latent factor impacts variable $Y_{1,t}$ as well. Moreover, from Section 3.2, condition $c_{11,0} \neq c_{22,0}$ is necessary for Assumption ID.2 to hold. Indeed, if $c_{11,0} = c_{22,0}$, then matrix $\Sigma_{u,0} C'_{1,0} \Sigma^{-1}_{u,0} = c_{11,0} I_2$ admits any nonzero vector as an eigenvector. Proposition 4 shows that the aforementioned conditions are both necessary and sufficient for identification.

3.5. Identification from Other Model Representations

It is instructive to contrast Propositions 1–4 and Theorem 1 with other identification results obtained in the literature on linear state-space models.

3.5.1. *VARMA Representation.* The state-space model in (2.1) and (2.2) admits a vector autoregressive moving average (VARMA) representation (Akaike, 1974). Darolles et al. (2014) show that, when $p = q = 1$, the VARMA representation has two lags in the autoregressive part and one lag in the moving average part. For generic orders $p,q \geq 1$ in Appendix A.2, we prove the causal VARMA($p+q$,q) representation

$$A(L)Y_t = \Psi(L)w_t, \tag{3.11}$$

where the AR matrix polynomial $A(L) = I_n - \sum_{j=1}^{p+q} A_j L^j$ has order $p+q$ and coefficients A_j given in Lemma 1, the MA polynomial $\Psi(L) = I_n + \sum_{j=1}^{q} \Psi_j L^j$ has order q, and $w_t \sim WN(0, \Sigma_w)$. It holds $\Psi_j = B\nu_j'$ for some $n \times K$ matrices ν_j. Matrices Σ_w and ν_j are defined in terms of the state-space parameters by equations (A.2) and (A.3), and are not in closed form.

Since $A_{p+q} = -B\Phi_q \bar{B}' C_p$ and $\Psi_q = B\nu_q'$, where $\bar{B} = B(B'B)^{-1}$, the VARMA($p+q$,q) representation in (3.11) is such that Rank$[A_{p+q} : \Psi_q] \leq K < n$. Thus, the sufficient and necessary rank condition for identification of parameters A_j, Ψ_j, and Σ_w (see Hannan, 1969, 1971, 1975) is not met. The parameter vector θ of the state-space model in (2.1) and (2.2) is not identifiable from the VARMA representation (3.11) unless all structural restrictions are imposed.[9] The complexity of these structural restrictions motivates us to consider a different approach to identification, estimation, and model selection presented in this paper.

3.5.2. *Identification from ABCD Representation.* In the statistical and econometric literature, the identification of linear state-space systems is often analyzed in the ABCD state-space representation. In this subsection, let the number of autoregressive matrices in the measurement and state equations be $p = q = 1$ to ease exposition. In this case, the state-space system in (2.1) and (2.2) admits an ABC(D) representation:

$$X_t = \mathscr{A} X_{t-1} + \mathscr{B} W_t, \qquad Y_t = \mathscr{C} X_t, \tag{3.12}$$

where the state vector is $X_t = (Y_t', f_{t+1}')'$, the innovation vector W_t follows a $WN(0, I_{n+K})$ process, and

[9] In Appendix E.1 of the Supplementary Material, we show the unidentifiability of the parameters of the restricted VARMA($p+q$,q) representation, in which the AR matrices A_j are replaced with their expressions in terms of B, C_j, and Φ_j, and the MA matrices Ψ_j are written as $B\nu_j'$ without imposing the structural constraints from (A.2) and (A.3) on matrices ν_j and Σ_w.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

$$\mathscr{A} = \begin{bmatrix} C_1 & B \\ 0 & \Phi_1 \end{bmatrix}, \quad \mathscr{B} = \begin{bmatrix} \Sigma_u^{1/2} & 0 \\ 0 & \Sigma_v^{1/2} \end{bmatrix}, \quad \mathscr{C} = [I_n : 0].$$

The ABCD representation yields a map between matrices $\mathscr{A}, \mathscr{B}$, and $\mathscr{C}$ and the parameter vector $\theta \in \Theta_{1,1,K}$ of the state-space representation, for given K. The results in Glover and Willems (1974) are among the most important ones to find conditions of global identification for the parameters of linear state-space systems in ABCD form. They study state-space systems in minimal representation, and consider global (and local) identification from the transfer function $G(s) = \mathscr{C}(I_{n+K} - \mathscr{A}s)^{-1}\mathscr{B}s$, where s is the complex argument of function $G(\cdot)$ such that $Y_t = G(L)W_t$, and from the spectral function $G(s)G(\bar{s})'$, where $\bar{s}$ denotes complex conjugate of s. By applying the results in Glover and Willems (1974) and the references therein, in Appendix D of the Supplementary Material, we show that the representation (3.12) is minimal and we prove the following proposition.

PROPOSITION 5. *Assume the Identification Restriction IR.1. (a) The state-space parameter vector $\theta \in \Theta_{1,1,K}$ is globally identifiable from the transfer function $G(s)$, for given K. (b) The parameter vector is globally identifiable from the spectral function $G(s)G(\bar{s})'$ if, and only if, the equation $\Sigma_{u,0}C'_{1,0}\Sigma_{u,0}^{-1}(B_0Q_{22}) = (B_0Q_{22})\Phi'_{1,0}$, for a $K \times K$ symmetric matrix Q_{22}, implies $Q_{22} = 0$.*

Global identification from the transfer function is a necessary condition for consistency of the QMLE (Hannan and Deistler, 1988). From Proposition 5(a), this condition does not require Assumptions ID.1–ID.3, once K is identified. Furthermore, global identification from the spectral function, and identification from the autocovariance function as in Theorem 1, are intimately connected. Indeed, there is a one-to-one map between the spectral density function and the autocovariance function of the process.[10] From Proposition 5(b), we deduce that: if the only linear subspace of the range $\mathscr{R}(B_0)$ that is mapped by $\Sigma_{u,0}C'_{1,0}\Sigma_{u,0}^{-1}$ into $\mathscr{R}(B_0)$ is the trivial subspace, then the parameter set is globally identified from the spectral function. As we argue in Section 3.2, the latter sufficient condition is implied by Assumption ID.2 when $p = 1$. In the example of Section 3.4, the parameters are identifiable from the spectral density function under the same conditions as those listed in Proposition 4.

We stress that Theorem 1 expands substantially the identification results in the state-space literature along several directions. First, our identification strategy presented in Sections 3.1–3.3 is constructive, in that it leads to a multistep procedure to estimate separately and sequentially parameters B, $\boldsymbol{C}$, and $\boldsymbol{\Phi}$ in closed form. This estimator is computationally more convenient and faster than the QMLE based, e.g., on Kalman filter and numerical optimization in a high-dimensional parameter space (see Section 6). Second, we identify parameters B

[10]Note, however, that Theorem 1 relies on a finite number of autocovariances for identification, while the spectral density function involves autocovariances at all lags.

and $\boldsymbol{C}$ semiparametrically, i.e., without specifying a parametric dynamics for the latent factor process. Third, Theorem 1 applies for the general framework with $p,q \geq 1$. Moreover, the arguments in Proposition 1 yield the identification of K as well.

Finally, other identification results for structural ABCD models are discussed in Komunjer and Ng (2011). These authors find rank conditions for which the parameters of a structural ABCD system are locally identified when the number of the shocks is less or greater than the number of state variables.

4. ESTIMATION AND SPECIFICATION TESTING

In this section, we assume the number of latent factors K and the VAR orders p, q known, and consider the estimation of model parameter vector $\theta_0 \in \Theta_{p,q,K}$ and testing of model specification.

4.1. A Three-Step Estimation Procedure

We estimate the parameters of the state-space model on a sample Y_t, for $t = 1,\ldots,T$, of the observable process by using the empirical counterpart of the identification strategy presented in Section 3. The estimation procedure for the matrix parameters B, $\boldsymbol{C}$, and $\boldsymbol{\Phi}$ consists of three steps and is in closed form up to an eigendecomposition (the estimators of Σ_u and Σ_v are given in Appendix E.3 of the Supplementary Material).

(i) Estimation of the factor loadings matrix B. We first estimate the matrix pseudo-parameters $\hat{A}_j^*$, for $j = 1,\ldots,p+1$, by ordinary least squares (OLS) applied to the VAR$(p+1)$ pseudo-model (3.2). Let $\hat{\boldsymbol{U}} = [\hat{U}_1 : \cdots : \hat{U}_K]$ be the matrix having for columns the standardized eigenvectors of $\hat{A}_{p+1}^*(\hat{A}_{p+1}^*)'$ associated with the K largest eigenvalues. We partition this matrix as $\hat{\boldsymbol{U}} = [\hat{\boldsymbol{U}}_1' : \hat{\boldsymbol{U}}_2']'$, where $\hat{\boldsymbol{U}}_2$ is the lower $K \times K$ block. Then, we estimate B under the identification restriction in Assumption IR.1 as

$$\hat{B} = [\hat{B}_1' : I_K]' = \hat{\boldsymbol{U}}\hat{\boldsymbol{U}}_2^{-1}. \tag{4.1}$$

(ii) Estimation of the stacked contagion matrix $\boldsymbol{C}$. Let $\hat{B}_\perp = [I_{n-K} : -\hat{B}_1]'$ be the estimator of $B_{0\perp}$ obtained using $\hat{B}_1$ from the first step under the normalization restriction introduced in Section 3.2. Building on equation (3.5), we estimate $\boldsymbol{\Delta}_0$ with OLS by $\hat{\boldsymbol{\Delta}} = \hat{B}_\perp'\left(\sum_{t=1}^T Y_t \boldsymbol{Y}_{t-1}'\right)\left(\sum_{t=1}^T \boldsymbol{Y}_{t-1}\boldsymbol{Y}_{t-1}'\right)^{-1}$. Given $\hat{B}_\perp$ and $\hat{\boldsymbol{\Delta}}$, we estimate matrix $\hat{\boldsymbol{C}}$ by a multivariate IV estimator built on the sample analog of the moment condition (3.7). The instrument vector is $\hat{\boldsymbol{Z}}_t = (\hat{\eta}_{t-1}', \ldots, \hat{\eta}_{t-M}')'$ with $\hat{\eta}_t = \hat{B}_\perp' Y_t - \hat{\boldsymbol{\Delta}}\boldsymbol{Y}_{t-1}$. By vectorizing the moment equation, we get the estimator

$$\text{vec}(\hat{\boldsymbol{C}}) = \left[(\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} \otimes I_n)\hat{\Omega}_c(\hat{Q}_{\boldsymbol{Z}\boldsymbol{Y}_{-1}} \otimes I_n)\right]^{-1}(\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} \otimes I_n)\hat{\Omega}_c \text{vec}(\hat{Q}_{Y\boldsymbol{Z}}), \tag{4.2}$$

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

where $\hat{Q}_{YZ} = \frac{1}{T}\sum_{t=1}^{T} Y_t \hat{\boldsymbol{Z}}_t'$, $\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Y}_{t-1}\hat{\boldsymbol{Z}}_t' = (\hat{Q}_{\boldsymbol{Z}\boldsymbol{Y}_{-1}})'$, and $\hat{\Omega}_c$ is an $Mn(n-K) \times Mn(n-K)$ positive definite weighting matrix. That matrix introduces a weighting of the orthogonality restrictions across the M instruments and the n series in the system. Since the IV criterion involves estimated parameters $\hat{B}_\perp$ and $\hat{\boldsymbol{\Delta}}$ in the instruments, the optimal weighting matrix differs from that of a standard generalized method of moments (GMM) estimator (see the next subsection). With the identity weighting matrix, the estimator becomes $\hat{\boldsymbol{C}} = \hat{Q}_{YZ}\hat{Q}_{\boldsymbol{Z}\boldsymbol{Y}_{-1}}\left(\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}\hat{Q}_{\boldsymbol{Z}\boldsymbol{Y}_{-1}}\right)^{-1}$.

(iii) Estimation of the stacked autoregressive matrix $\boldsymbol{\Phi}$ *in the factor dynamics.* We find a multivariate IV estimator based on the moment condition (3.8), which yields

$$\operatorname{vec}(\hat{\boldsymbol{\Phi}}) = \left[(\hat{Q}_{\boldsymbol{\xi}_{-1}\boldsymbol{W}} \otimes I_K)\hat{\Omega}_\phi(\hat{Q}_{\boldsymbol{W}\boldsymbol{\xi}_{-1}} \otimes I_K)\right]^{-1}(\hat{Q}_{\boldsymbol{\xi}_{-1}\boldsymbol{W}} \otimes I_K)\hat{\Omega}_\phi \operatorname{vec}(\hat{Q}_{\xi\boldsymbol{W}}), \qquad \textbf{(4.3)}$$

where $\hat{Q}_{\xi\boldsymbol{W}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t\hat{\boldsymbol{W}}_t'$ and $\hat{Q}_{\boldsymbol{\xi}_{-1}\boldsymbol{W}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\xi}}_{t-1}\hat{\boldsymbol{W}}_t' = (\hat{Q}_{\boldsymbol{W}\boldsymbol{\xi}_{-1}})'$, with $\hat{\boldsymbol{\xi}}_{t-1} = (\hat{\xi}_{t-1}', \ldots, \hat{\xi}_{t-q}')'$, $\hat{\boldsymbol{W}}_t = (\hat{\xi}_{t-q-1}', \ldots, \hat{\xi}_{t-q-L}')'$, and $\hat{\xi}_t := (\hat{B}'\hat{B})^{-1}\hat{B}'(Y_t - \hat{\boldsymbol{C}}\boldsymbol{Y}_{t-1})$, and $\hat{\Omega}_\phi$ is a $K^2L \times K^2L$ positive definite weighting matrix.

Estimators $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\Phi}}$ involve two tuning parameters, which are the numbers of lagged instruments M and L, subject to the order conditions $M \geq pn/(n-K)$ and $L \geq q$. Too large values of M and L for given sample size T may result in estimators with poor finite-sample properties. In our Monte Carlo experiments and empirical analysis, we implement ad hoc choices of M, L and check that the estimates are stable in a range around the selected values of the tuning parameters.

4.2. Large Sample Properties

We derive the asymptotic properties of the estimators when the sample size $T \to \infty$.

Assumption LS.1. The error process $z_t = (u_t', v_t')'$ is either (i) a conditionally homoskedastic martingale difference sequence (m.d.s.), i.e., $\boldsymbol{E}[z_t|\underline{z_{t-1}}] = 0$ and $V(z_t|\underline{z_{t-1}}) = \begin{pmatrix} \Sigma_u & 0 \\ 0 & \Sigma_v \end{pmatrix} =: \Sigma_z$, or (ii) an i.i.d. sequence with mean 0 and variance Σ_z, and the density functions f_u and f_v of u_t and v_t are positive on $\mathbb{R}^n$ and $\mathbb{R}^K$, respectively. Moreover, (iii) $\boldsymbol{E}(|u_t|^{2\beta}) < \infty$, $\boldsymbol{E}(|v_t|^{2\beta}) < \infty$ for $\beta > 2$, and (iv) the K nonzero eigenvalues of matrix $A_{p+1}^*(A_{p+1}^*)'$ are distinct. Furthermore, (v) the weighting matrices $\hat{\Omega}_c$ and $\hat{\Omega}_\phi$ converge a.s. to positive definite matrices Ω_c and Ω_ϕ, respectively.

Let $b = \operatorname{vec}(B_1)$, $c = vec(\boldsymbol{C})$ and $\phi = \operatorname{vec}(\boldsymbol{\Phi})$ denote the vectorized forms of the matrix parameters, and let $\hat{b}$, $\hat{c}$, and $\hat{\phi}$ be the corresponding estimators from Section 4.1.[11]

[11] We focus on estimators $\hat{b}$, $\hat{c}$, and $\hat{\phi}$ for conciseness. The large sample properties of estimators $\hat{\Sigma}_u$ and $\hat{\Sigma}_v$ defined in Appendix E.3 of the Supplementary Material can be established along similar lines.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

THEOREM 2. *(a) Under Assumptions M.1, M.2, IR.1, ID.1–ID.3, and LS.1(i) and (iii)–(v), the estimators* $\hat{b}$, $\hat{c}$, *and* $\hat{\phi}$ *are strongly consistent:* $\hat{b} \overset{a.s.}{\to} b_0$, $\hat{c} \overset{a.s.}{\to} c_0$, *and* $\hat{\phi} \overset{a.s.}{\to} \phi_0$, *as* $T \to \infty$. *(b) We have the asymptotic expansion*

$$\sqrt{T}\begin{pmatrix} \hat{b}-b_0 \\ \hat{c}-c_0 \\ \hat{\phi}-\phi_0 \end{pmatrix} = \begin{bmatrix} S_{b1} & 0 & 0 & 0 \\ S_{c1} & S_{c2} & S_{c3} & 0 \\ S_{\phi 1} & S_{\phi 2} & S_{\phi 3} & S_{\phi 4} \end{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \tilde{Y}_{t-p-1} \otimes u_t^* \\ \boldsymbol{Y}_{t-1} \otimes \eta_t \\ \boldsymbol{Z}_t \otimes \varepsilon_t \\ \boldsymbol{W}_t \otimes e_t \end{pmatrix}$$

$$+ o_p(1) =: S\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t + o_p(1), \tag{4.4}$$

where $\tilde{Y}_{t-p-1} = Y_{t-p-1} - EL(Y_{t-p-1}|Y_{t-1},\ldots,Y_{t-p})$, $u_t^* = Y_t - EL(Y_t|Y_{t-1},\ldots,Y_{t-p-1})$, $\varepsilon_t = Y_t - \boldsymbol{C}_0\boldsymbol{Y}_{t-1}$, $e_t = \xi_t - \boldsymbol{\Phi}_0\boldsymbol{\xi}_{t-1}$, $\boldsymbol{Z}_t = (\eta_{t-1}',\ldots,\eta_{t-M}')'$, $\boldsymbol{W}_t = (\xi_{t-q-1}',\ldots,\xi_{t-q-L}')'$, *and* η_t *and* ξ_t *are defined in (3.4). The matrices* S_{ak} *with* $a = b,c,\phi$ *and* $k = 1,2,3,4$ *are continuous functions of the autocovariances of process* $\{Y_t\}$ *and parameters* b_0, c_0, *and* ϕ_0, *and their expressions are given in (B.7) in Appendix B. (c) If Assumption LS.1(ii) holds instead of LS.1(i), we have*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t \overset{d}{\to} N(0,V_\psi), \tag{4.5}$$

as $T \to \infty$, *where* $V_\psi = \sum_{j=-\infty}^{\infty} Cov(\psi_t,\psi_{t-j})$. *In particular, the estimators* $\hat{b}$, $\hat{c}$, *and* $\hat{\phi}$ *are asymptotically normal* $\sqrt{T}\left((\hat{b}-b_0)',(\hat{c}-c_0)',(\hat{\phi}-\phi_0)'\right)' \overset{d}{\to} N(0,\Sigma_0)$, *as* $T \to \infty$, *with* $\Sigma_0 = SV_\psi S'$.

We establish the strong consistency of the estimators in Theorem 2(a) by using the fact that $\hat{b}$, $\hat{c}$, and $\hat{\phi}$ are continuous functions of the sample autocovariances of process $\{Y_t\}$, and the latter are strongly consistent under the conditions of Proposition 4.1.1 of Hannan and Deistler (1988). To check those conditions, we use the m.d.s. property of the errors in Assumption LS.1(i). The proof of strong consistency also uses an adapted version of a perturbation theory result in Izenman (1975) (Proposition 6 in Appendix B.1) to get the asymptotic expansion of the eigenvectors of matrix $\hat{A}_{p+1}^*(\hat{A}_{p+1}^*)'$ needed to estimate $\hat{b}$. To apply this result, we need the condition of distinct nonzero eigenvalues for matrix $A_{p+1}^*(A_{p+1}^*)'$ in Assumption LS.1(iv). Theorem 2(b) states that the estimators $\hat{b}$, $\hat{c}$, and $\hat{\phi}$ are asymptotically a linear transformation of the scaled sample average of process $\{\psi_t\}$. This process is the orthogonality vector in the representation of $(\hat{b}',\hat{c}',\hat{\phi}')'$ as a sequential GMM estimator (Newey and McFadden, 1994). Specifically, (i) $\tilde{Y}_{t-p-1} \otimes u_t^*$ is the orthogonality vector for OLS estimation of A_{p+1}^* in the VAR($p+1$) pseudo-model (3.2), and similarly (ii) $\boldsymbol{Y}_{t-1} \otimes \eta_t$ for OLS estimation of $\boldsymbol{\Delta}_0$ in equation (3.5), (iii) $\boldsymbol{Z}_t \otimes \varepsilon_t$ for IV estimation of $\boldsymbol{C}_0$ with equation (3.7), and (iv) $\boldsymbol{W}_t \otimes e_t$ for IV estimation of $\boldsymbol{\Phi}_0$ with equation (3.8). In matrix S, the blocks S_{cj}, with $j = 1, 2$, and $S_{\phi j}$, with $j = 1, 2, 3$, account for the multistep nature of the estimation procedure.[12]

The asymptotic variance Σ_0 of the estimator depends on the asymptotic weighting matrices Ω_c and Ω_ϕ. In Appendix B.4, we show that the optimal weighting matrix for estimating $\boldsymbol{C}$, that is, the weighting matrix yielding the smallest asymptotic variance for $\sqrt{T}(\hat{c} - c_0)$, is given by $\Omega_c^* = (D_c V_\psi D_c')^{-1}$, where matrix D_c is defined in equation (B.9) and accounts for first-step estimation of B_0 and $\boldsymbol{\Delta}_0$. Similarly, we can derive the optimal weighting matrix for estimating $\boldsymbol{\Phi}$ once the weighting matrix for estimating $\boldsymbol{C}$ is set, which is given in (B.11).[13]

The components of vector ψ_t, except $\boldsymbol{Y}_{t-1} \otimes \eta_t$ under Assumption LS.1(ii), are correlated across time; thus, the long-run matrix V_ψ involves autocovariances at all lags and leads. We can estimate the asymptotic variance Σ_0 by using a heteroskedasticity and autocorrelation consistent (HAC) estimator for matrix V_ψ. Specifically, let us write $\psi_t = \psi_t(\gamma_0, \theta_0)$ and $S = S(\gamma_0, \theta_0)$, where $\theta_0 = (b_0', c_0', \phi_0')'$ and vector γ_0 consists of a finite number of (cross) autocovariances of the components of process $\{Y_t\}$. Then, an estimator of Σ_0 is $\hat{\Sigma} = \hat{S}\hat{V}_\psi \hat{S}'$, where $\hat{S} = S(\hat{\gamma}, \hat{\theta})$, vector $\hat{\gamma}$ is obtained by the sample autocovariance function of $\{Y_t\}$, and $\hat{V}_\psi = \sum_{j=-(T-1)}^{T-1} \kappa(j/m_T)\hat{\Gamma}_\psi(j)$, with $\hat{\Gamma}_\psi(j) = \frac{1}{T}\sum_{t=j+1}^{T} \hat{\psi}_t \hat{\psi}_{t-j}'$, and $\hat{\psi}_t = \psi_t(\hat{\gamma}, \hat{\theta})$. The kernel function $\kappa(\cdot)$ and the bandwidth m_T satisfy the next assumption.

Assumption LS.2. Kernel function $\kappa : \mathbb{R} \to \mathbb{R}$ is such that (i) $\kappa(0) = 1$, $\kappa(-x) = \kappa(x)$ for all $x \in \mathbb{R}$, (ii) κ is continuous at 0 and at almost all $x \in \mathbb{R}$, $\int_{-\infty}^{\infty} |\kappa(x)| dx < \infty$, and (iv) $\hat{\kappa}(\lambda) \geq 0$, for all $\lambda \in \mathbb{R}$, where $\hat{\kappa}(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \kappa(x) e^{-ix\lambda} dx$. (vi) The bandwidth m_T is a sequence of positive integers such that $m_T \to \infty$ and $m_T = o(T^{1/2})$.

THEOREM 3. *Under Assumptions M.1, M.2, IR.1, ID.1–ID.3, LS.1, and LS.2, the estimator $\hat{\Sigma}$ is positive definite almost surely and converges in probability to Σ_0.*

We prove Theorem 3 in Appendix B.5 by applying the results in De Jong and Davidson (2000), which require mild conditions on the bandwidth parameter m_T. Andrews and Monahan (1992) and Newey and West (1994) discuss methods for

[12]The asymptotic normality in (4.5) could possibly be proved under milder conditions than the i.i.d. property of the errors in Assumption LS.1(ii) using versions of the central limit theorem (CLT) for the sample autocovariances of linear processes as those in, e.g., Hannan and Heyde (1972) and Phillips and Solo (1992). However, those results are stated for univariate processes and their extensions to a multivariate setting seem cumbersome. This explains why instead we establish that $\{(Y_t', f_t')'\}$ is a geometrically strongly mixing process under Assumptions M.1, M.2, and LS.1(ii) using Theorem 1 in Mokkadem (1988), and use the CLT for mixing processes in Herrndorf (1984). Furthermore, we can show consistency and asymptotic normality of $\hat{b}$ and $\hat{c}$ in a semiparametric framework without assuming linear factor dynamics, if we take the strong mixing property for process $\{(Y_t', f_t')'\}$ as a high-level assumption instead of proving it from primitive conditions. For the sake of space, we do not detail this alternative formulation.

[13]The definition of optimal weighting matrices for joint estimation of $\boldsymbol{C}$ and $\boldsymbol{\Phi}$ is more involved, since the two matrix parameters are estimated sequentially without cross-weighting of the corresponding sets of moment restrictions.

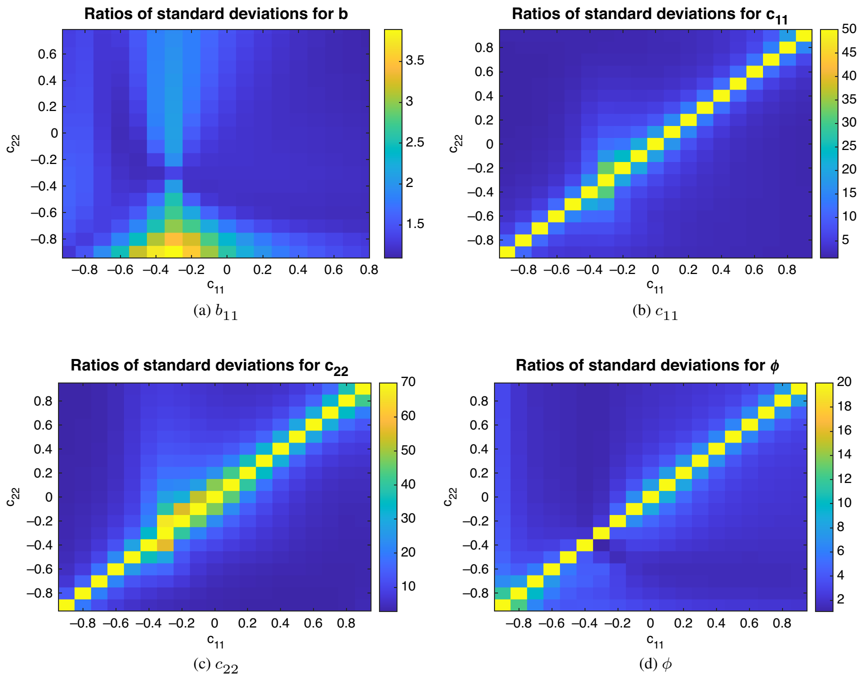

FIGURE 1. This figure plots the ratios between the asymptotic standard deviations of our estimators (4.1)–(4.3) computed with $M = L = 10$ and identity weighting matrix, and of the ML estimator of parameters b_{11}, c_{11}, c_{22}, and ϕ, for different values of DGP parameters c_{11} and c_{22}. The DGP is a model with $n = 2$ and $K = p = q = 1$ as in equation (3.10) with Gaussian errors. We fix $b_{1,0} = 1.5$, $\phi_0 = 0.9$, $\sigma^2_{v,0} = 1$, and $\Sigma_{u,0} = I_2$.

automatic selection of m_T. Estimators of $\boldsymbol{C}_0$ and $\boldsymbol{\Phi}_0$ based on optimal weighting matrices can be implemented with a two-step approach using HAC estimator $\hat{V}_\psi$ and first-step estimate of θ_0 built with identity weighting matrices.

As an application of Theorem 2, in Figure 1, we display the ratio of the asymptotic standard deviation of some components of the estimators $\hat{b}$, $\hat{c}$, and $\hat{\phi}$ to the asymptotic standard deviation of the corresponding ML estimators in a simple example through heat maps. The DGP is the same as in Section 3.4 (see equation (3.10)) with Gaussian innovations. We fix $b_{1,0} = 1.5$, $\phi_0 = 0.9$, $\Sigma_{u,0} = I_2$, and $\sigma^2_{v,0} = 1$, and display the asymptotic standard deviation ratio as a function of the parameters $c_{11,0}$ and $c_{22,0}$ in the DGP. We use a closed-form expression for the asymptotic variance of our estimators $\hat{b}$, $\hat{c}$, and $\hat{\phi}$ with Gaussian innovations and identity weighting matrix that we obtain from Theorem 2, and an algorithm based on the Kalman filter to compute numerically the asymptotic variance of the maximum likelihood estimator (MLE; see Appendixes E.5 and E.6 of the Supplementary Material). As expected, the asymptotic standard deviation ratios are larger than 1 for any DGP parameter values. For the loading parameter b_1,

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

the asymptotic efficiency loss of our estimator compared to MLE is moderate, with the asymptotic standard deviation ratio below 1.5 for the most part of the DGP parameter space (see Panel (a) of Figure 1). The efficiency loss is maximal for $(c_{11,0}, c_{22,0})$ near $(-0.3, -1)$. In Panels (b)–(d) of Figure 1, for the autoregressive parameters c_{11}, c_{22}, and ϕ, the asymptotic standard deviation ratio is large (unbounded) near the line $c_{11,0} = c_{22,0}$. In fact, on this line, the asymptotic variance of our estimator diverges since Assumption ID.2 is not met (see Proposition 4).[14] For parameters c_{11}, c_{22}, and ϕ, the minimal value of the asymptotic standard deviation ratio is 1.23, 1.42, and 1.75, respectively. Overall, Figure 1 shows that, in this example, the asymptotic standard deviation of our estimator is larger than the standard deviation of the MLE (i.e., the parametric efficiency bound) by a factor about 2, for DGP parameter values that are not close to the ones for which Assumptions ID.1–ID.3 do not hold.

4.3. Model Specification Testing

We use the Sargan–Hansen statistic to test the validity of the moment restrictions (3.6) underlying the identification of matrix $\boldsymbol{C}_0$. This is tantamount to testing the correct specification of the measurement equation, namely the property that $Y_t - \boldsymbol{C}_0\boldsymbol{Y}_{t-1}$ is the sum of a transformation of K dynamic latent factors plus a White Noise process. The Hansen J-statistic is $J_1 = T\text{vec}[\hat{Q}_{YZ} - \hat{\boldsymbol{C}}\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}]'(\hat{D}_c\hat{V}_\psi\hat{D}_c')^{-1}\text{vec}[\hat{Q}_{YZ} - \hat{\boldsymbol{C}}\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}]$, where $\hat{\boldsymbol{C}}$ is obtained with a consistent estimator $(\hat{D}_c\hat{V}_\psi\hat{D}_c')^{-1}$ of the optimal weighting matrix defined in Section 4.2. Vector $vec[\hat{Q}_{YZ} - \hat{\boldsymbol{C}}\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}]$ is a one-to-one transformation of a vector whose entries are sample autocovariances of the $\hat{\eta}_t$, and sample cross-covariances of the $\hat{\eta}_t$ and $\hat{\xi}_{t-j}$ with $j \neq 0$. In Appendix B.4, we show that, under the assumptions of Theorems 2 and 3 and the null hypothesis of correct specification of the moment restrictions (3.6), the statistic J_1 is asymptotically distributed as $\chi^2[Mn(n-K) - pn^2]$ when $T \to \infty$. The null hypothesis is rejected at the asymptotic level α when J_1 is above the $1-\alpha$ quantile of that chi-square distribution. We can use a similar approach to test the validity of the moment restrictions (3.8) implied by the VAR(q) dynamics of the factor. The Hansen statistic is $J_2 = T\text{vec}[\hat{Q}_{\xi W} - \hat{\boldsymbol{\Phi}}\hat{Q}_{\boldsymbol{\xi}_{-1}W}]'(\hat{D}_\phi\hat{V}_\psi\hat{D}_\phi')^{-1}\text{vec}[\hat{Q}_{\xi W} - \hat{\boldsymbol{\Phi}}\hat{Q}_{\boldsymbol{\xi}_{-1}W}]$, where $\hat{\boldsymbol{\Phi}}$ is obtained with a consistent estimator $(\hat{D}_\phi\hat{V}_\psi\hat{D}_\phi')^{-1}$ of the optimal weighting matrix defined in (B.11). The critical value is deduced from the asymptotic $\chi^2[K^2(L-q)]$ distribution.

The degrees of freedom for test statistics J_1 and J_2 grow with the numbers of instruments M and L. When those are large, the chi-square approximation may be poor in small samples. To cope with that issue, one could consider testing for a subset of the moment restrictions used for estimation.

[14] In the heat maps in Panels (b)–(d) of Figure 1, the maximal value on the diagonal blocks is set to 50, 70, and 20 to improve the visualization.

5. MODEL SELECTION

This section deals with determination of the factor space dimension K and the VAR orders p and q.

5.1. Selection of the Number of Dynamic Latent Factors

We first develop a test for hypotheses on the number of latent factors K, assuming that the number of lags p of process Y_t is known. From Assumption ID.1, integer K is equal to the rank of matrix A^*_{p+1}. Therefore, we can build on the extensive literature on rank testing (see, e.g., Anderson, 1951b; Cragg and Donald, 1996; Robin and Smith, 2000). Here, we follow more closely the F-test of Kleibergen and Paap (2006) and its generalizations in Al-Sadoon (2017). For a given integer $r < n$, we consider the null hypothesis $\mathscr{H}_0(r)$ of $K = r$ latent factors against the alternative hypothesis $\mathscr{H}_1(r)$ of more than r latent factors, i.e., $K > r$. The test statistic is

$$\mathscr{F}(r) = T \cdot vec(P_{\hat{N}_r}\hat{A}^*_{p+1}P_{\hat{M}_r})' \left\{(P_{\hat{M}_r} \otimes P_{\hat{N}_r})\hat{\Sigma}_{11}(P_{\hat{M}_r} \otimes P_{\hat{N}_r})\right\}^{\dagger} vec(P_{\hat{N}_r}\hat{A}^*_{p+1}P_{\hat{M}_r}), \tag{5.1}$$

where $P_{\hat{M}_r}$ (resp. $P_{\hat{N}_r}$) denotes the orthogonal projection onto the right (resp. left) null space of the rank-r approximation of matrix $\hat{A}^*_{p+1}$, and $\Sigma^{\dagger}$ is the Moore–Penrose inverse of matrix Σ. The rank-r approximation of matrix $\hat{A}^*_{p+1}$ can be obtained by singular value decomposition (SVD) as in Kleibergen and Paap (2006), or any other decomposition-based or norm-based approximations as in Definitions 2 and 3 of Al-Sadoon (2017).[15] Under the conditions of Theorem 2, we have $\sqrt{T}vec(\hat{A}^*_{p+1} - A^*_{p+1}) \xrightarrow{d} N(0, \Sigma_{11})$, where $\Sigma_{11} = S_{11}V_{\psi,11}S'_{11}$ with $S_{11} = (E[\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1}])^{-1} \otimes I_n$ and $V_{\psi,11} = \sum_{j=-\infty}^{\infty} cov(\tilde{Y}_{t-p-1} \otimes u^*_t, \tilde{Y}_{t-p-1-j} \otimes u^*_{t-j})$ (see Appendix B.2). Matrix $\hat{\Sigma}_{11}$ in (5.1) denotes a consistent HAC estimator of Σ_{11} as in Theorem 3. Using Theorem 1 and Corollary 1 in Al-Sadoon (2017), in Appendix B.6, we show that under the null hypothesis $\mathscr{H}_0(K)$, the statistic $\mathscr{F}(K)$ is asymptotically distributed as a $\chi^2[(n-K)^2]$ variate as $T \to \infty$. The number of degrees of freedom $(n-K)^2$ is the product of the dimensions of the left and right null spaces of A^*_{p+1}, both equal to $n-K$ under the null of K latent factors. We reject the null at the asymptotic significance level $\alpha \in (0,1)$ if $\mathscr{F}(K)$ is above $\chi^2_{1-\alpha}[(n-K)^2]$, i.e., the $1-\alpha$ quantile of the $\chi^2[(n-K)^2]$ distribution.

[15]Let $\hat{A}^*_{p+1} = \hat{U}\hat{\Lambda}\hat{V}'$ be the SVD of the $n \times n$ matrix $\hat{A}^*_{p+1}$, where $\hat{\Lambda} = diag(\hat{\sigma}_1, \ldots, \hat{\sigma}_n)$ is the diagonal matrix of singular values ranked in descending order, and $\hat{U}$ and $\hat{V}$ are orthogonal matrices of eigenvectors of $\hat{A}^*_{p+1}(\hat{A}^*_{p+1})'$ and $(\hat{A}^*_{p+1})'\hat{A}^*_{p+1}$, respectively. The rank-r approximation of matrix $\hat{A}^*_{p+1}$ is $\hat{U}_r\hat{\Lambda}_r\hat{V}'_r$, and the orthogonal projections on the left and right null spaces are $P_{\hat{N}_r} = \hat{U}_{n-r}\hat{U}'_{n-r}$ and $P_{\hat{M}_r} = \hat{V}_{n-r}\hat{V}'_{n-r}$, where $\hat{\Lambda}_r = diag(\hat{\sigma}_1, \ldots, \hat{\sigma}_r)$ and $\hat{U} = [\hat{U}_r \; : \; \hat{U}_{n-r}]$ and $\hat{V} = [\hat{V}_r \; : \; \hat{V}_{n-r}]$ are written in blocks of r and $n-r$ columns. For $r = 0$, we have $P_{\hat{N}_r} = P_{\hat{M}_r} = I_n$.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

THEOREM 4. *Under Assumptions M.1, IR.1, ID.1–ID.3, LS.1, and LS.2:*
(a) the test has asymptotic size α, *i.e.,* $\lim_{T\to\infty} \mathbb{P}\big(\mathscr{F}(K) > \chi^2_{1-\alpha}[(n-K)^2] | \mathscr{H}_0(K)\big) = \alpha$, *for any* $\alpha \in (0,1)$, *and (b) is consistent, i.e.,* $\lim_{T\to\infty} \mathbb{P}\big(\mathscr{F}(K) > \chi^2_{1-\alpha} [(n-K)^2] | \mathscr{H}_1(K)\big) = 1$.

Moreover, (c) let $\hat{K} = \min_{k\in\{0,\ldots,n-1\}} \Big\{k : \mathscr{F}(r) > \chi^2_{1-\alpha_T}[(n-r)^2]$ *for* $r < k$ *and* $\mathscr{F}(k) \le \chi^2_{1-\alpha_T}[(n-k)^2]\Big\}$, *where* $\alpha_T \to 0$ *s.t.* $\frac{1}{T}\log(\alpha_T) \to 0$. *Then,* $\hat{K} = K_0$ *w.p.a. 1 as* $T \to \infty$, *where* K_0 *is the true number of factors.*

In Theorem 4(c), as with other consistent rank tests (see, e.g., Robin and Smith, 2000), to estimate the number of factors K, we adopt a sequential testing procedure for $\mathscr{H}_0(r)$ against $\mathscr{H}_1(r)$ for $r = 0, 1, \ldots, n-1$. The estimator $\hat{K}$ is the smallest integer $k \le n-1$ for which the statistic $\mathscr{F}(k)$ is not above the critical value. It is well known that such an estimator with fixed significance level α is not consistent because of type-I error at the step of testing $\mathscr{H}_0(K)$ against $\mathscr{H}_1(K)$. To recover consistency in Theorem 4, we let $\alpha = \alpha_T \to 0$ at a suitable rate implied by Theorem 5.8 in Poetscher (1983).

In the sequential procedure, when testing $\mathscr{H}_0(r)$ against $\mathscr{H}_1(r)$ with $0 < r < n$, the null hypothesis contains DGPs with exactly $K = r$ common factors but not those with $K < r$ common factors. Chen and Fang (2019) stresses some potential shortcomings and investigates alternative procedures. In our Monte Carlo experiments in Section 6.4, we find a good performance of the selection procedure.

In the literature on reduced rank regression (e.g., Anderson, 1951a; Robinson, 1973; Reinsel and Velu, 1998), the Likelihood Ratio statistic has a limiting $\chi^2[(n-K)^2]$ distribution. In our case, the VAR($p+1$) regression in Section 3.1 is a pseudo-model and such results cannot be invoked. For the same reason, the asymptotic variance matrix of $vec(\hat{A}^*_{p+1})$ does not admit a Kronecker product form in general, even under conditionally homoskedastic DGP. Hence, we cannot use the weighting in Corollary 3.1 of Robin and Smith (2000) to get a $\chi^2[(n-K)^2]$ distribution for their test.

5.2. Selection of the Number of Lags *p* and *q*

In this subsection, we first provide a consistent selection method for the number of lags p in the VAR dynamics of the endogenous vector Y_t. Let p_0 be the true (unknown) number of lags. For a given integer $p^\star \ge p_0 + 2$, we consider the VAR($p^\star$) pseudo-model

$$Y_t = \sum_{j=1}^{p^\star} A^\star_j Y_{t-j} + u^\star_t, \tag{5.2}$$

with pseudo-true matrix parameters $A^\star_j = C_{j,0} + B_0 F_j^{(p^\star)}$, for $j \le p_0$, and $A^\star_j = B_0 F_j^{(p^\star)}$, for $p_0 + 1 \le j \le p^\star$, where $F_j^{(p^\star)}$ is the matrix coefficient of Y_{t-j}

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

is the linear projection of f_t onto $Y_{t-1},\dots,Y_{t-p^\star}$, namely $EL(f_t|Y_{t-1},\dots,Y_{t-p^\star}) = \sum_{i=1}^{p^\star} F_i^{(p^\star)} Y_{t-i}$, and $u_t^\star$ is the residual of the linear projection of Y_t onto $Y_{t-1},\dots,Y_{t-p^\star}$. The selection method is based on the fact that $B'_{0\perp}[A^\star_{p_0+1} : \cdots : A^\star_{p^\star}] = 0$ for any $p^\star \geq p_0+2$, i.e., the matrices $A^\star_j$ for $j \geq p_0+1$ share a common left null space. We provide next the iterative algorithm for selecting p. We start the procedure with $p = 0$.

(1) Estimate the pseudo-parameters $\hat{A}^\star_i$, for $i = 1,\dots,p^\star$, in model (5.2) using multivariate OLS.
(2) Estimate the VAR pseudo-model $Y_t = \sum_{j=1}^{p+1} A_j^* Y_{t-j} + u_t^*$ and select $\hat{K}(p)$ from $\hat{A}^*_{p+1}$ with the method proposed in Section 5.1.[16] If $\hat{K}(p) = n$, then set $\mathscr{S}_p = +\infty$, update $p \curvearrowright p+1$, and redo Step 2. If $\hat{K}(p) < n$, proceed with Step 3.
(3) Estimate matrix $\hat{B}(p)$ from $\hat{A}^*_{p+1}$ with the procedure in Section 4.1, compute $\hat{B}_\perp(p)$, and construct the statistic $\mathscr{S}_p = T \cdot vec\left(\hat{B}_\perp(p)'[\hat{A}^\star_{p+2} : \cdots : \hat{A}^\star_{p^\star}]\right)' \hat{\Sigma}_{22}^{-1}$ $vec\left(\hat{B}_\perp(p)'[\hat{A}^\star_{p+2} : \cdots : \hat{A}^\star_{p^\star}]\right)$, where $\hat{\Sigma}_{22}$ is the consistent estimator of the asymptotic covariance matrix of vector $\sqrt{T} vec\left(\hat{B}_\perp(p)'[\hat{A}^\star_{p+2} : \cdots : \hat{A}^\star_{p^\star}]\right)$ defined in (B.14). Given an asymptotic significance level $\alpha \in (0,1)$ and the critical value $c_{1-\alpha}(p) := \chi^2_{1-\alpha}[(n-\hat{K}(p))n(p^\star - p - 1)]$, if $\mathscr{S}_p \leq c_{1-\alpha}(p)$, then $\hat{p} = p$ is the selected number of lags. If $\mathscr{S}_p > c_{1-\alpha}(p)$, then update $p \curvearrowright p+1$ and repeat Steps 2 and 3 until the condition $\mathscr{S}_p \leq c_{1-\alpha}(p)$ is met.

Assumption SEL.1. The $n \times n$ matrix $D_i = [C'_{i,0}\bar{B}_0 + F_i^{(p^\star)\prime} \;:\; C'_{i,0}B_{0\perp}]$ is either nonsingular, or such that the projection of its kernel on the first K_0 components of $\mathbb{R}^n$ is nonnull, for all $i = 1,\dots,p_0$.

In Appendix B.7, we show that $\mathscr{S}_p$ has a chi-square distribution under the null hypothesis $p = p_0$ and that Assumption SEL.1 is sufficient for the consistency of the testing procedure. It ensures that the left null space of A^*_{p+1} does not annihilate all columns of matrix $[A^\star_{p+2} : \cdots : A^\star_{p^\star}]$ when $p < p_0$.

THEOREM 5. *Under Assumptions M.1, IR.1, ID.1–ID.3, LS.1, LS.2, and SEL.1, we have: (a) under the null hypothesis $p = p_0$, the asymptotic distribution of the statistic is $\mathscr{S}_p \xrightarrow{d} \chi^2[n(n-K_0)(p^\star - p_0 - 1)]$ as $T \to \infty$; (b) the test is consistent, i.e., under the alternative hypothesis $p < p_0$, we have $\mathscr{S}_p \xrightarrow{p} +\infty$; and (c) if α_T is such that $\alpha_T = o(1)$ and $T^{-1}\log \alpha_T = o(1)$, the selection procedure is consistent: $\mathbb{P}[\hat{p} = p_0] \to 1$, as $T \to \infty$.*

The proposed selection procedure for p requires the knowledge of an upper bound on p_0. It does not require knowledge of the true number of latent factors K_0. The latter number is estimated along the selection procedure, and $\hat{K}(\hat{p})$ coincides with the selection obtained in Theorem 4(c) based on p_0 w.p.a. 1 and is consistent

[16]The notation $\hat{K}(p)$ and $\hat{B}_\perp(p)$ highlights the fact that, in the proposed algorithm, $\hat{K}$ and $\hat{B}_\perp$ are functions of p.

for K_0. The selection procedure for K and p is semiparametric, i.e., is valid with unspecified factor dynamics. Furthermore, note that the algorithm starts with $p = 0$. Hence, if $\hat{p} > 0$, we reject a model without contagion.

Finally, the selection for the number of autoregressive lags q in the factor dynamics follows by the fact that ξ_t is a VARMA(q,q) process (see Section 3.3). Given the results in Hannan and Deistler (1988, p. 205), the selection of q is performed by using information criteria.

6. MONTE CARLO ANALYSIS

We conduct Monte Carlo experiments in order to investigate the finite-sample properties of the parameter estimators of the state-space model, of the test on the number of latent factors, and of the selection procedures for the model orders K,p,q. The DGPs are defined in Section 6.1. The results for the estimators, test statistics, and model selection procedure are described in Sections 6.2–6.4, respectively.

6.1. Data Generating Processes

We consider the following DGPs.

(i) DGPs 1–2: We start with a specification with $n = 4$ observed variables and a single dynamic latent factor, i.e., $K = 1$. The VAR orders are $p = q = 1$. In DGP 1, the innovations are Gaussians $u_t \sim iiN(0, I_5)$ and $v_t \sim iiN(0,1)$. In DGP 2, the innovations are Student $u_{i,t} \sim t_4$, independent across i, and $v_t \sim t_4$. The values of parameters in C_0, B_0, and ϕ_0 are the same in DGPs 1 and 2, and are reported in the second column of Table 2 in Appendix G of the Supplementary Material.

(ii) DGPs 3–7: We next turn to multifactor specifications. The number of observables and latent factors is $n = 5$ and $K = 2$, respectively, the innovations are Gaussians $u_t \sim iiN(0, I_5)$ and $v_t \sim iiN(0, \sigma_v^2 I_2)$, the VAR orders are $p = q = 1$, and the state-space representation is

$$Y_t = C_0 Y_{t-1} + B_0 f_t + u_t, \qquad f_t = \begin{bmatrix} \phi_0 & 0 \\ 0 & -\phi_0 \end{bmatrix} f_{t-1} + \sqrt{1-\phi_0^2}\, v_t. \tag{6.1}$$

The parameterization of DGPs 3–7 in (6.1) allows us to disentangle the effect of the factor variance, controlled by parameter σ_v^2, from the effect of the factor persistence, given by autocorrelation parameter ϕ_0. The five DGPs differ in terms of the values of parameters σ_v^2 and ϕ_0, as reported in Table 1.

The values of the other DGP parameters are provided in Tables 7–15 in Appendix G.1 of the Supplementary Material. Matrices C_0 and Φ_0 have eigenvalues inside the unit circle, and matrix B_0 meets the normalization restriction in Assumption IR.1. We have checked (numerically) that the identification Assumptions ID.1–ID.3 hold. The matrices B_0 and C_0 are kept constant across DGPs 3–7 for comparability reasons and for focusing the Monte Carlo analysis on the effects of variability and persistence of the common factor.

Table 1. Parameter values of DGPs 3–7.

Parameter	DGP 3	DGP 4	DGP 5	DGP 6	DGP 7
ϕ_0	0.9	0.9	0.7	0.7	0.4
σ_v	1	0.6	1	0.6	0.6

(iii) DGP 8: We have $n = 5$, $K = 2$, $u_t \sim iiN(0, I_5)$, and $v_t \sim iiN(0, I_2)$, and $Y_t = C_0 Y_{t-1} + B_0 f_t + u_t$,

$$\begin{bmatrix} f_{1,t} \\ f_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_0 & 0 \\ 0 & -\phi_0 \end{bmatrix} \begin{bmatrix} f_{1,t-1} \\ f_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0.6 \cdot \sqrt{1 + 0.8 \cdot f_{1,t-1}^2} v_{1t} \\ 0.6 \cdot \sqrt{1 + 0.8 \cdot f_{2,t-1}^2} v_{2t} \end{bmatrix},$$

with $\phi_0 = 0.7$. The bivariate common factor f_t in DGP 8 follows a VAR process of order 1 with autoregressive conditionally heteroscedastic (ARCH) dynamics in the errors. This specification is compatible with the state-space model in (2.1) and (2.2) since the error terms are White Noise processes. The matrices B_0 and C_0 are as in DGPs 3–7.

(iv) DGP 9: Finally, we consider a model with $p = 2$ lags in the VAR specification. For the rest, $n = 4$, $K = q = 1$, $u_t \sim iiN(0, I_5)$, and $v_t \sim iiN(0, 1)$ as in DGP1.

6.2. Finite-Sample Properties of the Estimators

In this subsection, we describe the results for the finite-sample bias and standard deviations of the estimators $\hat{B}$, $\hat{C}$, and $\hat{\Phi}$ defined in equations 4.1–4.3, with $M = L = 10$ and identity weighting matrices. For each DGP, we consider sample sizes $T = 500$, $1,000$, and $5,000$. In each setting, the number of Monte Carlo replications is $N_{rep} = 1,000$. The results are shown in the tables provided in Appendix G.1 of the Supplementary Material. We have checked that the results are similar under moderate changes of tuning parameters M and L.

Let us start with the results for DGP 1 displayed in Table 2 in the Supplementary Material. As expected, the bias and the standard deviation of the estimators decrease as the sample size T increases. The bias is typically smaller for the coefficients in loadings vector B. The estimator of the factor autoregressive coefficient ϕ is negatively biased. As a comparison, in Table 3 in the Supplementary Material, we provide the results for the ML estimator. As expected, the standard deviations for the ML estimator in a model with Gaussian errors are smaller than the standard errors of our estimator. The standard deviations differ by a factor about 2. In Figures 3 and 4 in the Supplementary Material, we display the distribution of execution times for computing the estimates with our procedure, and with ML, respectively ($T = 5,000$).[17] The former is in closed form and requires about 1 second per simulated sample. For the latter, the median execution time is

[17] We use a server with 2.10 GHz processor and 512 GB of RAM.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

approximately 60 seconds. The difference is larger in specifications with a larger number of parameters. In Table 4 in the Supplementary Material, we investigate the impact of the initial value for computing the ML estimates. We compare bias and standard deviation when the maximization algorithm is initialized at the true parameter values versus the case when a random initial value is chosen.[18] With random initial values, the bias and standard deviation of MLE are substantially larger, and larger than for our estimator.

The bias and standard deviation for our estimators and the Gaussian QMLE in DGP 2 with Student innovations are provided in Tables 5 and 6 in the Supplementary Material. The results are similar as those for DGP 1. Heavy tails do not affect substantially the performance of the estimators in this DGP.

Let us now consider the results in the two-factor specifications DGPs 3–7 with Gaussian errors. We observe that, for a given sample size, the accuracy of the estimator $\hat{B}$ is higher when the latent factors are persistent and/or sufficiently volatile (relative to the idiosyncratic shocks). Indeed, the smallest standard deviations for the elements of $\hat{B}$ are found in DGP 3 (see Table 7 in the Supplementary Material). In this DGP, the estimates of the coefficients in $\hat{B}$ are rather accurate even for sample sizes as small as $T = 500$, with root mean square error (RMSE) about 20% of the parameter absolute value. The standard deviations of the estimates in $\hat{B}$ increase as we move to DGPs 4–6 characterized by less volatile and/or less persistent latent factors (Tables 9–11 in the Supplementary Material). In DGP 7 with small factor variance and low persistence, estimator $\hat{B}$ features large bias and standard deviation in our simulations for sample sizes $T = 500$ and $T = 1,000$, while it works well for sample size $T = 5,000$ (see Table 12 in the Supplementary Material). The poor performance for the smaller sample sizes is driven mostly from some simulations with extreme estimates. For this reason, in Table 13 in the Supplementary Material, we report the difference between the median of the estimates and the true parameter value, and a standard deviation measure obtained from the interquartile range. These findings show that the distribution of estimates is more centered and less dispersed than what Table 12 in the Supplementary Material suggests.

The performance of estimators $\hat{C}$ and $\hat{\Phi}$ is quite homogenous across DGPs 3–7, typically with smaller standard deviations when the latent factors are less volatile (DGPs 4, 6, and 7). The RMSEs of the estimators in $\hat{C}$ are in the range of 0.20–0.30 for the smaller sample size $T = 500$, and in the range of 0.10–0.20 for the intermediate sample size $T = 1,000$. The biases and standard deviations of estimator $\hat{\Phi}$ are large for $T = 500$, and decrease with sample size. For instance, in DGP 3 with a more volatile latent factor, the RMSEs of the estimators of parameters ϕ_{11} and ϕ_{22} are close to 0.55 and 0.90, that are about 60% and 100% of the parameter absolute value, respectively (see Table 7 in the Supplementary Material). In Table 8 in the Supplementary Material, we report

[18] In the second case, the initial value is the true parameter value plus a uniform random draw in interval $(-0.5, 0.5)$. The maximum number of iterations is fixed to 10,000 in both cases.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

the results using the median and interquartile range to assess the distortion and dispersion of the estimates. Distortion and dispersion are smaller when assessed from the quantiles of the distribution especially for $T = 1,000$ and $T = 5,000$, and the decrease of standard deviations with sample size is more evident and in accordance with root-T scaling from asymptotic theory. In fact, the distributions of the elements of $\hat{\Phi}$ with $T = 500$ are bimodal. This finding may appear quite surprising. Bimodal distributions may result from, e.g., incorrect order selection, or parameter values close to the region of identification failure. Because we use the true orders $K = 2$ and $p = q = 1$ for estimation, the first explanation is ruled out. To assess the possibility of being near to identification failure, we remark that the second eigenvalue of matrix $A_2^*(A_2^*)'$ involved in Assumption ID.1 is one order of magnitude smaller than the first eigenvalue in DGPs 3–7. A similar remark applies to Assumption ID.2. This fact may explain the relatively low precision of estimators $\hat{B}$ and $\hat{C}$ for small sample size $T = 500$ in some DGPs. In turn, this induces noisy estimates $\hat{\xi}_t$ of ξ_t. In unreported simulations, we notice that, when using the true values of ξ_t instead of the estimates $\hat{\xi}_t$, the infeasible estimator of Φ performs well with unimodal distributions peaked near the true value. We deduce that the estimation error of $\hat{B}$ and $\hat{C}$ is mostly responsible for the poor performance of the feasible estimator $\hat{\Phi}$ with $T = 500$.

Finally, we comment on the results for the last two DGPs. The Monte Carlo results for the parameter estimators $\hat{B}$ and $\hat{C}$ when data are generated from DGP 8 with ARCH errors (Tables 14 and 15 in the Supplementary Material) feature small bias. Standard deviations are rather large for some elements of $\hat{C}$ with $T = 1,000$, but generally decrease with the sample size, confirming that the methodology works also when the errors are neither i.i.d. nor Gaussian. The estimator $\hat{\Phi}$ is reliable for the larger sample size $T = 5,000$ only. Table 16 in the Supplementary Material shows the results for DGP 9 with $p = 2$ lags. The bias and standard deviations are small, and comparable with the results in DGP 1. Hence, the inclusion of a second lag does not make the performance of the estimators worse in this DGP.

Overall, in our Monte Carlo experiments, we find good results for estimators $\hat{B}$ and $\hat{C}$ with $T = 1,000$, if not already with $T = 500$, in most of the considered DGPs. For estimator $\hat{\Phi}$, the larger sample size $T = 5,000$ is needed to get reliable results in some multifactor models.

6.3. Finite-Sample Properties of the Rank Test for the Number of Factors

We conduct a Monte Carlo analysis of the finite-sample size and power properties of the test for the number of unobservable factors defined in Section 5.1. We simulate the processes with the two-factor specifications in DGPs 3–8. The number of Monte Carlo replications is $N_{rep} = 1,000$.

The results for the empirical size refer to the test of the null hypothesis of $K = 2$ unobservable factors, against the alternative with $K > 2$. We collect the results

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

in Table 17 in Appendix G.2 in the Supplementary Material. Overall, the results show that the empirical size is close to the nominal size α, for $\alpha = 10\%, 5\%, 1\%$, or 0.5%. Size distortions are smaller than 3% even for sample size $T = 500$ in DGPs 3–6 and 8. When data are drawn from DGP 7, size distortions are bigger and the empirical and nominal sizes are close only for the largest considered sample size $T = 5,000$. This finding is in line with the relatively poor performance of the factor loading estimators in small samples reported in Section 6.2 for this DGP with low persistency and small volatility of the latent factors.[19]

The results for the empirical power displayed in Table 18 in the Supplementary Material refer to the test of the null hypothesis of $K = 1$ unobservable factors, against the alternative $K > 1$, when data are generated from DGPs 3–8 with $K = 2$. The test features overall good power properties. For nominal size 5%, the power is approximately 70% or more for most DGPs, already with sample size $T = 500$. Again, for DGP 7, the power is satisfactory only for the largest sample size.

6.4. Finite-Sample Properties of the Selection Procedure for Model Orders

In Table 19 in the Supplementary Material, we report the percentages of selected model orders combinations (p,K). Orders (p,K) are selected according to the procedure in Section 5.2 with different nominal levels $\alpha = 5\%, 1\%, 0.5\%$, and 0.1%. Data are generated according to DGP 1 with sample sizes $T = 500$, $T = 1,000$, and $T = 5,000$. For $T = 1,000$ and $T = 5,000$, the proportion of times we select the correct model $(1,1)$ is the largest and above 99%, when the smallest nominal size $\alpha = 0.1\%$ is adopted. For $T = 500$, the best performance is obtained with $\alpha = 0.5\%$, with the correct model selected in approximately 85% of the repetitions. With the smallest size $\alpha = 0.1\%$, the performance deteriorates, because a model with no factors, i.e., $K = 0$, is selected approximately 30% of the time given the large critical value. The results are in line with the theoretical analysis requiring the size α to shrink to 0 with growing sample size.

In order to expand the analysis and include the selection of q, in Table 20 in the Supplementary Material, we report the percentages of selected model orders combinations (p,K,q). Orders (p,K) are selected according to the procedure presented in Section 5.2 with $\alpha = 0.01$, and q is selected using the Bayesian information criterium (BIC). Due to the numerical burden in the selection of q, in this exercise, we focus on DGP 1 with sample size $T = 1,000$. In 66% of our simulations, we select the correct model orders combination $p = q = K = 1$. In approximately 20% of our simulations, we overestimate the order q of the factor dynamics. The selection of (p,K) is correct in approximately 90% of the simulations in line with the results found in Table 19 in the Supplementary Material.

[19] We consider DGPs with a number of observables $n \leq 5$ because the finite-sample properties of the rank test when $n > 5$ are rather poor.

7. EMPIRICAL APPLICATION

In this section, we conduct an empirical analysis on the interconnectedness among the realized volatility series of four European stock market indices. We use our state-space model to separate the effects of systematic volatility shocks from volatility spillovers across markets.

7.1. Data and State-Space Model

We collect the daily 5-minute realized volatilities of four European stock market indices, which are the French CAC 40, the OMX Stockholm 30, the Spanish IBEX 35, and the German DAX 30, from June 1, 2009 to November 19, 2021, i.e., $T = 3,160$ observations.[20] Figure 2 displays the four realized volatility series. The series feature substantial co-movements, and one observes the generalized increase of volatilities during the pandemics. We transform the series by taking the log for the rest of the analysis. In Appendix F of the Supplementary Material, we provide summary statistics (Table 1) and auto- and cross-correlation functions of the four time series (Figure 1). In all markets, realized volatilities display positive autocorrelations slowly decaying with the lag, and the cross-correlation is positive and long-lasting for all pairs of markets. Moreover, the stationary distribution of the individual series display non-Gaussianity features with negative skewness and kurtosis larger than 3.[21]

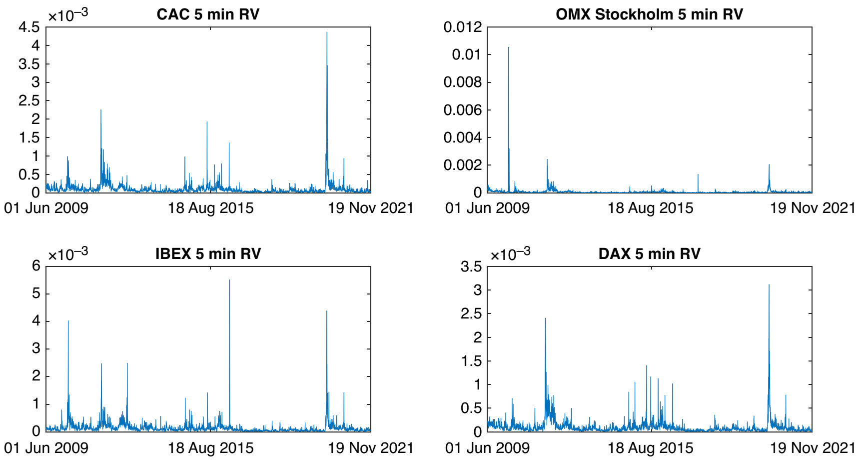

FIGURE 2. This figure plots the daily series of 5-minute realized volatilities of CAC 40, OMX 30, IBEX 35, and DAX 30 market indices from June 1, 2009 to November 19, 2021.

[20] The data can be downloaded from the website https://realized.oxford-man.ox.ac.uk/data.

[21] The Jarque–Bera test rejects the normality assumption at the 5% level for all series of log realized volatilities. For each series, the Dickey–Fuller test rejects the null of unit root.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

We estimate the state-space model in equations (2.1) and (2.2), where the vector of observables $Y_t = (Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})'$ contains the demeaned values of the log realized volatilities of the French, Swedish, Spanish, and German stock market indices. First, the autoregressive matrices $C_1, \dots, C_p$ represent volatility spillover effects among the European financial markets. The reaction of a given stock market volatility to shocks in volatilities of other markets can be due to either financial or macroeconomic linkages, or correlation in assets and strategies of the firms in these markets. Second, the K-dimensional unobservable common factors vector f_t represents variables that have a pervasive impact on the European stock market volatility. This systematic source of uncertainty has a differentiated impact across countries as a function of the loadings in matrix B.

7.2. Model Selection and Parameter Estimates

The first empirical task that we tackle is the joint selection of the number of lags p and the number of unobservable factors K. For the estimation of these quantities, we adopt a VAR pseudo-model with $p^\star = 8$ lags (see Section 5.2). The procedure outlined in Theorem 5 with $\alpha = 0.01$ selects two unobservable factors and four contagion lags, i.e., $\hat{K} = 2$ and $\hat{p} = 4$. The data-driven selection procedure finds evidence of the necessity of both contagion and latent factors to explain co-movements in the log volatilities series. Once orders p and K are selected, we estimate the loadings matrix B and the contagion matrices C_j with the procedure in Section 4.1. The estimates are

$$\hat{B} = \begin{bmatrix} \underset{(0.1695)}{\mathbf{-0.5417}} & \underset{(0.2367)}{0.4457} \\ \underset{(0.1497)}{-0.1594} & \underset{(0.2171)}{\mathbf{0.6040}} \\ \underset{(-)}{\mathbf{1.0000}} & \underset{(-)}{0.0000} \\ \underset{(-)}{0.0000} & \underset{(-)}{\mathbf{1.0000}} \end{bmatrix},$$

$$\hat{C}_1 = \begin{bmatrix} \underset{(0.0458)}{\mathbf{0.3736}} & \underset{(0.0462)}{0.0151} & \underset{(0.0471)}{\mathbf{0.2235}} & \underset{(0.0493)}{0.0502} \\ \underset{(0.0415)}{-0.0090} & \underset{(0.0408)}{\mathbf{0.2457}} & \underset{(0.0361)}{\mathbf{-0.1402}} & \underset{(0.0370)}{\mathbf{-0.1061}} \\ \underset{(0.0394)}{\mathbf{0.1655}} & \underset{(0.0406)}{0.0452} & \underset{(0.0404)}{\mathbf{0.1122}} & \underset{(0.0483)}{-0.0187} \\ \underset{(0.0418)}{0.0214} & \underset{(0.0413)}{-0.0155} & \underset{(0.0453)}{\mathbf{-0.1227}} & \underset{(0.0440)}{\mathbf{0.2437}} \end{bmatrix},$$

$$\hat{C}_2 = \begin{bmatrix} \underset{(0.0622)}{\mathbf{0.1229}} & \underset{(0.0625)}{0.0136} & \underset{(0.0548)}{0.0693} & \underset{(0.0700)}{0.0964} \\ \underset{(0.0541)}{0.0118} & \underset{(0.0552)}{\mathbf{0.2493}} & \underset{(0.0474)}{0.0820} & \underset{(0.0517)}{0.1066} \\ \underset{(0.0495)}{0.0343} & \underset{(0.0533)}{0.0760} & \underset{(0.0452)}{\mathbf{0.1677}} & \underset{(0.0559)}{\mathbf{0.1363}} \\ \underset{(0.0588)}{0.0418} & \underset{(0.0569)}{0.0909} & \underset{(0.0501)}{\mathbf{0.1345}} & \underset{(0.0612)}{\mathbf{0.3639}} \end{bmatrix},$$

$$\hat{C}_3 = \begin{bmatrix} \underset{(0.0440)}{\mathbf{0.1332}} & \underset{(0.0454)}{-0.0105} & \underset{(0.0422)}{-0.0690} & \underset{(0.0444)}{-0.0032} \\ \underset{(0.0525)}{0.0603} & \underset{(0.0482)}{\mathbf{0.2085}} & \underset{(0.0484)}{0.0517} & \underset{(0.0549)}{\mathbf{0.1672}} \\ \underset{(0.0483)}{\mathbf{0.1754}} & \underset{(0.0524)}{-0.0029} & \underset{(0.0442)}{-0.0687} & \underset{(0.0432)}{-0.0420} \\ \underset{(0.0428)}{0.0706} & \underset{(0.0447)}{0.0805} & \underset{(0.0431)}{0.0012} & \underset{(0.0522)}{\mathbf{0.2955}} \end{bmatrix},$$

$$\hat{C}_4 = \begin{bmatrix} \underset{(0.0949)}{-0.0045} & \underset{(0.1094)}{0.0283} & \underset{(0.0898)}{0.0001} & \underset{(0.1098)}{-0.0169} \\ \underset{(0.0876)}{-0.0338} & \underset{(0.0911)}{-0.0416} & \underset{(0.0811)}{0.0190} & \underset{(0.0830)}{-0.0945} \\ \underset{(0.0784)}{\mathbf{0.3831}} & \underset{(0.0877)}{0.0132} & \underset{(0.0779)}{0.0621} & \underset{(0.0950)}{0.0457} \\ \underset{(0.1007)}{-0.0425} & \underset{(0.1022)}{-0.0696} & \underset{(0.0836)}{-0.0027} & \underset{(0.1040)}{-0.0836} \end{bmatrix}.$$

Below each estimate, we display the standard errors from Theorem 3.[22] The coefficient estimates displayed in bold are statistically significant at the 5% level. The standard errors for the elements of B are rather large (greater than 0.10) despite the sample consisting of more than 3,000 observations. In our normalization, the dynamic latent factors are rotated such that the first factor impacts on the IBEX 35 index (with unit loading), but not on the DAX 30 index, and vice versa for the second factor. From the estimated loadings matrix $\hat{B}$, the first factor has opposite effects on the French and Spanish stock market volatilities, whereas the second factor impacts positively on both the German and Swedish indices (as well as on the French one, although the loading coefficient is significant at the 10% level only). Focusing on the contagion matrix at lag one $\hat{C}_1$, the third column displays statistically significant volatility spillover effects from the Spanish index to all the other stock markets over 1-day horizon. That spillover coefficient is positive for the French CAC 40, but negative for the Swedish OMX 30 and the German DAX 30 indices. We find also a positive spillover effect from the French stock market to the Spanish market index, and a negative effect from the German market to the Swedish one. Overall, these estimates suggest a certain geographic segmentation, with the German and Swedish stock markets loading positively on the second latent factor, and the Spanish and French markets loading on the first factor with opposite signs while featuring positive volatility spillover effects among each other, as well as negative spillover effects from the Spanish index toward the German and Swedish stock markets.

For a preliminary analysis of the volatility spillover effects at horizons longer than 1 day, we consider the matrices $\mathfrak{C}_h$ for $h \geq 1$, i.e., the matrix coefficients in the series expansion $C(L)^{-1} = \sum_{h=0}^{\infty} \mathfrak{C}_h L^h$. Indeed, $(\mathfrak{C}_h)_{i,j}$ yields the linear response of $Y_{i,t+h}$ to a unit change in $Y_{j,t}$ holding fixed the path of the latent factor,

[22] For the HAC estimator $\hat{V}_\psi$, we use the Bartlett kernel with 10 lags. For the estimation of matrix C, we use $M = 25$ lags of $\hat{\eta}_t$ in the instrument vector $\hat{\boldsymbol{Z}}_t$ and implement the optimal weighting matrix. This choice is dictated by the necessity of having a condition number for matrix $\hat{Q}_{Y_{-1}Z}$ smaller than 20, say. The estimates are rather stable to changes in M, when M is greater than 25.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

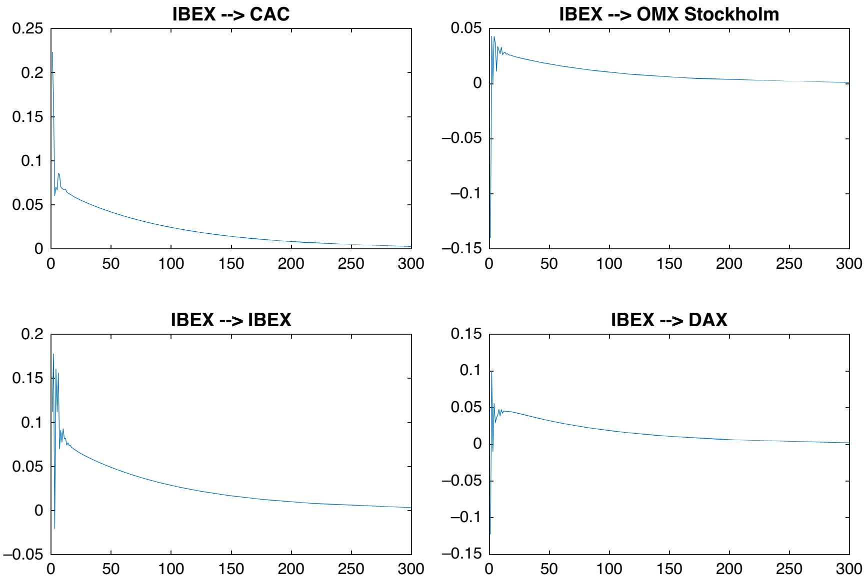

FIGURE 3. The panels of this figure plot the IRFs to shocks in realized volatility of IBEX 35 index, for the DAX, OMX Stockholm, IBEX, and DAX indices. The dynamic latent factor is held fixed in the definition of the impulse responses.

namely $(\mathfrak{C}_h)_{i,j} = \frac{\partial EL[Y_{i,t+h}|\Omega_t]}{\partial Y_{j,t}}$, where the information set Ω_t contains $Y_t, Y_{t-1}, \ldots$ and $\ldots, f_{t+1}, f_t, f_{t-1}, \ldots$. In Figure 3, we display the coefficients $(\mathfrak{C}_h)_{i,3}$ as a function of lag h, for $i = 1, \ldots, 4$, i.e., the percentage increase in realized volatility of market i at horizon h in response to 1% increase in the realized volatility of IBEX 35. The four panels show long-lasting positive volatility spillover effects from the Spanish stock market index to the other ones. In fact, the largest estimated eigenvalue of the companion matrix corresponding to the VAR(4) dynamics is equal to 0.98. The definition of IRFs for idiosyncratic shocks and systematic shocks separately is beyond the scope of this paper and is left for future research.

We now deal with model selection and parameter estimation for the VAR factor dynamics. The BIC criterion selects $\hat{q} = 6$ lags. The estimates of the autoregressive matrix coefficients and their standard errors are[23]

$$\hat{\Phi}_1 = \begin{bmatrix} \underset{(0.2448)}{\mathbf{0.5189}} & \underset{(0.2909)}{-0.0557} \\ \underset{(0.3672)}{0.5427} & \underset{(0.3344)}{0.5137} \end{bmatrix}, \quad \hat{\Phi}_2 = \begin{bmatrix} \underset{(0.2437)}{-0.2294} & \underset{(0.2758)}{-0.0823} \\ \underset{(0.1517)}{\mathbf{-0.5029}} & \underset{(0.2376)}{-0.3128} \end{bmatrix},$$

[23] We use $L = 50$ lagged instruments. Choices of L smaller than 45 lead to unstable estimates, whereas larger values of L may result in poorer finite-sample properties.

$$\hat{\Phi}_3 = \begin{bmatrix} \underset{(0.3247)}{0.1630} & \underset{(0.2804)}{-0.0828} \\ \underset{(0.2545)}{0.0367} & \underset{(0.2619)}{-0.0642} \end{bmatrix}, \quad \hat{\Phi}_4 = \begin{bmatrix} \underset{(0.2288)}{0.1268} & \underset{(0.2630)}{0.0317} \\ \underset{(0.2395)}{-0.3393} & \underset{(0.2424)}{0.0202} \end{bmatrix},$$

$$\hat{\Phi}_5 = \begin{bmatrix} \underset{(0.4334)}{0.3522} & \underset{(0.5198)}{-0.0803} \\ \underset{(0.1665)}{0.0647} & \underset{(0.3119)}{-0.0815} \end{bmatrix}, \quad \hat{\Phi}_6 = \begin{bmatrix} \underset{(0.2824)}{-0.0040} & \underset{(0.3349)}{0.0409} \\ \underset{(0.1793)}{0.0810} & \underset{(0.1651)}{0.0345} \end{bmatrix}.$$

The interpretation of the elements of matrices Φ_j is not straightforward because the factors are latent and their dynamic parameters are conditional on the chosen normalization for B. The estimated autoregressive matrix at lag one displays positive coefficients among the numerically large entries, although only one is statistically significant at the 5% level—the AR coefficient for the factor with opposite effects on the French and Spanish stock markets. The moduli of the eigenvalues of the estimated companion matrix $\tilde{\Phi}$ in the FAVAR representation (Appendix A.1), which are invariant to the chosen factor normalization, range between 0.3050 and 0.9867, matching the stationarity condition and highlighting the persistence in the latent factor process. Moreover, the estimate of the unconditional variance $V(f_t)$ of the latent factor vector is $\begin{bmatrix} 0.3055 & -0.1154 \\ -0.1154 & 0.6994 \end{bmatrix}$. The two latent factors have a similar scale, and are weakly correlated unconditionally.

7.3. Model Specification Testing

In this subsection, we deal with specification testing for the state-space model. In Figure 2 in Appendix F of the Supplementary Material, we plot the auto- and cross-correlograms of the bivariate serially dependent component $\hat{\xi}_t$, and the two-dimensional White Noise component $\hat{\eta}_t$. We use the estimated processes obtained from $\hat{B}$ and $\hat{C}$. The estimated autocorrelation functions of the White Noise components η_{it}, $i = 1, 2$, are almost flat and close to zero, consistently with the population properties. Moreover, we observe almost zero correlation between $\hat{\xi}_{jt}$, for $j = 1, 2$, and lags and leads of $\hat{\eta}_{it}$, for $i = 1, 2$. These findings provide evidence that the bivariate latent factor f_t is able to capture most of the long-range persistence displayed by the individual volatility series in Figure 2. Standard confidence bands on the sample autocorrelation function (ACF) are invalid since they do not account for the statistical error in estimating B and C when obtaining $\hat{\xi}_{jt}$ and $\hat{\eta}_{it}$. Therefore, for a formal specification test, we use the Hansen statistic introduced in Section 4.3. The statistic value is $J_1 = 73.64$, which is below the asymptotic critical value at the 5% level 119.87. This result corroborates the evidence of correct semiparametric specification of the measurement equation in our state-space model. Finally, the Hansen statistic value $J_2 = 216.75$ is below the critical value at 5% (equal to 233.99), suggesting correct specification of the VAR(6) dynamics for the bivariate latent factor process.

8. CONCLUDING REMARKS

In this paper, we study identification and statistical inference in a small-dimensional VAR model with dynamic unobservable factors. The novelty in our approach consists in the use of internal instrumental variables, which are (parameter-dependent) linear transformations of the observable variables and their lags, instead of large cross sections of data. The constructive identification approach leads to a multistep estimation procedure which does not require numerical optimization. We establish asymptotic normality of the estimators with a large number of serial observations T and finite number of series n. In some numerical examples, we show that the efficiency loss with respect to Gaussian QMLE is moderate, except for DGPs near the parameter values that are not identified (a zero-measure subset of the parameter space). We complete our theoretical analysis with consistent methods for selecting the number of latent factors and the VAR orders, and J-tests for correct model specification. In an empirical illustration, we use our state-space specification to disentangle dynamic common factors from spillover effects in the daily realized volatility series of four European stock markets.

Our analysis can be extended along several directions. First, we could combine our estimator and QMLE. Indeed, the estimates from our procedure can serve as consistent initial values for likelihood maximization. Moreover, the results in Section 3.3 imply that process ξ_t follows a state-space model (see (3.9)). Hence, we could use $\hat{\xi}_t$ and QMLE to estimate the parameters $\boldsymbol{\Phi}$ of the VAR factor dynamics. While requiring numerical optimization, compared to QMLE on the full model, it would have the advantage to reduce substantially the number of parameters. This estimator does not require the invertibility condition for $\Phi_{q,0}$ and might have better finite-sample properties compared to the estimator of $\boldsymbol{\Phi}$ studied in this paper. Second, we could extend the model specification to include lags of the latent factor, as well as external observable factors, in the measurement equation. Third, we could investigate the IRFs in our state-space model, by distinguishing the responses to idiosyncratic volatility shocks from responses to shocks in the latent common factor. These and other extensions are beyond the scope of this paper, and we leave them for future research.

APPENDIX A. FAVAR and VARMA Representations and Proofs of Identification

A.1. FAVAR Representation

Let $\boldsymbol{Y}_t = [Y_t', \ldots, Y_{t-p+1}']'$ and $\boldsymbol{f}_t = [f_t', \ldots, f_{t-q+1}']'$. Then, equations (2.1) and (2.2) can be written as a VAR(1) model:

$$\begin{bmatrix} \boldsymbol{Y}_t \\ \boldsymbol{f}_t \end{bmatrix} = \begin{bmatrix} \tilde{C} & \tilde{B}\tilde{\Phi} \\ 0 & \tilde{\Phi} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_{t-1} \\ \boldsymbol{f}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{w}_t \\ \boldsymbol{v}_t \end{bmatrix}, \tag{A.1}$$

where $\boldsymbol{w}_t = [(u_t + Bv_t)', 0', \ldots, 0']'$ and $\boldsymbol{v}_t = [v_t', 0', \ldots, 0']'$. The parameters $\tilde{C}$, $\tilde{B}$, and $\tilde{\Phi}$ are $np \times np$, $np \times Kq$, and $Kq \times Kq$ matrices, respectively, defined in companion form as

$$\tilde{C} = \begin{bmatrix} C_1 & \cdots & C_{p-1} & C_p \\ I_n & 0 & 0 & 0 \\ 0 & \ddots & \vdots & \vdots \\ 0 & \cdots & I_n & 0 \end{bmatrix}, \quad \tilde{\Phi} = \begin{bmatrix} \Phi_1 & \cdots & \Phi_{q-1} & \Phi_q \\ I_K & 0 & 0 & 0 \\ 0 & \ddots & \vdots & \vdots \\ 0 & \cdots & I_K & 0 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

The eigenvalues of matrices $\tilde{C}$ and $\tilde{\Phi}$ are the roots of polynomials $\det[z^p C(z^{-1})]$ and $\det[z^q \Phi(z^{-1})]$ with complex argument z. Hence, these eigenvalues are smaller than 1 in modulus under Assumptions M.1(iii) and M.2(ii).

A.2. VARMA Representation

The VARMA representation for observable process Y_t is given in the next lemma, which is proved in the Supplementary Material.

LEMMA 1. *The vector Y_t in the state-space model in equations (2.1) and (2.2) admits the VARMA(p + q,q) representation $A(L)Y_t = \Psi(L)w_t$ with $w_t \sim WN(0, \Sigma_w)$. The AR polynomial $A(L) = I_n - \sum_{j=1}^{p+q} A_j L^j$ has order $p+q$, and the matrix coefficients are $A_j = C_j + B\Phi_j \bar{B}' - \sum_{i=1}^{j-1} B\Phi_i \bar{B}' C_{j-i}$, $j = 1, \ldots, p+q$, where $\bar{B} = B(B'B)^{-1}$, $C_j \equiv 0$ for $j > p$, and $\Phi_j \equiv 0$ for $j > q$. The coefficients Ψ_j of the MA polynomial $\Psi(L) = I_n + \sum_{j=1}^{q} \Psi_j L^j$ and the variance Σ_w of the innovation satisfy*

$$\Sigma_w + \sum_{j=1}^{q} \Psi_j \Sigma_w \Psi_j' = B\Sigma_v B' + \Sigma_u + \sum_{j=1}^{q} B\Phi_j \bar{B}' \Sigma_u \bar{B} \Phi_j' B', \tag{A.2}$$

$$\Psi_i \Sigma_w + \sum_{j=i+1}^{q} \Psi_j \Sigma_w \Psi_{j-i}' = -B\Phi_i \bar{B}' \Sigma_u + \sum_{j=i+1}^{q} B\Phi_j \bar{B}' \Sigma_u \bar{B} \Phi_{j-i}' B', \quad i = 1, \ldots, q. \tag{A.3}$$

Moreover, under Assumptions M.1(iii) and M.2(ii), the roots of polynomial $\det A(z)$ *are outside the unit circle.*

A.3. Proof of Proposition 1

From measurement equation (2.1), we have $EL(Y_t | Y_{t-1}, \ldots, Y_{t-p-1}) = \sum_{j=1}^{p} C_{j,0} Y_{t-j} + B_0 EL(f_t | Y_{t-1}, \ldots, Y_{t-p-1}) = \sum_{j=1}^{p} (C_{j,0} + B_0 F_j) Y_{t-j} + B_0 F_{p+1} Y_{t-p-1}$. Thus,

$$A_{p+1}^* = B_0 F_{p+1}. \tag{A.4}$$

Then, under Assumption ID.1, matrix F_{p+1} has full rank and $F_{p+1} F_{p+1}'$ is nonsingular. Thus, we identify the column space of B_0 from the eigenvectors of matrix $(A_{p+1}^*)(A_{p+1}^*)' = B_0 (F_{p+1} F_{p+1}') B_0'$ associated with the K nonzero eigenvalues.

A.4. Proof of Proposition 2

Let us first prove Part (a). Let us write matrix $Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}$ in terms of the structural parameters. Matrix $Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} = \boldsymbol{E}[\boldsymbol{Y}_{t-1} \eta_{t-1}' : \cdots : \boldsymbol{Y}_{t-1} \eta_{t-M}']$ has blocks $\boldsymbol{E}[\boldsymbol{Y}_{t-1} \eta_{t-j}'] =$

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

$\boldsymbol{E}[\boldsymbol{Y}_{t-1}u'_{t-j}]B_{0\perp}$. From the FAVAR representation (A.1), we have $\boldsymbol{E}[\boldsymbol{Y}_{t-1}u'_{t-j}]B_{0\perp} = \tilde{C}_0^{j-1}\boldsymbol{E}[\boldsymbol{Y}_{t-j}u'_{t-j}]B_{0\perp} = \tilde{C}_0^{j-1}\boldsymbol{E}[\boldsymbol{w}_{t-j}u'_{t-j}]B_{0\perp} = \tilde{C}_0^{j-1}\tilde{\Sigma}_{u,0}B_{0\perp}$ for $j = 1,\ldots,M$, where $\tilde{\Sigma}_{u,0} = [\Sigma_{u,0}, 0_{n\times n},\ldots,0_{n\times n}]'$. By denoting $(A)_{[1]}$ the first n columns of an $np \times np$ matrix A, we get

$$
\begin{aligned}
Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} &= \left[(I_{np})_{[1]}\Sigma_{u,0}B_{0\perp} : (\tilde{C}_0)_{[1]}\Sigma_{u,0}B_{0\perp} : \cdots : (\tilde{C}_0^{M-1})_{[1]}\Sigma_{u,0}B_{0\perp}\right] \\
&= \left[(I_{np})_{[1]} : (\tilde{C}_0)_{[1]} : \cdots : (\tilde{C}_0^{M-1})_{[1]}\right]\left[I_M \otimes (\Sigma_{u,0}B_{0\perp})\right]. \qquad \textbf{(A.5)}
\end{aligned}
$$

We have $(\tilde{C}_0^{m})_{[1]} = \tilde{C}_0(\tilde{C}_0^{m-1})_{[1]}$, and the upper $n\times n$ block of matrix $(\tilde{C}_0^{m})_{[1]}$ is $\mathfrak{C}_m$, for any $m \geq 1$. Thus, we have $\left[(I_{np})_{[1]} : (\tilde{C}_0)_{[1]} : \cdots : (\tilde{C}_0^{M-1})_{[1]}\right] = [H : \mathcal{O}]$, where

$$
H = \begin{bmatrix}
I_n & \mathfrak{C}_1 & \mathfrak{C}_2 & \cdots & \mathfrak{C}_{p-1} \\
 & I_n & \mathfrak{C}_1 & \cdots & \mathfrak{C}_{p-2} \\
 & & I_n & \ddots & \vdots \\
 & & & \ddots & \mathfrak{C}_1 \\
 & & & & I_n
\end{bmatrix}, \quad
\mathcal{O} = \begin{bmatrix}
\mathfrak{C}_p & \mathfrak{C}_{p+1} & \mathfrak{C}_{p+2} & \cdots & \mathfrak{C}_{M-1} \\
\mathfrak{C}_{p-1} & \mathfrak{C}_p & \mathfrak{C}_{p+1} & \cdots & \mathfrak{C}_{M-2} \\
\vdots & \vdots & & & \vdots \\
\mathfrak{C}_2 & \mathfrak{C}_3 & & & \mathfrak{C}_{M-p+1} \\
\mathfrak{C}_1 & \mathfrak{C}_2 & \cdots & & \mathfrak{C}_{M-p}
\end{bmatrix}.
$$

The matrix coefficients $\mathfrak{C}_j$ satisfy the recursive relations $\mathfrak{C}_m = \sum_j C_j\mathfrak{C}_{m-j}$, $m \geq 1$, where the summation extends over $j = 1,2,\ldots$, with $\mathfrak{C}_0 = I_n$, $\mathfrak{C}_j = 0$, for $j < 0$, and $C_j = 0$, for $j > p$. Hence, from (A.5), the matrix $Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}$ written in terms of the structural parameters is $Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} = [H : \mathcal{O}]\left[I_M \otimes (\Sigma_{u,0}B_{0\perp})\right]$.

Let us now characterize the rank of matrix $Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}$. Let us left-multiply it by the nonsingular matrix $\left[I_p \otimes (\mathcal{Q}'\Sigma_{u,0}^{-1})\right]H^{-1}$, where $\mathcal{Q} = [B_0 : B_{0,\perp}]$. We get

$$
\begin{aligned}
\left[I_p \otimes (\mathcal{Q}'\Sigma_{u,0}^{-1})\right]H^{-1}Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} &= \left[I_p \otimes (\mathcal{Q}'\Sigma_{u,0}^{-1})\right]\left[I_{np} : H^{-1}\mathcal{O}\right]\left[I_M \otimes (\Sigma_{u,0}B_{0\perp})\right] \\
&= \left(I_p \otimes (\mathcal{Q}'B_{0,\perp}) : \left[I_p \otimes (\mathcal{Q}'\Sigma_{u,0}^{-1})\right]H^{-1}\mathcal{O}\left[I_{M-p} \otimes (\Sigma_{u,0}B_{0\perp})\right]\right).
\end{aligned}
$$

By writing this matrix explicitly, we have

$$
\begin{aligned}
&\left[I_p \otimes (\mathcal{Q}'\Sigma_{u,0}^{-1})\right]H^{-1}Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} \\
&= \begin{bmatrix}
0_{K\times(n-K)} & & & & B_0'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{11}\Sigma_{u,0}B_{0,\perp} & \cdots & B_0'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{1,M-p}\Sigma_{u,0}B_{0,\perp} \\
I_{n-K} & & & & B_{0,\perp}'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{11}\Sigma_{u,0}B_{0,\perp} & & B_{0,\perp}'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{1,M-p}\Sigma_{u,0}B_{0,\perp} \\
& 0_{K\times(n-K)} & & & B_0'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{21}\Sigma_{u,0}B_{0,\perp} & \cdots & B_0'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{2,M-p}\Sigma_{u,0}B_{0,\perp} \\
& I_{n-K} & & & B_{0,\perp}'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{21}\Sigma_{u,0}B_{0,\perp} & & B_{0,\perp}'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{2,M-p}\Sigma_{u,0}B_{0,\perp} \\
& & \ddots & & \vdots & & \vdots \\
& & & 0_{K\times(n-K)} & B_0'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{p,1}\Sigma_{u,0}B_{0,\perp} & \cdots & B_0'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{p,M-p}\Sigma_{u,0}B_{0,\perp} \\
& & & I_{n-K} & B_{0,\perp}'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{p,1}\Sigma_{u,0}B_{0,\perp} & & B_{0,\perp}'\Sigma_{u,0}^{-1}[H^{-1}\mathcal{O}]_{p,M-p}\Sigma_{u,0}B_{0,\perp}
\end{bmatrix},
\end{aligned}
$$

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

where $[A]_{i,j}$ denotes the $n \times n$ block in position (i,j) of the $np \times n(M-p)$ matrix A. We deduce that matrix $Q_{Y_{-1}Z}$ has full row rank if, and only if, the matrix

$$\Lambda = \begin{bmatrix} B_0' \Sigma_{u,0}^{-1} [H^{-1}\mathcal{O}]_{11} \Sigma_{u,0} B_{0,\perp} & \cdots & B_0' \Sigma_{u,0}^{-1} [H^{-1}\mathcal{O}]_{1,M-p} \Sigma_{u,0} B_{0,\perp} \\ B_0' \Sigma_{u,0}^{-1} [H^{-1}\mathcal{O}]_{21} \Sigma_{u,0} B_{0,\perp} & \cdots & B_0' \Sigma_{u,0}^{-1} [H^{-1}\mathcal{O}]_{2,M-p} \Sigma_{u,0} B_{0,\perp} \\ \vdots & & \vdots \\ B_0' \Sigma_{u,0}^{-1} [H^{-1}\mathcal{O}]_{p,1} \Sigma_{u,0} B_{0,\perp} & \cdots & B_0' \Sigma_{u,0}^{-1} [H^{-1}\mathcal{O}]_{p,M-p} \Sigma_{u,0} B_{0,\perp} \end{bmatrix} \qquad \textbf{(A.6)}$$

has full row rank. Let us write explicitly matrix $H^{-1}\mathcal{O}$. We have

$$H^{-1} = \begin{bmatrix} I_n & -C_1 & -C_2 & \cdots & -C_{p-1} \\ & I_n & -C_1 & \cdots & -C_{p-2} \\ & & \ddots & & \vdots \\ & & & I_n & -C_1 \\ & & & & I_n \end{bmatrix} \quad \text{and}$$

$$H^{-1}\mathcal{O} = \begin{bmatrix} G_{1,p} & G_{2,p} & \cdots & G_{M-p,p} \\ G_{1,p-1} & G_{2,p-1} & \cdots & G_{M-p,p-1} \\ \vdots & \vdots & & \vdots \\ G_{1,2} & G_{2,2} & & G_{M-p,2} \\ G_{1,1} & G_{2,1} & \cdots & G_{M-p,1} \end{bmatrix}, \qquad \textbf{(A.7)}$$

where $G_{m,i} = \mathfrak{C}_{m+i-1} - C_1 \mathfrak{C}_{m+i-2} - \cdots - C_{i-2}\mathfrak{C}_{m+1} - C_{i-1}\mathfrak{C}_m = \sum_{j\geq 1} C_j \mathfrak{C}_{m+i-1-j} - \sum_{j<i} C_j \mathfrak{C}_{m+i-1-j} = \sum_{j\geq i} C_j \mathfrak{C}_{m+i-1-j} = \sum_{j\geq 1} C_{j+i-1}\mathfrak{C}_{m-j}$. By plugging (A.7) into (A.6), transposing the matrix Λ and interchanging the order of the columns, the statement in Part (a) follows. Finally, the proof of Part (b) follows from equation (3.7) and Part (a).

A.5. Proof of Proposition 3

We have $Q_{W\xi_{-1}} = E(W_t \xi_{t-1}')$, where $W_t = (\xi_{t-q-1}', \ldots, \xi_{t-q-L}')'$ and $\boldsymbol{\xi}_{t-1}' = (\xi_{t-1}', \ldots, \xi_{t-q}')'$, with $\xi_t = \bar{B}_0'(Y_t - C_0 Y_{t-1}) = f_t + \bar{B}_0' u_t$. Thus, $Q_{W\xi_{-1}} = E(\boldsymbol{f}_{t-q-1:t-q-L}\boldsymbol{f}_{t-1}') = E(\boldsymbol{f}_{t-q-1:t-q-L}\boldsymbol{f}_{t-q-1}')(\tilde{\Phi}_0')^q$, where $\tilde{\Phi}_0$ is the matrix in the FAVAR representation in companion form (Appendix A.1) and $\boldsymbol{f}_{t-q-1:t-q-L} = (f_{t-q-1}', \ldots, f_{t-q-L}')'$. If $L \geq q$, the upper $qK \times qK$ block of matrix $Q_{W\xi_{-1}}$ is equal to $V(\boldsymbol{f}_{t-q-1})(\tilde{\Phi}_0')^q$. Therefore, we deduce that $Q_{W\xi_{-1}}$ is full column rank if, and only if, $\tilde{\Phi}_0$ is nonsingular. By using that $\tilde{\Phi}_0$ is nonsingular if and only if $\Phi_{q,0}$ is nonsingular, the statement in Part (a) follows. Part (b) follows from equation (3.8).

APPENDIX B. Proofs of Asymptotic Results

B.1. Consistency of Estimators $\hat{b}$, $\hat{c}$, $\hat{\phi}$: Proof of Theorem 2(a)

In the proof, we use repeatedly the consistency of sample autocovariances. The next lemma is proved in the Supplementary Material using Theorem 4.1.1 in Hannan and Deistler (1988).

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

LEMMA 2. *Under Assumptions M.1, M.2, and LS.1(i) and (iii), we have* $\frac{1}{T}\sum_{t=1}^{T} Y_t Y_{t-i}' \overset{a.s.}{\to} \Gamma(i)$, *as* $T \to \infty$, *for any* $i \geq 0$, *where* $\Gamma(i) = E(Y_t Y_{t-i}')$.

(i) Consistency of $\hat{b}$. Estimator $\hat{b} = vec(\hat{B}_1)$ is obtained from the eigendecomposition of matrix $\hat{R} = \hat{A}_{p+1}^* \hat{A}_{p+1}^{*\prime}$, which is an estimator of $R = A_{p+1}^* A_{p+1}^{*\prime}$. So let us first find the spectrum of matrix R. We recall that $A_{p+1}^* = B_0 F_{p+1}$ where F_{p+1} is full rank under Assumption ID.1 (see equation (A.4)). The normalized eigenvectors of matrix R associated with the K nonzero eigenvalues $\lambda_1, \dots, \lambda_K$ are the columns of the matrix

$$\boldsymbol{U} = [U_1 : \cdots : U_K] = B_0 Q, \tag{B.1}$$

where Q is the matrix such that $Q'(B_0' B_0) Q = I_K$. Under Assumption LS.1(iv), the nonzero eigenvalues are distinct, so that they are ranked as $\lambda_1 > \lambda_2 > \cdots > \lambda_K > 0$. Furthermore, matrix R has eigenvalue $\lambda_0 = 0$ with multiplicity $n - K$. The corresponding eigenspace is spanned by the columns of $B_{0\perp}$.

Now, $\hat{A}_{p+1}^* \overset{a.s.}{\to} A_{p+1}^*$, since $\hat{A}_{p+1}^*$ is a continuous function of sample autocovariances and the latter are strongly consistent from Lemma 2. Then, $\hat{R} \overset{a.s.}{\to} R$. The strong consistency of estimator $\hat{B}$ is proved by perturbation theory methods. Here, we use a version of Theorem 3 of Izenman (1975) adapted to prove almost sure convergence of the spectrum.

PROPOSITION 6. *Let R be an $n \times n$ symmetric positive-semidefinite matrix of rank $K \leq n$, and let its K strictly positive eigenvalues be distinct: $\lambda_1 > \cdots > \lambda_K > 0$. Let U_j, for $j = 1, \dots, K$, be the corresponding normalized eigenvectors. Let matrix $\hat{R}$ be a strongly consistent estimator of R, namely $\hat{R} \overset{a.s.}{\to} R$, and let $\hat{\lambda}_j$ and $\hat{U}_j$, for $j = 1, \dots, K$, be its K largest eigenvalues and associated normalized eigenvectors. Then, $\hat{\lambda}_j = \lambda_j + U_j'(\hat{R} - R)U_j + O_{a.s.}(|\hat{R} - R|^2)$ and*

$$\hat{U}_j = U_j + \sum_{i=0,\, i \neq j}^{K} \frac{1}{\lambda_j - \lambda_i} P_i (\hat{R} - R) U_j + O_{a.s.}(|\hat{R} - R|^2) \quad j = 1, \dots, K, \tag{B.2}$$

where matrix $P_i = U_i U_i'$ is the orthogonal projector onto the eigenspace associated with eigenvalue λ_i, for $i = 1, \dots, K$, matrix P_0 is the eigenprojector on the null space associated with eigenvalue $\lambda_0 = 0$, and $X = O_{a.s.}(1)$ means $X = O(1)$ a.s.

Under Assumptions M.1 and LS.1(i), (iii), and (iv), Proposition 6 applies in our case. Note that the sum in the r.h.s. of (B.2) extends also to include the zero eigenvalue $\lambda_0 = 0$ with eigenprojector given by $P_0 = B_{0\perp}(B_{0\perp}' B_{0\perp})^{-1} B_{0\perp}' = I_n - B_0 (B_0' B_0)^{-1} B_0'$. Given that $\hat{R} \overset{a.s.}{\to} R$, then $\hat{U}_j \overset{a.s.}{\to} U_j$, for $j = 1, \dots, K$. This means that $\hat{\boldsymbol{U}} = [\hat{U}_1 : \cdots : \hat{U}_K] = \left[\hat{\boldsymbol{U}}_1' : \hat{\boldsymbol{U}}_2'\right]'$ converges a.s. to $\boldsymbol{U} = B_0 Q = \left[(B_{1,0} Q)' : Q'\right]'$ (see equation (B.1)). It then follows that $\hat{B} = \hat{\boldsymbol{U}} \hat{\boldsymbol{U}}_2^{-1}$ converges a.s. to $B_0 = \left[B_{1,0}' : I_K\right]'$.

(ii) Consistency of $\hat{c}$. From equation (4.2), estimator $\hat{c}$ is function of matrices $\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}$, $\hat{Q}_{YZ}$ and $\hat{\Omega}_c$. The first two matrices depend on a set of estimated autocovariances of process $\{Y_t\}$, whose elements we gather in vector $\hat{\gamma}$ with true values γ_0, and on estimators $\hat{\boldsymbol{\Delta}}$ and $\hat{B}_\perp$. Thus, we have $\hat{c} = c(\hat{\gamma}, \hat{\boldsymbol{\Delta}}, \hat{B}_\perp, \hat{\Omega}_c)$, where $c(\cdot, \cdot, \cdot, \cdot)$ is a continuous function such that $c_0 = c(\gamma_0, \boldsymbol{\Delta}_0, B_{0\perp}, \Omega_c)$. Then, the strong consistency of estimator $\hat{c}$ follows, if we show that $\hat{\gamma}$, $\hat{B}_\perp$, and $\hat{\boldsymbol{\Delta}}$ are strongly consistent ($\hat{\Omega}_c \overset{a.s.}{\to} \Omega_c$ from Assumption LS.1(v)). We have $\hat{\gamma} \overset{a.s.}{\to} \gamma_0$ from Lemma 2. From $\hat{B}_\perp = [I_{n-K} : -\hat{B}_1]'$ and the strong consistency of $\hat{B}_1$ proved in Subsection (i), we get $\hat{B}_\perp \overset{a.s.}{\to} B_{0\perp}$. From equation $\hat{\boldsymbol{\Delta}} = \hat{B}'_\perp \left(\frac{1}{T}\sum_t Y_t \boldsymbol{Y}'_{t-1}\right)\left(\frac{1}{T}\sum_t \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}\right)^{-1}$, the strong consistency of $\hat{B}_\perp$, and Lemma 2, we get $\hat{\boldsymbol{\Delta}} \overset{a.s.}{\to} \boldsymbol{\Delta}_0$. Then, the conclusion follows.

(iii) Consistency of $\hat{\phi}$. From equation (4.3), we can write $\hat{\phi} = \phi(\hat{\gamma}, \hat{b}, \hat{c}, \hat{\Omega}_\phi)$, where $\phi(\cdot,\cdot,\cdot,\cdot)$ is a continuous function such that $\phi_0 = \phi(\gamma_0, b_0, c_0, \Omega_c)$ and $\hat{\gamma}$ denotes a vector whose elements are sample (cross-)autocovariances of components of process $\{Y_t\}$ with true value γ_0. By the strong consistency of $\hat{\gamma}$ in Lemma 2, that of estimators $\hat{b}$ and $\hat{c}$ proved in Subsections (i) and (ii), respectively, and the convergence $\hat{\Omega}_\phi \overset{a.s.}{\to} \Omega_\phi$ from Assumption LS.1(v), we get $\hat{\phi} \overset{a.s.}{\to} \phi_0$.

B.2. Asymptotic Expansions for Estimators $\hat{b}$, $\hat{c}$, $\hat{\phi}$: Proof of Theorem 2(b)

Let us first provide the asymptotic expansion of estimator $\hat{b} = vec(\hat{B}_1)$.

PROPOSITION 7. *Under Assumptions M.1, IR.1, ID.1–ID.3, and LS.1(i), (iii), and (iv), the asymptotic expansion of estimator $\hat{b} = vec(\hat{B}_1)$ is*

$$\sqrt{T}(\hat{b} - b_0) = \left\{Q \otimes B'_{0\perp}\right\}\sqrt{T}\left[\ (\hat{U}_1 - U_1)' : \quad \cdots : (\hat{U}_K - U_K)' \ \right]' + o_p(1), \tag{B.3}$$

where

$$\sqrt{T}(\hat{U}_j - U_j) = \sum_{i=0,\, i\neq j}^{K} \frac{1}{\lambda_j - \lambda_i}(U'_j \otimes P_i)\sqrt{T} vec(\hat{R} - R) + o_p(1), \quad j = 1, \ldots, K,$$

$$\sqrt{T} vec(\hat{R} - R) = (I_{n^2} + \mathcal{K}_{n,n})(A^*_{p+1} \otimes I_n)\sqrt{T} vec(\hat{A}^*_{p+1} - A^*_{p+1}) + o_p(1),$$

$$\sqrt{T} vec(\hat{A}^*_{p+1} - A^*_{p+1}) = \left[\left(\boldsymbol{E}(\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1})\right)^{-1} \otimes I_n\right]\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{Y}_{t-p-1} \otimes u^*_t)\right) + o_p(1). \tag{B.4}$$

Matrix $\mathcal{K}_{m,n}$ is the commutation matrix such that $vec(M') = \mathcal{K}_{m,n} vec(M)$ for an $m \times n$ matrix M and the vector $\tilde{Y}_{t-p-1} = Y_{t-p-1} - EL(Y_{t-p-1}|Y_{t-1}, \ldots, Y_{t-p}) = Y_{t-p-1} - \tilde{\boldsymbol{\Gamma}}(p)'\boldsymbol{\Gamma}(0)^{-1}\boldsymbol{Y}_{t-1}$ is the population residual of the regression of Y_{t-p-1} onto $Y_{t-1}, \ldots, Y_{t-p}$, and $\boldsymbol{E}(\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1}) = \Gamma(0) - \tilde{\boldsymbol{\Gamma}}(p)'\boldsymbol{\Gamma}(0)^{-1}\tilde{\boldsymbol{\Gamma}}(p)$, with $\boldsymbol{\Gamma}(0) = \boldsymbol{E}(\boldsymbol{Y}_t\boldsymbol{Y}'_t)$ and $\tilde{\boldsymbol{\Gamma}}(p) = \boldsymbol{E}(\boldsymbol{Y}_t Y'_{t-p})$.

In Proposition 7, the asymptotic expansion of estimator $\hat{b}$ follows from that of the eigenvectors of matrix $\hat{R}$ using Proposition 6. In turn, the asymptotic expansion of $\hat{R}$ follows from that of the OLS estimator $\hat{A}^*_{p+1}$ on the VAR($p+1$) pseudo-model.

Let us now derive the asymptotic expansion for estimator $\hat{d} = vec(\hat{\boldsymbol{\Delta}})$ of parameter vector $d_0 = vec(\boldsymbol{\Delta}_0)$.

PROPOSITION 8. *Under Assumptions M.1, IR.1, ID.1–ID.3, and LS.1(i), (iii), and (iv), the asymptotic expansion for estimator $\hat{d}$ is given by*

$$\sqrt{T}(\hat{d} - d_0) = \left[\boldsymbol{\Gamma}(0)^{-1} \otimes I_{n-K}\right] \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \otimes \eta_t - \left[\boldsymbol{\Gamma}(0)^{-1} \tilde{\boldsymbol{\Gamma}}_{[K]}(-1) \otimes I_{n-K}\right] \sqrt{T}(\hat{b} - b_0) + o_p(1),$$

where $\tilde{\boldsymbol{\Gamma}}_{[K]}(j) = \boldsymbol{E}[\boldsymbol{Y}_t Y'_{[K],t-j}]$ and $Y_{[K],t} = (Y_{n-K+1,t}, \ldots, Y_{n,t})'$.

We give the asymptotic expansions for estimators $\hat{c} = vec(\hat{\boldsymbol{C}})$ and $\hat{\phi} = vec(\hat{\Phi})$ in the next propositions.

PROPOSITION 9. *Under Assumptions M.1, IR.1, ID.1–ID.3, and LS.1(i), (iii), and (v), estimator $\hat{c} = vec(\hat{\boldsymbol{C}})$ admits the asymptotic expansion*

$$\sqrt{T}(\hat{c} - c_0) = -(J'_c \Omega_c J_c)^{-1} J'_c \Omega_c \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h_t + J_d \sqrt{T}(\hat{d} - d_0) + J_b \sqrt{T}(\hat{b} - b_0) \right\} + o_p(1), \tag{B.5}$$

where $h_t = vec\left[(Y_t - \boldsymbol{C}_0 \boldsymbol{Y}_{t-1}) \boldsymbol{Z}'_t\right] = \boldsymbol{Z}_t \otimes \varepsilon_t$ and $J_c = -Q'_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} \otimes I_n$,

$$J_d = -(I_M \otimes \mathcal{K}_{n-K,n}) \left[\begin{pmatrix} \boldsymbol{E}(\varepsilon_t \boldsymbol{Y}'_{t-2}) \\ \vdots \\ \boldsymbol{E}(\varepsilon_t \boldsymbol{Y}'_{t-M-1}) \end{pmatrix} \otimes I_{n-K} \right],$$

$$J_b = -(I_M \otimes \mathcal{K}_{n-K,n}) \left[\begin{pmatrix} \boldsymbol{E}(\varepsilon_t Y'_{[K],t-1}) \\ \vdots \\ \boldsymbol{E}(\varepsilon_t Y'_{[K],t-M}) \end{pmatrix} \otimes I_{n-K} \right],$$

where $Q_{\boldsymbol{Y}_{-1}\boldsymbol{Z}} = \boldsymbol{E}(\boldsymbol{Y}_{t-1}\boldsymbol{Z}_t)$, $\mathcal{K}_{n-K,n}$ is the commutation matrix for orders $n-K, n$, and $\varepsilon_t = Y_t - \boldsymbol{C}_0 \boldsymbol{Y}_{t-1}$.

PROPOSITION 10. *Under Assumptions M.1, M.2, IR.1, ID.1–ID.3, and LS.1(i), (iii), and (v), the asymptotic expansion for estimator $\hat{\phi}$ is*

$$\sqrt{T}(\hat{\phi} - \phi_0) = -(\mathscr{J}'_\phi \Omega_\phi \mathscr{J}_\phi)^{-1} \mathscr{J}'_\phi \Omega_\phi \times \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} l_t + \mathscr{J}_b \sqrt{T}(\hat{b} - b_0) + \mathscr{J}_c \sqrt{T}(\hat{c} - c_0) \right\} + o_p(1), \tag{B.6}$$

where $l_t = vec\left[(\xi_t - \mathbf{\Phi}_0 \boldsymbol{\xi}_{t-1})\boldsymbol{W}_t'\right] = \boldsymbol{W}_t \otimes e_t$, $\boldsymbol{W}_t = (\xi_{t-q-1}', \ldots, \xi_{t-q-L}')'$, *and* $\mathscr{J}_\phi = -Q_{\boldsymbol{W}\boldsymbol{\xi}_{-1}} \otimes I_K$,

$$\mathscr{J}_b = \left\{ \begin{bmatrix} \boldsymbol{E}(e_t \varepsilon_{t-q-1}') \\ \vdots \\ \boldsymbol{E}(e_t \varepsilon_{t-q-L}') \end{bmatrix} \otimes I_K + \boldsymbol{E}(\boldsymbol{W}_t \varepsilon_t') \otimes I_K - \sum_{i=1}^{q} \boldsymbol{E}(\boldsymbol{W}_t \varepsilon_{t-i}') \otimes \Phi_{i,0} \right\} D_0,$$

$$\mathscr{J}_c = -\begin{bmatrix} \boldsymbol{E}(e_t \boldsymbol{Y}_{t-q-2}') \\ \vdots \\ \boldsymbol{E}(e_t \boldsymbol{Y}_{t-q-L-1}') \end{bmatrix} \otimes \bar{B}_0' - \boldsymbol{E}(\boldsymbol{W}_t \boldsymbol{Y}_{t-1}') \otimes \bar{B}_0' + \sum_{i=1}^{q} \boldsymbol{E}(\boldsymbol{W}_t \boldsymbol{Y}_{t-i-1}') \otimes (\Phi_{i,0} \bar{B}_0'),$$

$e_t := \xi_t - \mathbf{\Phi}_0 \boldsymbol{\xi}_{t-1}$, *and* $D_0 = \partial vec(\bar{B}_0)/\partial b_0' = \left(P_{0,[n-K]} \otimes (B_0' B_0)^{-1}\right) \mathcal{K}_{n-K,K} - \bar{B}_0 \otimes \left((B_0' B_0)^{-1} B_{1,0}'\right)$ *and* $P_{0,[n-K]}$ *denotes the first* $n-K$ *columns of matrix* P_0.

By combining the results in Propositions 7–10, we get the asymptotic expansions in (4.4) where the blocks of matrix S are given by

$$\begin{aligned} S_{b1} &= \left[Q \otimes B_{0\perp}'\right] \left[\sum_{i=0, i\neq 1}^{K} \frac{1}{\lambda_1 - \lambda_i} (U_1 \otimes P_i) : \cdots : \sum_{i=0, i\neq K}^{K} \frac{1}{\lambda_K - \lambda_i} (U_K \otimes P_i) \right]' \\ &\quad \times (I_{n^2} + \mathcal{K}_{n,n})(A_{p+1}^* \otimes I_n) \left[\left(\Gamma(0) - \tilde{\mathbf{\Gamma}}(p)' \mathbf{\Gamma}(0)^{-1} \tilde{\mathbf{\Gamma}}(p) \right)^{-1} \otimes I_n \right], \\ \left[S_{c1} : S_{c2} : S_{c3}\right] &= -(J_c' \Omega_c J_c)^{-1} J_c' \Omega_c \left[\left(J_b - J_d \left(\mathbf{\Gamma}(0)^{-1} \tilde{\mathbf{\Gamma}}_{[K]}(-1) \otimes I_{n-K} \right) \right) \right. \\ &\qquad \left. \times S_{b1} : J_d \left(\mathbf{\Gamma}(0)^{-1} \otimes I_{n-K} \right) : I_{n(n-K)M} \right], \\ \left[S_{\phi 1} : S_{\phi 2} : S_{\phi 3} : S_{\phi 4}\right] &= -(\mathscr{J}_\phi' \Omega_\phi \mathscr{J}_\phi)^{-1} \mathscr{J}_\phi' \Omega_\phi \\ &\qquad \times \left[\left(\mathscr{J}_b S_{b1} + \mathscr{J}_c S_{c1} \right) : \mathscr{J}_c S_{c2} : \mathscr{J}_c S_{c3} : I_{K^2 L} \right], \end{aligned} \tag{B.7}$$

and all matrices are defined in Propositions 7–10.

B.3. Asymptotic Normality of Estimators $\hat{b}, \hat{c}, \hat{\phi}$: Proof of Theorem 2(c)

The asymptotic normality of vector $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_t$ follows from a multivariate version of Corollary 1 in Herrndorf (1984) and the next lemma proved in the Supplementary Material.

LEMMA 3. *Under Assumptions M.1, M.2, and LS.1(ii) and (iii), we have: (a)* $\boldsymbol{E}(|\psi_t|^\beta) < \infty$, *for* $\beta > 2$; *(b) the process* ψ_t *is* α*-mixing, and the* α*-mixing coefficients are such that* $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\beta-2}{\beta}} < \infty$; *and (c)* $\frac{1}{T} V\left(\sum_{t=1}^{T} \psi_t\right) \to V_\psi$ *as* $T \to \infty$, *where* $V_\psi = \sum_{j=-\infty}^{\infty} Cov(\psi_t, \psi_{t-j})$.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

The asymptotic expansions in Theorem 2(b), and the distributional convergence of vector $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t$, imply that the estimators $\hat{b}$, $\hat{c}$, and $\hat{\phi}$ are asymptotically normal and yield Theorem 2(c).

B.4. Optimal Weighting Matrix and Specification Test

Let us first derive the optimal weighting matrix for the estimation of $\boldsymbol{C}$. From Proposition 9, we have

$$\sqrt{T}(\hat{c}-c_0) = -(J_c'\Omega_c J_c)^{-1}J_c'\Omega_c D_c \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t, \quad \text{where} \tag{B.8}$$

$$D_c = \Big[\Big(J_b - J_d\Big(\boldsymbol{\Gamma}(0)^{-1}\tilde{\boldsymbol{\Gamma}}_{[K]}(-1)\otimes I_{n-K}\Big)\Big) \times S_{b1} \;:\; J_d\Big[\boldsymbol{\Gamma}(0)^{-1}\otimes I_{n-K}\Big] \;:\; I_{n(n-K)M} \;:\; 0_{n(n-K)M\times K^2 S}\Big]. \tag{B.9}$$

Moreover, from Section B.3, vector $D_c\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t$ is asymptotically normal with asymptotic variance $D_c V_\psi D_c'$. From standard theory of GMM, we deduce that the optimal weighting matrix of estimator $\hat{c}$ is

$$\Omega_c^* = (D_c V_\psi D_c')^{-1}. \tag{B.10}$$

Let us now show the asymptotic distribution of the Hansen statistic under the null hypothesis of correct specification of the orthogonality restrictions for identification of $\boldsymbol{C}$. The Hansen statistic is $J_1 = T\left(\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\right)'\hat{\Omega}_c^*\left(\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\right)$, where $\hat{h}_t = \mathrm{vec}[(Y_t - \hat{\boldsymbol{C}}\boldsymbol{Y}_{t-1})\hat{\boldsymbol{Z}}_t']$ and $\hat{\Omega}_c^* = (\hat{D}_c\hat{V}_\psi\hat{D}_c')^{-1}$ is a consistent estimator of Ω_c^*. By standard arguments as in the proof of Proposition 9, we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{h}_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_t + J_c\sqrt{T}(\hat{c}-c_0) + J_d\sqrt{T}(\hat{d}-d_0) + J_b\sqrt{T}(\hat{b}-b_0) + o_p(1) = D_c\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t + J_c\sqrt{T}(\hat{c}-c_0) + o_p(1)$. By plugging the asymptotic expansion (B.8) with optimal weighting matrix Ω_c^*, we get $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{h}_t = \left[I_{n(n-K)M} - J_c(J_c'\Omega_c^*J_c)^{-1}J_c'\Omega_c^*\right]D_c\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t + o_p(1)$. Since $D_c\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t$ is asymptotically normal with asymptotic variance $(\Omega_c^*)^{-1}$, from standard arguments in GMM theory, we get that J_1 is asymptotically distributed as $\chi^2[n(n-K)M - n^2 p]$. Finally, by using that $\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t = \mathrm{vec}[\hat{Q}_{Y\boldsymbol{Z}} - \hat{\boldsymbol{C}}\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}]$, the expression for the J-statistic in Section 4.3 follows.

Given the optimal choice Ω_c^* for estimating $\boldsymbol{C}$, the optimal weighting matrix for estimating $\boldsymbol{\Phi}$ is the one minimizing the asymptotic variance of $\hat{\phi}$. We have $\sqrt{T}(\hat{\phi}-\phi_0) = -(\mathscr{J}_\phi'\Omega_\phi\mathscr{J}_\phi)^{-1}\mathscr{J}_\phi'\Omega_\phi D_\phi\frac{1}{\sqrt{T}}\sum_t\psi_t$ from Proposition 10, where $D_\phi = \left[\mathscr{J}_b S_{b1} + \mathscr{J}_c S_{c1} : \mathscr{J}_c S_{c2} : \mathscr{J}_c S_{c3} : I_{K^2 S}\right]$, and the $S_{c,i}$, for $i = 1,2,3$, are evaluated with Ω_c^*. Then, the optimal weighting matrix for $\hat{\phi}$ is

$$\Omega_\phi^* = (D_\phi V_\psi D_\phi')^{-1}. \tag{B.11}$$

The Hansen statistic for testing the validity of the moment restrictions (3.8) is $J_2 = T\left(\frac{1}{T}\sum_t\hat{l}_t\right)'\hat{\Omega}_\phi^*\left(\frac{1}{T}\sum_t\hat{l}_t\right)$, where $\hat{l}_t = vec[(\hat{\xi}_t - \hat{\boldsymbol{\Phi}}\hat{\boldsymbol{\xi}}_{t-1})\hat{\boldsymbol{W}}_t']$ and $\hat{\Omega}_\phi^* = (\hat{D}_\phi\hat{V}_\psi\hat{D}_\phi')^{-1}$. Under the null hypothesis, J_2 admits a $\chi^2[K^2(S-p)]$ distribution.

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

B.5. Proof of Theorem 3

We have $\hat{S} \overset{a.s.}{\to} S$ because function $S(\gamma,\theta)$ is continuous w.r.t. γ and θ, $\hat{\gamma} \overset{a.s.}{\to} \gamma_0$ from Lemma 2, and $\hat{\theta} \overset{a.s.}{\to} \theta_0$ from Theorem 2(a). We show $\hat{V}_\psi \overset{p}{\to} V_\psi$ using the results in De Jong and Davidson (2000). Let us check the regularity conditions. We can write $\psi_t = \psi_t(\gamma_0,\theta_0) = \Psi_1(\gamma_0,\theta_0)X_t$, where vectors X_t and γ_0 have elements like $Y_{i,t}Y_{j,t-h}$ (resp. $\boldsymbol{E}[Y_{i,t}Y_{j,t-h}]$), for $i,j = 1,\ldots,n$ and $|h| \leq h^*$, vector θ_0 contains elements of matrices $\bar{B}_0$, $\boldsymbol{C}_0$, and $\boldsymbol{\Phi}_0$, and matrix function $\Psi_1(\cdot)$ is differentiable. Process $\{X_t\}$ is geometrically strong mixing and $E[|\psi_t|^\beta] < \infty$ for $\beta > 2$ by the arguments in Appendix B.3, and similarly $\sup_{t\geq 1} \boldsymbol{E}[\sup_{\theta,\gamma\in\tilde{\Theta}} |\frac{\partial \psi_t(\gamma,\theta)}{\partial(\gamma',\theta')'}|^2] < \infty$. Estimators $\hat{\gamma}$ and $\hat{\theta}$ are root-T consistent by Theorem 2. Finally, $\kappa(\cdot)$ and m_T satisfy the regularity conditions in Assumption LS.2. Then, the results in De Jong and Davidson (2000) imply that $\hat{V}_\psi \overset{p}{\to} V_\psi$. Moreover, matrix $\hat{V}_\psi$ is positive definite a.s. because of Assumption LS.2(iv).

B.6. Proof of Theorem 4

Let us first prove Part (a). When the rank-K approximation of $\hat{A}^*_{p+1}$ is based on SVD, the result follows from Theorem 1 and Corollary 1 in Kleibergen and Paap (2006). Indeed, from (B.4), we have $\sqrt{T} vec(\hat{A}^*_{p+1} - A^*_{p+1}) \overset{p}{\to} N(0,\Sigma_{11})$, which yields their Assumption 1. Note that their Assumption 2 is equivalent to $(P_{M_K} \otimes P_{N_K})\Sigma_{11}(P_{M_K} \otimes P_{N_K})$ being of rank $(n-K)^2$, which is satisfied in our framework since matrix Σ_{11} is positive definite. Moreover, the HAC estimator $\hat{\Sigma}_{11}$ is consistent for Σ_{11} under Theorem 3.

In the case of other reduced-rank approximations, we use Theorem 1 and Corollary 1 in Al-Sadoon (2017). We keep as maintained hypothesis that the rank-r approximation of $\hat{A}^*_{p+1}$ to obtain left and right null-space projections $P_{\hat{M}_r}$ and $P_{\hat{N}_r}$ for $r \leq K$, and of A^*_{p+1} to obtain P_{M_r} and P_{N_r} for $r < K$, are either decomposition-based approximations (DBA) or Cragg Donald approximations (CDA) in the sense of Definitions 2 and 3 of Al-Sadoon (2017). Let us write the test statistic as $\mathscr{F}(r) = T \cdot \tau(\hat{A}^*_{p+1}, \hat{\Sigma}_{11}, P_{\hat{N}_r}, P_{\hat{M}_r})$ where $\tau(\hat{A}^*_{p+1}, \hat{\Sigma}_{11}, P_{\hat{N}_r}, P_{\hat{M}_r}) = vec(P_{\hat{N}_r}\hat{A}^*_{p+1}P_{\hat{M}_r})'\left[(P_{\hat{M}_r} \otimes P_{\hat{N}_r})\hat{\Sigma}_{11}(P_{\hat{M}_r} \otimes P_{\hat{N}_r})\right]^\dagger vec(P_{\hat{N}_r}\hat{A}^*_{p+1}P_{\hat{M}_r})$, and let $\tilde{\mathscr{F}}(r) = T \cdot \tau(\hat{A}^*_{p+1}, \hat{\Sigma}_{11}, P_{N_r}, P_{M_r})$ be the infeasible version of the statistic based on the null-space projections associated with the true matrix value A^*_{p+1}. Assumption A in Al-Sadoon (2017) is satisfied by estimators $\hat{A}^*_{p+1}$ and $\hat{\Sigma}_{11}$ under the conditions of Theorems 2 and 3. The functional form of τ satisfies Assumption K in Al-Sadoon (2017). Then, by Theorem 1 in Al-Sadoon (2017) statistic $\mathscr{F}(r)$ satisfies the (weak) plug-in principle:

$$\mathscr{F}(r) = \tilde{\mathscr{F}}(r) + O_p(T^{-1/2}), \qquad \text{if } r = K, \tag{B.12}$$

$$\tau(\hat{A}^*_{p+1}, \hat{\Sigma}_{11}, P_{N_r}, P_{M_r})^{-1} = O_p(1) \Rightarrow \tau(\hat{A}^*_{p+1}, \hat{\Sigma}_{11}, P_{\hat{N}_r}, P_{\hat{M}_r})^{-1} = O_p(1), \qquad \text{if } r < K, \tag{B.13}$$

that is, the feasible and infeasible statistics are asymptotically equivalent at order $O_p(T^{-1/2})$ under the null hypothesis, and if the unscaled infeasible statistic is bounded away from zero under the alternative hypothesis so is the unscaled feasible statistic. Then, along the lines of Corollary 1 in Al-Sadoon (2017), under the null hypothesis using (B.12) and

$\tilde{\mathscr{F}}(K) = T \cdot vec(P_{N_r}\hat{A}^*_{p+1}P_{M_r})'\left[(P_{M_r}\otimes P_{N_r})\hat{\Sigma}_{11}(P_{M_r}\otimes P_{N_r})\right]^{\dagger} vec(P_{N_r}\hat{A}^*_{p+1}P_{M_r}) \xrightarrow{d} \chi^2[(n-K)^2]$we get $\mathscr{F}(K) \xrightarrow{d} \chi^2[(n-K)^2]$, which yields Part (a).

Let us now prove the consistency of the test in Part (b) of Theorem 4. Under the alternative hypothesis ($r < K$), vector $vec(P_{N_r}\hat{A}^*_{p+1}P_{M_r})$ converges in probability to a nonnull vector and $\tau(\hat{A}^*_{p+1}, \hat{\Sigma}_{11}, P_{N_r}, P_{M_r})$ converges in probability to a strictly positive constant. Then, from the plug-in principle (B.13), the statistic $\mathscr{F}(r)$ diverges in probability at rate T, which yields Part (b). Finally, Part (c) follows by paralleling the argument in the proof of Theorem 5.2 in Robin and Smith (2000).

B.7. Proof of Theorem 5

We start with Part (a). Under the null hypothesis, $p = p_0$. Then, $vec(B'_{0\perp}\boldsymbol{A}^\star_0) = 0$, where $\boldsymbol{A}^\star_0 := [A^\star_{p+2} : \cdots : A^\star_{p^\star}]$. Similarly, $\hat{\boldsymbol{A}}^\star := [\hat{A}^\star_{p+2} : \cdots : \hat{A}^\star_{p^\star}]$. By using the asymptotic expansions of the estimators, in the Supplementary Material, we show the next lemma.

LEMMA 4. *Under the Assumptions of Theorem 2, we have* $\sqrt{T}vec(\hat{B}_\perp(p)'\hat{\boldsymbol{A}}^\star) \xrightarrow{d} N(0, \Sigma_{22})$ *as* $T \to \infty$, *where* $\Sigma_{22} = \mathcal{M}V_{\psi,22}\mathcal{M}'$ *with* $V_{\psi,22} = \sum_{j=-\infty}^{\infty} Cov(\psi_{2,t}, \psi_{2,t-j})$ *for* $\psi_{2,t} = [(\tilde{Y}_{t-p-1}\otimes u^*_t)', (\tilde{X}_{t-p-2}\otimes u^\star_t)']'$, $\tilde{X}_{t-p-2} = X_{t-p-2} - EL(X_{t-p-2}|Y_{t-1}, \ldots, Y_{t-p-1})$, $X_{t-p-2} = (Y'_{t-p-2}, \ldots, Y'_{t-p^\star})'$ *and* $\mathcal{M} = [\mathcal{M}_1 : \mathcal{M}_2]$ *for* $\mathcal{M}_1 = -(\tilde{\boldsymbol{A}}^{\star\prime}_0 \otimes I_{n-K_0})S_{b1}$ *and* $\mathcal{M}_2 = (I_{(p^\star-p-1)n}\otimes B'_{0\perp})\{\boldsymbol{E}(\tilde{X}_{t-p-2}\tilde{X}'_{t-p-2})^{-1}\otimes I_n\}$, *and* $\tilde{\boldsymbol{A}}^\star_0 = [A^\star_{p+2,0,[K_0]} : \cdots : A^\star_{p^\star,0,[K_0]}]$ *and* $A^\star_{i,0,[K_0]}$ *is the matrix with the last* K_0 *columns of* $A^\star_{i,0}$.

Let

$$\hat{\Sigma}_{22} = \hat{\mathcal{M}}\hat{V}_{\psi,22}\hat{\mathcal{M}}' \tag{B.14}$$

be a consistent estimator of Σ_{22}, where $\hat{V}_{\psi,22}$ is an HAC estimator of $V_{\psi,22}$, and $\hat{\mathcal{M}}$ is obtained by replacing unknown quantities in $\mathcal{M}$ with the sample analogs, as in the proof of Theorem 3. Then, under the null hypothesis, the asymptotic distribution of the test statistic is given by $Tvec(\hat{B}_\perp(p)'\hat{\boldsymbol{A}}^\star)'\hat{\Sigma}^{-1}_{22}vec(\hat{B}_\perp(p)'\hat{\boldsymbol{A}}^\star) \xrightarrow{d} \chi^2[(n-K_0)n(p^\star-p-1)]$ as $T \to \infty$.

Let us now prove Part (b). Under the alternative hypothesis, $p < p_0$. First, we note that $\hat{K}(p) = K(p)$ w.p.a. 1 as $T \to \infty$, where $K(p)$ is the rank of matrix A^*_{p+1}. Indeed, the arguments in the proof of Theorem 4 applies also to matrix A^*_{p+1} for $p < p_0$. If matrix $A^\star_{p+1}$ is nonsingular, then w.p.a. 1 we have $\hat{K}(p) = n$ and $\mathscr{S}_p = +\infty$ by definition. If matrix $A^\star_{p+1}$ is singular, with rank $K(p) < n$, let us denote $B_\perp(p)$ the $n\times(n-K(p))$ full-rank matrix, whose columns span the orthogonal complement of the range $\mathscr{R}(A^\star_{p+1})$ with the normalization in Section 3. Then, we have $\mathscr{S}_p \geq CT$, w.p.a. 1, for a constant $C > 0$ and T large, if $B_\perp(p)'\boldsymbol{A}^\star_0(p) \neq 0$, where $\boldsymbol{A}^\star_0(p) = [A^\star_{p+2} : \cdots : A^\star_{p+2}]$. Since $A^\star_{p+1} = C_{p+1,0} + B_0F^{(p^\star)}_{p+1}$, and one of the blocks in $\boldsymbol{A}^\star_0$ is $A^\star_{p_0+1} = B_0F^{(p^\star)}_{p_0+1}$, a sufficient condition for $B_\perp(p)'\boldsymbol{A}^\star_0(0) \neq 0$ to hold is that the kernel of matrix $(C_{p+1,0} + B_0F^{(p^\star)}_{p+1})'$ is not a subset of the kernel of B'_0. Since under the alternative hypothesis the possible values

https://doi.org/10.1017/S0266466622000536 Published online by Cambridge University Press

of p are $p = 0, 1, \ldots, p_0 - 1$, we deduce that a sufficient condition for consistency of the test statistic is the following: the kernel of matrix $(C_{i,0} + B_0 F_i^{(p^\star)})'$ is either (i) empty, or (ii) not a subset of the kernel of B_0', for any $i = 1, \ldots, p_0$. By using $(C_{i,0} + B_0 F_i^{(p^\star)})'[\bar{B}_0 \ : \ \bar{B}_{0\perp}] = [C_{0,i}' \bar{B}_0 + F_i^{(p^\star)\prime} \ : \ C_{i,0}' B_{0\perp}]$ and $B_0'[\bar{B}_0 \ : \ \bar{B}_{0\perp}] = [I_K \ : \ 0_{K \times (n-K)}]$, the above condition for consistency is equivalent to: the kernel of matrix $[C_{0,i}' \bar{B}_0 + F_i^{(p^\star)\prime} \ : \ C_{i,0}' B_{0\perp}]$ either (i) is empty, or (ii) has a nontrivial projection on the linear space $\{(v', 0'_{(n-K_0)\times 1})' : v \in \mathbb{R}^{K_0}\}$, for any $i = 1, \ldots, p_0$. This is the condition in Assumption SEL.1. Finally, the proof of Part (c) parallels that of Theorem 5.2 in Robin and Smith (2000) with some adjustments for the randomness of $\hat{K}(n)$; it is provided in the the Supplementary Material for completeness.

SUPPLEMENTARY MATERIAL

Carlini, F. and Gagliardini, P. (2022). Supplement to "Instrumental variables inference in a small-dimensional VAR model with dynamic latent factors," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466622000536.

REFERENCES

Acemoglu, D., A. Ozdaglar, & A. Tahbaz-Salehi (2015) Systemic risk and stability in financial networks. *American Economic Review* 105(2), 564–608.

Acharya, V., L. Pedersen, T. Philippon, & M. Richardson (2017) Measuring systemic risk. *Review of Financial Studies* 30, 2–47.

Adrian, T. & M. Brunnermeier (2016) CoVaR. *American Economic Review* 106, 1705–1741.

Ait-Sahalia, Y., J. Cacho-Diaz, & R. Laeven (2015) Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics* 117, 585–606.

Akaike, H. (1974) Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics* 26, 363–387.

Allen, F. & D. Gale (2000) Financial contagion. *Journal of Political Economy* 108(1), 1–33.

Al-Sadoon, M.M. (2017) A unifying theory of tests of rank. *Journal of Econometrics* 199(1), 49–62.

Anderson, T.W. (1951a) The asymptotic distribution of certain characteristic roots and vectors. In J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, pp. 103–130.

Anderson, T.W. (1951b) Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327–351.

Andrews, D. & C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

Bai, J., K. Li, & L. Lu (2016) Estimation and inference of FAVAR models. *Journal of Business & Economic Statistics* 34(4), 620–641.

Bernanke, B., J. Boivin, & P. Eliasz (2005) Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120, 387–422.

Billio, M., M. Getmansky, A.W. Lo, & L. Pelizzon (2012) Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104(3), 535–559.

Brownlees, C. & R. Engle (2017) SRISK: A conditional capital shortfall measure of systemic risk. *Review of Financial Studies* 30, 48–79.

Chan, J., E. Eisenstat, & G. Koop (2016) Large Bayesian VARMAs. *Journal of Econometrics* 192, 374–390.

Chan, J., R. Leon-Gonzalez, & R.W. Strachan (2018) Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association* 113(522), 819–828.

Chen, Q. & Z. Fang (2019) Improved inference on the rank of a matrix. *Quantitative Economics* 10(4), 1787–1824.

Cragg, J. & S. Donald (1996) On the asymptotic properties of LDU-based tests of the rank of a matrix. *Journal of the American Statistical Association* 91, 1301–1309.

Darolles, S., S. Dubecq, & C. Gouriéroux (2014). Contagion Analysis in the Banking Sector. Working paper.

Darolles, S. & C. Gourieroux (2015) *Contagion Phenomena with Applications in Finance*. Elsevier.

De Jong, R.M. & J. Davidson (2000) Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica* 68(2), 407–423.

Diebold, F.X. & K. Yılmaz (2014) On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics* 182(1), 119–134.

Eisenberg, L. & T. Noe (2001) Systemic risk in financial systems. *Management Science* 47, 236–249.

Elliott, M., B. Golub, & M.O. Jackson (2014) Financial networks and contagion. *American Economic Review* 104(10), 3115–3153.

Engle, R., T. Ito, & W.-L. Lin (1990) Meteor showers or heat waves? Heteroskedastic intra-day volatility in foreign exchange market. *Econometrica* 58, 525–542.

Foerster, A.T., P.-D.G. Sarte, & M.W. Watson (2011) Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy* 119(1), 1–38.

Forbes, K.J. & R. Rigobon (2002) No contagion, only interdependence: Measuring stock market comovements. *The Journal of Finance* 57(5), 2223–2261.

Forni, M., D. Giannone, M. Lippi, & L. Reichlin (2009) Opening the black box: Structural factor models with large cross sections. *Econometric Theory* 25(5), 1319–1347.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2000) The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics* 82(4), 540–554.

Forni, M. & M. Lippi (2001) The generalized dynamic factor model: Representation theory. *Econometric Theory* 82(4), 1113–1141.

Gagliardini, P. & C. Gouriéroux (2019) Identification by Laplace transforms in nonlinear time series and panel models with unobserved stochastic dynamic effects. *Journal of Econometrics* 208(2), 613–637.

Gallant, R., R. Giacomini, & G. Ragusa (2017) Bayesian estimation of state space models using moment conditions. *Journal of Econometrics* 201, 198–211.

Glover, K. & J. Willems (1974) Parametrizations of linear dynamical systems: Canonical forms and identifiability. *IEEE Transactions on Automatic Control* 19(6), 640–646.

Hallin, M. & M. Lippi (2013) Factor models in high dimensional time series. A time domain approach. *Stochastic Processes and their Applications* 123(7), 2678–2695.

Hamilton, J. D. (1994) State-space models. In R. F. Engle, D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4. Elsevier, pp. 3041–3080.

Hannan, E. (1969) The identification of vector mixed autoregressive-moving average system. *Biometrika* 56(1), 223–225.

Hannan, E. (1971) The identification problem for multiple equation systems with moving average errors. *Econometrica* 39(5), 751–756.

Hannan, E. (1975) The estimation of ARMA models. *Annals of Statistics* 3, 975–981.

Hannan, E. & C. Heyde (1972) On limit theorems for quadratic functions of discrete time series. *The Annals of Mathematical Statistics* 43, 2058–2066.

Hannan, E.J. & M. Deistler (1988) *The Statistical Theory of Linear Systems*. SIAM.

Hansen, L.P. & K.J. Singleton (1991) Computing semi-parametric efficiency bounds for linear time series models. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 387–412.

Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability* 12(1), 141–153.

Izenman, A.J. (1975) Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis* 5(2), 248–264.

King, M.A. & S. Wadhwani (1990) Transmission of volatility between stock markets. *The Review of Financial Studies* 3(1), 5–33.

Kleibergen, F. & R. Paap (2006) Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133(1), 97–126.

Komunjer, I. & S. Ng (2011) Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica* 79(6), 1995–2032.

Koop, G., M.H. Pesaran, & S.M. Potter (1996) Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74(1), 119–147.

Long, J.B.J. & C.I. Plosser (1983) Real business cycles. *Journal of Political Economy* 91(1), 39–69.

Manski, C.F. (1993) Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3), 531–542.

Masson, P. (1999) Contagion: Macroeconomic models with multiple equilibria. *Journal of International Money and Finance* 18(4), 587–602.

Mokkadem, A. (1988) Mixing properties of ARMA processes. *Stochastic Processes and their Applications* 29(2), 309–315.

Newey, W. & D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle, D. L. McFadden (eds.), *Handbook of Econometrics*, chapter 36, vol. 4. Elsevier.

Newey, W. & K. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–653.

Pesaran, M.H. & A. Pick (2007) Econometric issues in the analysis of contagion. *Journal of Economic Dynamics and Control* 31(4), 1245–1277.

Pesaran, M.H. & Y. Shin (1998) Generalized impulse response analysis in linear multivariate models. *Economics Letters* 58(1), 17–29.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Poetscher, B. (1983) Order estimation in ARMA-models by Lagrangian multiplier tests. *Annals of Statistics* 11, 872–885.

Reinsel, G. & R. Velu (1998). *Multivariate Reduced-Rank Regression*. Lecture Notes in Statistics, Springer.

Robin, J.-M. & R.J. Smith (2000) Tests of rank. *Econometric Theory* 16(2), 151–175.

Robinson, P.M. (1973) Generalized canonical analysis for time series. *Journal of Multivariate Analysis* 3(2), 141–160.

Schennach, S.M. (2014) Entropic latent variable integration via simulation. *Econometrica* 82(1), 345–385.

Shiu, J.-L. & Y. Hu (2013) Identification and estimation of nonlinear dynamic panel data models with unobserved covariates. *Journal of Econometrics* 175(2), 116–131.

Stock, J.H. & M.W. Watson (2005) Understanding changes in international business cycle dynamics. *Journal of the European Economic Association* 3(5), 968–1006.

Stock, J.H. & M.W. Watson (2018) Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal* 128(610), 917–948.

Trevino, I. (2020) Informational channels of financial contagion. *Econometrica* 88(1), 297–335.