Structural VAR models in the Frequency Domain[*]

Alain Guay[†]

Florian Pelgrin[‡]

Abstract

This paper proposes a joint methodology for the identification and inference of structural vector autoregressive models in the frequency domain. We show that identifying restrictions can be written naturally as an asymptotic least squares problem (Gouriéroux et al., 1985) in which there is a continuum of nonlinear estimating equations. Following Carrasco and Florens (2000), we then propose a continuum asymptotic least squares estimator (C-ALS) that efficiently exploits the continuum of estimating equations, thereby allowing to obtain optimal consistent estimates of impulse responses and reliable confidence intervals. Moreover, the identifying restrictions can be formally tested using an appropriate J-stat, and the frequency band can be selected using a data-driven procedure. Finally, we provide some Monte Carlo simulations and an application regarding the hours–productivity debate.

JEL classification: C12, C32, C51.

Keywords: SVARs, Frequency domain, Asymptotic least squares, Continuum of identifying restrictions.

[*]The first author gratefully acknowledges the financial support of SSHRC and FRQSC. We thank the editor, the associate editor and two anonymous referees for very helpful remarks and suggestions. We would also like to thank Claire Alessandrini, Jean-Pierre Florens, René Garcia, John Galbraith, Stéphane Gregoir, Christophe Hurlin, Patrick Fève, Nour Meddahi, Serena Ng, Benoît Perron, Simon Van Norden, Roberto Pancrazi, Eric Renault and Eva Ortega.

[†]Université du Québec à Montréal, Chaire en macroéconomie et prévisions ESG-UQAM, and CIREQ, *e-mail*: guay.alain@uqam.ca.

[‡]EDHEC Business School, *e-mail*: florian.pelgrin@edhec.edu.

1 Introduction

While the ability of vector autoregressive models (VAR) models as descriptive and/or forecasting tools is well established, structural interpretation of VAR models is still subject to lively debates. Following the seminal works of Sims (1980a, 1980b), moving from atheoretical VAR models to structural VAR models requires identifying assumptions that rest on economic theory (among others); VAR results cannot be interpreted independently of a more structural macroeconomic model (Cooley and Leroy, 1985; Bernanke, 1986). This paper highlights the relevance of identifying restrictions in the frequency domain and proposes a general framework and its practical inference implementation for structural VAR models.

Our starting point is that identifying restrictions in the frequency domain can be particularly appealing in various contexts and can be motivated by economic intuition and some statistical arguments.[1] Imposing identifying restrictions at a given frequency has a long tradition in macroeconomics: a long-run identifying restriction à la Blanchard-Quah (1989) is nothing more than a low-frequency restriction at the frequency $\omega = 0$, or seasonal business cycles can be identified by imposing a frequency-identifying restriction, say at $\omega = \frac{\pi}{2}$ for quarterly data (Wen, 2001; 2002). Similarly, imposing restrictions on a frequency interval can be justified in many interesting contexts. Indeed, it is widely recognized that most macroeconomic series are generally considered to reflect both business cycle forces and lower-frequency forces unrelated to business cycles (demographic factors, trend growth, etc). For instance, Francis and Ramey (2009) convincingly argue that the conflicting results regarding the effect of a technological shock on hours worked can be explained by the fact that demographic trends and sectoral allocation are important sources of low-frequency movements in hours worked and labor productivity, and not only at $\omega = 0$, thereby suggesting to impose restrictions on a frequency interval around $\omega = 0$. Meanwhile, the recent literature on the identification of a news shock (e.g., Barsky and Sims, 2011; Kurmann and Sims, 2021) has emphasized the importance of isolating short-term and business cycle fluctuations from medium- and long-term effects to support the view that a news shock is the only shock driving productivity growth. Provided that one can define or test the relevant ranges of periodicities, that is, cycle length, to focus on, it makes sense to assume that some structural shocks are sources of medium- or low-frequency movements. This leads to imposing identifying restrictions on a frequency band and, possibly, to testing their empirical relevance and selecting the frequency band of interest.

Such a frequency-based identification is also consistent with the recent bulk of the literature (Angeletos et al., 2020) that aims at identifying a single structural shock explaining the fluctuations of a wide set of macroeconomic variables in a given range of periodicities, say 6–32 quarters, in the frequency domain. Indeed, it can be shown that common features (e.g., common business cycle components) can be identified, inferred, and evaluated using SVAR models in the frequency domain.[2] Meanwhile, our analysis is motivated by the seminal contributions of Watson (1993) and King and Watson (1996), which suggest plotting the

[1]Note that imposing frequency identifying restrictions is not mutually exclusive to other exclusion restrictions (e.g., short-run restrictions).

[2]See Section 8 of the online Appendix.

model and data spectra as one of the most informative diagnostics. Given that the macro variables of interest have a vector moving average representation in terms of structural shocks, reproducing and capturing characteristics of the spectral density matrix presuppose that (some of) these structural shocks have informative content to explain peak ranges arising at some frequency intervals.

As a last, but not least argument, imposing identifying restrictions on *low frequency* intervals offers an alternative to the so-called unreliability of long-run restrictions (Sims, 1971, 1972; Faust and Leeper, 1997; Faust, 1998) due to the sampling uncertainty of long-run multipliers irrespective of the sample size. The fact that confidence intervals for the structural impulse responses are generally too wide, thus making it difficult to reject misspecified models, is directly inherited from the zero Lebesgue measure of the long-run identifying restriction at $\omega = 0$ (Faust, 1996). Accordingly, our approach is fully consistent with the propositions of Faust and Leeper (1997), Faust (1998, 1999), Hauser et al. (1999) and Pötscher (2002), that is, to impose restrictions on a frequency interval around zero, but not too small relative to the sample size.[3] Moreover, restrictions based on a frequency interval surrounding zero is probably more robust to a structural change in the deterministic components of the series in the VAR than in the case of long-run restrictions.[4] To summarize, all these arguments underline the relevance of imposing some identifying restrictions in the frequency domain, testing their empirical plausibility, and selecting the frequency interval of interest. This is precisely what our methodological framework allows us to do.

In this respect, our starting point is to view the mapping between the structural and the reduced-form representation of the VAR model as a system of nonlinear estimating equations in which the structural parameters of interest depend on auxiliary parameters, those of the reduced-form representation. Capitalizing on Gouriéroux et al. (1985), we first cast the identification and estimation of structural VAR models into an appropriate asymptotic least squares (ALS) problem in the time or frequency domain.[5] While the ALS framework has already been used by Pesendorfer and Schmidt-Dengler (2008) for dynamic games with finite actions, it has not yet been considered for the identification, estimation, and evaluation of macro models in the frequency or time domain. Moreover, it turns out that our ALS-based procedure embeds the method-of-moment estimator and the estimation by a nonlinear equation solver discussed by Kilian and Lütkepohl (2017, chapters 11 and 12) or the sequential two-step GMM procedure of Bernanke and Mihov (1998) in the presence of short-run restrictions.

Defining the frequency identifying restrictions like a Fourier transform on a frequency interval, say $\omega \in (\underline{\omega}, \overline{\omega})$, one can fully exploit the fact that these restrictions are now defined by some functional equations and thus there exists a continuum of nonlinear estimating equations on $(\underline{\omega}, \overline{\omega})$. Building on the results of Carrasco

[3] Note that our approach is different from the alternative nonparametric estimators of the spectral density matrix at $\omega = 0$ suggested by Christiano et al. (2006a, 2006b). The comparison with Christiano et al. (2006a, 2006b) is further discussed in the online Appendix.

[4] Simulation results available upon request corroborate this conjecture.

[5] See also Szroeter (1983).

and Florens (2000) and Carrasco et al. (2007a), namely the continuum-generalized method of moments (C-GMM), we then propose an efficient two-step continuum-ALS (C-ALS) estimator.[6] Importantly, our approach is optimal compared to an alternative strategy based on the selection of a sufficiently refined grid through a discretization of the frequency band.

One key feature of our procedure is that (structural) identifying restrictions are not written as a generalized method of moments problem. This stems from the fact that, in general, identifying restrictions, that is, nonlinear (functional) estimating equations in our framework, do not necessarily involve any (conditional) expectation operator. We instead set the inference problem as a well-posed minimum distance one that rests on some (functional) estimating equations. As shown in Section 3, such a departure from the continuum-GMM approach (Carrasco and Florens, 2000; Carrasco et al., 2007a) has substantial implications regarding the properties of the covariance (kernel) operators, the identification of the structural parameters, and the asymptotic theory.

Our new methodology for the identification and inference of structural VAR models has appealing features. On the estimation side, Monte-Carlo simulations highlight that the C-ALS estimator has very interesting finite sample properties and performs better than traditional alternatives in terms of bias and (root) mean squared error irrespective of the impulse response function (IRF) horizon. On the inference side, our framework offers the opportunity to implement testing procedures and thus to provide new insights on the validity of both the identifying restrictions and the frequency band of interest. In particular, we show that the derivation of the overidentification J-based tests and their corresponding asymptotic distributions are different from the results of Carrasco and Florens (2000) and Amengual et al. (2020). The modified J-stat allows for selecting the frequency band, thereby permitting us to conduct a data-driven procedure to evaluate frequency intervals on which the imposed restrictions might be satisfied. We illustrate the usefulness of the overidentification test and the interval selection in an application regarding the identification of a technology shock.

The remainder of the paper is organized as follows. Section 2 reviews some notations and illustrates the idea of our approach in a simple setting, namely a bivariate structural VAR model. Section 3 presents the (optimal) C-ALS estimator while Section 4 applies these results in the case of any N-dimensional structural VAR model. Section 5 proposes a comparative study of competing identification schemes using some Monte-Carlo simulations. Section 6 revisits the empirical evidence of the productivity–hours debate. Section 7 contains concluding comments and possible future extensions. Proofs are given in the Appendix.

[6] Note that Pesendorfer and Schmidt-Dengler (2008, p. 917) mention a continuum ALS procedure as a possible extension of their work but do not derive it.

2 Identification of SVAR models and ALS

2.1 Notations

We first introduce some preliminary notations and the structural VAR models. Suppose an N-dimensional multiple time series $X_1, X_2, \cdots, X_T$ with $X_t = (X_{1t}, \cdots, X_{Nt})'$ is available and that these variables are second-order stationary. The vector X_t can include level stationary variables, integrated variables in first-difference or stationary linear combinations of integrated variables. To simplify the notation, presample values for each variable are assumed to be available. X_t is assumed to be approximated as the sum of a deterministic part ϑ_t and by a stationary, stable, reduced-form $\mathtt{VAR}(p)$ process:

$$X_t \quad = \quad \vartheta_t + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + u_t, \tag{2.1}$$

where the ϕ_i are fixed $(N \times N)$ coefficient matrices for lag i, $\Phi_p = [\phi_1 \; \phi_2 \; \ldots \; \phi_p]$ is a $(N \times (pN))$ matrix of all autoregressive coefficients, and $u_t = (u_{1t}, \cdots, u_{Nt})'$ is the vector of innovations such that $E(u_t) = 0_{N\times 1}$, $E(u_t u_t') = \Sigma_u$ is nonsingular and $E(u_t u_s') = 0_{N\times N}$ for $s \neq t$.

The corresponding vector moving average representation in terms of innovations is defined by:

$$X_t = \vartheta_t^* + \sum_{i=0}^{\infty} C_i u_{t-i} = \vartheta_t^* + C(L)u_t, \tag{2.2}$$

where $C(L) = \sum_{i=0}^{\infty} C_i L^i$, $C_0 = C(0) = I_N$, $C(1) = \sum_{i=0}^{\infty} C_i$, $C_i = \sum_{j=1}^{i} C_{i-j}\phi_j$ and ϑ_t^* is the corresponding deterministic part. Notably, when ϑ_t is a vector of intercepts, then $\vartheta_t = \vartheta$ and $\vartheta_t^* = C(1)\vartheta$. Meanwhile, the moving average representation in terms of structural shocks is written as:

$$X_t = \vartheta_t^* + \sum_{i=0}^{\infty} A_i \epsilon_{t-i} = \vartheta_t^* + A(L)\epsilon_t, \tag{2.3}$$

where $A(L) = \sum_{i=0}^{\infty} A_i L^i$, $A_0 \equiv A(0)$, and ϵ_t is an $N \times 1$ random vector of structural shocks with $E(\epsilon_t) = 0$ and $E(\epsilon_t \epsilon_t') = \mathrm{I}_N$ (normalization assumption).

Taking equations (2.2) and (2.3), the error terms of the reduced-form model are related to the structural shocks as follows:[7]

$$u_t = A(0)\epsilon_t, \tag{2.4}$$

with $C(L)A(0) = A(L)$ and $\Sigma = A(0)A(0)'$. In general, for an argument z, we have the following mapping between the reduced-form moving average matrix $C(z)$ and the structural moving average matrices $A(0)$ and $A(z)$:

$$C(z)A(0) = A(z), \tag{2.5}$$

which implies $C(z)\Sigma C(z)' = A(z)A(z)'$. The central question is then how to recover the elements of $A(0)$ from consistent estimates of the reduced-form parameters. In so doing, one generally needs to impose some identifying restrictions on equation (2.5).

[7] For a more general presentation, see Lütkepohl (2007), Kilian (2013), and Kilian and Lütkepohl (2017).

2.2 Identifying restrictions and asymptotic least squares

We now discuss standard identifying restrictions as being the estimating equations of the ALS procedure proposed by Gouriéroux et al. (1985). We then turn to their generalization in the frequency domain. In particular, to illustrate the intuition of our approach in a simple setting, our starting point is the bivariate VAR specification of Beaudry and Portier (2006) that describes the joint behavior of stock prices and measured total factor productivity. Let $X_t = (\Delta \mathrm{tfp}_t, \Delta \mathrm{sp}_t)'$, where tfp_t is the log of total factor productivity, and sp_t is the log of an index of stock market value. For the sake of simplicity, we assume that the two log-level variables are not cointegrated. The vector of structural shocks is $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})'$.

(a) **Identification scheme** In the absence of further information, recovering the vector moving average representation in terms of structural shocks requires finding the matrix $A(0) = A_0$ that solves the following system of (estimating) equations:

$$\mathrm{vech}\left(\Sigma_u - A(0)A(0)'\right) = \mathbf{0}, \tag{2.6}$$

where vech is the half-vectorization operator. We then must compute the remaining structural matrices:

$$A_k = C_k A(0) \quad \text{for } k > 0. \tag{2.7}$$

Looking at the system of estimating equations (2.6), there is one more structural parameter than independent (estimating) nonlinear equations. Therefore, it is necessary to add one (respectively, at least two) identifying restriction to get a just-identified (respectively, an overidentified) system of estimating equations. In this respect, one can impose that only the second shock ϵ_{2t} has a long-run effect on the log-level of total factor productivity (Assumption 1), and this shock has no contemporaneous impact on $\Delta \mathrm{tfp}_t$ (Assumption 2). As shown by Beaudry and Portier (2006), the two assumptions are compatible with a model in which a permanent technology shock is characterized by a diffusion process having no immediate impact on productive capacity. This is what they called a news shock. While Beaudry and Portier (2006) sequentially consider the two (long-run and short-run) identifying restrictions, we merge them with the aim of illustrating our approach in the case of overidentification.

Assumption 1 implies that the (1,1) element of the matrix of long-run cumulative effects of the structural shocks, $A(1)$, is zero and Assumption 2 that the (1,2) element of A(0) is zero:

$$\begin{cases} a_{11}(1) = 0 \\ a_{12}(0) = 0 \end{cases} \Leftrightarrow \begin{cases} c_{11}(1)a_{11}(0) + c_{12}(1)a_{21}(0) = 0 \\ a_{12}(0) = 0 \end{cases}$$

because $a_{11}(1) = \left[C(1)A(0)\right]_{11}$. Therefore, the system of estimating equations is overidentified:

$$\begin{cases} a_{11}^2(0) + a_{12}^2(0) = \sigma_{11} \\ a_{11}(0)a_{21}(0) + a_{12}(0)a_{22}(0) = \sigma_{12} \\ a_{21}^2(0) + a_{22}^2(0) = \sigma_{22} \\ c_{11}(1)a_{11}(0) + c_{12}(1)a_{21}(0) = 0 \\ a_{12}(0) = 0, \end{cases} \tag{2.8}$$

where $\sigma_{ij} = [\Sigma_u]_{ij}$ is the (i,j) element of Σ_u. However, in the absence of Assumption 2, the system of estimating equations is given by the first four equations of (2.8), and there are as many unknowns as equations. In both cases, the system of estimating equations can be written as:

$$g(a(0), \beta_0) = \mathbf{0}, \tag{2.9}$$

where $\beta_0 = (\text{vec}(\Phi_p)', \text{vech}(\Sigma_u)')'$ is the vector of true autoregressive parameters and covariance parameters—the so-called vector of auxiliary parameters (equation 2.1).

***(b)* Asymptotic least squares procedure in the time domain** Following Gouriéroux et al. (1985) and Gouriéroux and Monfort (1995), the starting point of the ALS procedure is to pass through the auxiliary parameters, which can be easily estimated, to solve the estimation problem of the parameters of interest $a(0)$. More specifically, replacing β_0 by some consistent and asymptotically normally distributed estimate $\widehat{\beta}_T$, the core principle consists of estimating the parameters of interest $a(0)$ by forcing the constraints (2.9) to be as close as possible to zero.

Then, an ALS estimator associated with the symmetric positive definite matrix S_T is a solution $\widehat{a}_T(0)$ to the minimization problem:

$$\min_{a(0)\in\mathcal{A}} \quad g(a(0), \widehat{\beta}_T)' S_T g(a(0), \widehat{\beta}_T).$$

The ALS estimator $\widehat{a}_T(0)$ brings the constraints closest to zero using the metric associated with the scalar product defined by the weighting matrix S_T. Under suitable regularity conditions, there exists an *optimal choice* for the sequence of matrices S_T (Gouriéroux et al., 1985).

In the context of SVAR, the ALS estimator can be implemented as a two-step procedure. Given a consistent estimator $\widehat{\beta}_T$ of the auxiliary parameters vector $\beta_0 = (\text{vec}(\Phi_p)', \text{vech}(\Sigma_u)')'$ of the reduced-form VAR(p) model, the optimal ALS estimator of the parameters of interest $a(0)$ is obtained as follows. In a first step, a consistent estimate of $a(0)$, denoted by $\widehat{a}_T^1$, is usually obtained by minimizing the square norm of the vector of estimating equations; that is, the weighting matrix is the identity matrix. In a second step, a consistent and efficient estimate of $a(0)$, denoted by $\widehat{a}_T$, minimizes a quadratic form with the use of a consistent estimate of the optimal weighting matrix evaluated at $\widehat{a}_T^1$ and $\widehat{\beta}_T = \left(\text{vec}\left(\widehat{\Phi}_{p,T}\right)', \text{vech}\left(\widehat{\Omega}_T\right)'\right)'$.

As a final remark, note that the first set of estimating equations (equation 2.6), which describes the mapping between the variance-covariance matrix of the innovations and of the structural shocks and possibly some short-run restrictions, can be written as moment conditions. It is also the case when imposing only long-run restrictions such that the matrix of long-run cumulative effects of the structural shocks, $A(1)$, is lower triangular and the system of estimating equations is just-identified. However, it is no longer true for general long-run identifying restrictions or for combinations of short, long, and any other zero identifying restrictions.[8] Instead, it can be written as a system of nonlinear equations with some equations that do not

[8] See Bernanke and Mihov (1998) and Kilian and Lütkepohl (2017), Sections 11.2.1 and 11.3.

involve any expectation operator. This corresponds to a minimum distance problem, which can be embedded into an ALS framework. This is particularly the case when considering overidentifying frequency restrictions.

(c) **Identifying restrictions on a frequency interval** Taking the just-identified case from now on, a credible alternative to the long-run restriction $a_{11}(1) = 0$ is to impose frequency restrictions at $\omega = 0$ but also its neighborhood in order to better capture the lower-frequency forces unrelated to medium-term fluctuations and business cycles.[9] The identifying restriction is then given by $a_{11}(e^{-i\omega}) = 0$, for all $\omega \in I_\omega = [-\omega; \omega]$, or equivalently:

$$c_{11}(e^{-i\omega})a_{11}(0) + c_{12}(e^{-i\omega})a_{21}(0) = 0. \tag{2.10}$$

This identifying restriction is similar to a Fourier transform of the standard long-run identification restriction on a (symmetric) frequency interval around $\omega = 0$. Consequently, there is now a continuum of identifying restrictions defined on I_ω. It turns out that the functional estimating equation (2.10) leads now to a case of overidentification. This point is all the more important as it opens up the possibility of testing the relevance of identifying restrictions and selecting the frequency band (see Section 3).

Finally, such a continuum of identifying restrictions can also be justified in this simple example by looking at the marginal spectral density of the first variable (Δtfp_t), denoted by $f_{X,11}(\omega)$. Indeed, it is straightforward to show that:

$$f_{X,11}(\omega) = \frac{1}{2\pi}\left\{\left|c_{11}(e^{-i\omega})a_{11}(0) + c_{12}(e^{-i\omega})a_{21}(0)\right|^2 + \left|c_{11}(e^{-i\omega})a_{12}(0) + c_{12}(e^{-i\omega})a_{22}(0)\right|^2\right\},$$

where $|\Psi(\omega)|^2$ is the modulus of the complex-valued function Ψ. Accordingly, the first (respectively, second) right-hand side term captures the contribution of the first (respectively, second) structural shock to the (marginal) spectral density of Δtfp_t at frequency ω. Therefore, imposing the identifying restriction on the frequency interval $[-\omega; \omega]$ is equivalent to minimizing the contribution of the first structural shock to the marginal spectral density $f_{X,11}(\omega)$ for all $\omega \in [-\omega; \omega]$.

In this respect, the inference of $a(0)$ can be solved in two ways. On the one hand, one can proceed with a discretization of the frequency interval and make use of the standard ALS estimator on a system of discretized estimating equations. On the other hand, instead of selecting a finite number of grid points of the frequency band, a continuum ALS procedure that efficiently exploits the functional estimating equation is proposed in this paper. Building on Carrasco and Florens (2000), we show that the presence of a continuum of estimating equations naturally leads to proposing a continuum ALS estimator that closely mimics the efficient two-step ALS of Gouriéroux et al. (1985).

[9] See Hamilton (1994) for a comprehensive overview of spectral analysis and the online Appendix.

3 Asymptotic Least Squares in the frequency domain

In this section, we propose a general ALS estimator in the presence of a continuum of identifying restrictions (C-ALS estimator). First, we define the class of C-ALS estimators for every sequence of random bounded linear operators. Second, the optimal C-ALS estimator is presented. Finally, a test of overidentification and a data-driven procedure for the choice of the frequency interval are discussed.

3.1 The class of C-ALS estimators

The stochastic process Z_t is an $N \times 1$ vector of random variables that belongs to a family of true unknown probability distributions, denoted by P_0, and is a second-order stationary and weakly dependent process. We first suppose that:

A.1 There exists a consistent and asymptotically normally distributed estimator of the true q-vector of auxiliary parameters β_0, denoted by β_T, that is, $\widehat{\beta}_T \xrightarrow{p} \beta_0$ and $\sqrt{T}\left(\widehat{\beta}_T - \beta_0\right) \xrightarrow{d} \mathcal{N}(0, \Omega)$.

We consider a system of J constraints, which is function of a r-vector of structural parameters α, defined on a continuum of frequencies such that:

$$g(\alpha_0, \widehat{\beta}_T, \omega) = 0, \tag{3.11}$$

where $g(\cdot,\cdot,\omega)$ takes its values in $H = (L^2(\mathcal{I}))^J$, a Hilbert space with the inner product $< .,. >$ and the norm $\|\cdot\|$ with $\alpha = \alpha_0$ under P_0.[10] $L^2(\mathcal{I}, \varphi) \equiv L^2(\varphi)$ is the space of complex valued functions that are uniformly square integrable with respect to the interval $\mathcal{I}$ for ω.[11] Let S denote a bounded linear operator defined on H or a subspace of H and $\overline{g(\cdot,\cdot,\omega)}$ denote the complex conjugate of $g(\cdot,\cdot,\omega)$. Let S_T be a sequence of random bounded linear operators converging in probability to S. The C-ALS estimator is defined by:

$$\widehat{\alpha}_T(S_T) = \arg\min_{\alpha \in \mathcal{A}} \left\| S_T g(\alpha, \widehat{\beta}_T, \omega) \right\|^2 .$$

Therefore, the C-ALS estimator forces the constraints, $g(\alpha, \widehat{\beta}_T, \omega) = 0$ for $\omega \in [\underline{\omega}, \overline{\omega}]$, to be as close as possible to zero by using the metric associated with the inner product defined by S_T on the interval $[-\pi, \pi]$. We now present the other assumptions for the asymptotic properties of the C-ALS estimator.

A.2 The J-vector of functions $g(\alpha_0, \beta_0, \omega) = 0$ $\forall \omega \in \mathcal{I}$ has a unique solution α_0, which is an interior point of a compact set $\mathcal{A}$, and α_0 and β_0 denote the unknown values under P_0.

A.3 Let $N(S)$ denote the null space of S, $N(S) = \{f \in H | Sf = 0\}$. We have that $g \in N(S)$ implies $g(\alpha_0, \beta_0, \omega) = 0$.

[10] Alternatively, the J constraint functions can be rewritten as a scalar function $\tilde{g}(\alpha_0, \widehat{\beta}_T, \tilde{\omega}_j)$ with $\tilde{\omega}_j = (\omega, j)$ where $\omega \in [-\pi, \pi]$ and $j \in \{1, 2, \ldots, J\}$, which takes its value in a suitably defined Hilbert space of a scalar function (see Kailath, 1971 and Carrasco et al., 2007a, p.534).

[11] See Assumption A.2 and Definition A.1 of Carrasco et al. (2007a) for Hilbert space of complex-valued functions.

Assumption **A.3** is an identification condition similar to the one in Carrasco and Florens (2000). In contrast to Carrasco et al. (2007a), we cannot suppose that the null space $N(S)$ reduces to $\{0\}$ as it will be clear hereafter because the optimal operator is of finite range, which implies that the null space of the optimal operator is a non-trivial closed subspace of H.

A.4 (i) $g(\alpha, \beta, \omega)$ is continuously differentiable with respect to α and β and $g(\alpha, \beta, \omega) \in (L_\infty(\mathcal{I} \otimes P_0))^J$ where $L_\infty(\mathcal{I} \otimes P_0)$ is the set of measurable bounded functions of (ω, Z_t).
(ii) $\sup_{\alpha \in \mathcal{A}} \|g(\alpha, \beta, \omega) - g(\alpha_0, \beta, \omega)\| = O_p\big(\frac{1}{\sqrt{T}}\big)$ for all $\beta \in \mathcal{B}$ and $\omega \in \mathcal{I}$.
(iii) $\sup_{\alpha \in \mathcal{A}_0} \|\partial g(\alpha, \beta, \omega)/\partial \alpha' - \partial g(\alpha_0, \beta, \omega)/\partial \alpha'\| = O_p\big(\frac{1}{\sqrt{T}}\big)$ for all $\beta \in \mathcal{B}$ and $\omega \in \mathcal{I}$ where $\mathcal{A}_0$ is some neighborhood about α_0.
(iv) $\sup_{\beta \in \mathcal{B}_0} \|\partial g(\alpha, \beta, \omega)/\partial \beta' - \partial g(\alpha, \beta_0, \omega)/\partial \beta'\| = O_p\big(\frac{1}{\sqrt{T}}\big)$ for all $\alpha \in \mathcal{A}$ and $\omega \in \mathcal{I}$ where $\mathcal{B}_0$ is some neighborhood about β_0.

In the context of structural VAR models, $g(\alpha, \beta, \omega)$ is bounded (Assumption **A.4** (i)) for any interval under the assumption of weakly stationary processes. Note that this assumption is not satisfied in the case of nonstationary (integrated of order one) processes when $\omega = 0$ belongs to the frequency interval. We can also easily show that Assumption **A.4** ii), iii), and iv) hold for the continuum of estimating equations (3.11) in the case of weakly stationary processes. Indeed, as shown in Section 4, the function $g(\alpha, \beta, \omega)$ and its derivatives depend on the estimates of the moving average representation in terms of innovations, which converge in probability for weakly stationary processes (see Phillips, 1998).

The following Lemma establishes the asymptotic normality of the functional

$$\frac{\partial g}{\partial \beta'}(\alpha_0, \widehat{\beta}_T, \omega)\sqrt{T}(\widehat{\beta}_T - \beta_0),$$

which is necessary to obtain the asymptotic normality of C-ALS estimators.

Lemma 3.1. *Under Assumptions **A.1**, **A.2** and **A.4** we have*

$$\frac{\partial g}{\partial \beta'}(\alpha_0, \widehat{\beta}_T, \omega)\sqrt{T}(\widehat{\beta}_T - \beta_0) \Rightarrow \xi \sim \mathcal{N}(0, K)$$

as $T \to \infty$ in H, where $N(0, K)$ is the Gaussian random vector of H with the covariance operator $K : H \to H$ satisfying for all f in H:

$$(Kf)(\omega_1) = \int_{\mathcal{I}} k(\omega_1, \omega_2) f(\omega_2) d\omega_2$$

where, under P_0, the kernel of K, denoted by k, is:

$$k(\omega_1, \omega_2) = \frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0, \omega_1)\Omega\overline{\frac{\partial g'}{\partial \beta}(\alpha_0, \beta_0, \omega_2)}$$

with $\int_{\mathcal{I}}\int_{\mathcal{I}} k(\omega_1, \omega_2) d\omega_1 d\omega_2 < \infty$. Moreover, K is a compact Hilbert–Schmidt operator, is self-adjoint ($K = K^*$), and $\xi \in \mathcal{H}(K)$ where $\mathcal{H}(K) \subset H$ is the so-called reproducing kernel Hilbert space (RKHS).[12]

[12]See Carrasco et al. (2007b) for definitions and properties of RKHS. See Wahba (1990) for RKHS of vector-valued functions.

Lemma 3.1 results directly from Theorem 2.47 in Carrasco et al. (2007b) under the assumption of convergence in probability of the random bounded linear operator $\frac{\partial g}{\partial \beta'}(\alpha_0, \widehat{\beta}_T, \omega)$ to $\frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0, \omega)$ (Assumption **A.4** (iv)) and $\sqrt{T}\left(\widehat{\beta}_T - \beta_0\right) \xrightarrow{d} \mathcal{N}(0, \Omega)$. Here, the Hilbert–Schmidt operator K has a finite dimensional closed range equal to q, the dimension of the (asymptotic) covariance matrix of the parameter vector β. Indeed the operator K, which depends on the first step estimation $\widehat{\beta}_T$, is degenerate and thus can be written as a finite sum of products of functions depending on ω_1 and ω_2 (see Groetsch, 1984, p.8). Accordingly, the range of K, denoted by $\mathcal{R}(K)$, is finite dimensional. This is different from the framework of Carrasco and Florens (2000) and Carrasco et al. (2007a) in which the inverse of K is not bounded because the range of K is not closed.[13] Since the dimension of the range of K is q, the number of its eigenvalues different from zero is finite and, according to the Mercer's Theorem, K admits the following spectral decomposition:

$$k(\omega_1, \omega_2) = \sum_{i=1}^{q} \lambda_i \gamma_i(\omega_1) \overline{\gamma_i(\omega_2)'},$$

where λ_i, $i = 1, \ldots, q$ denote the q eigenvalues of K different from zero and $\gamma_i(\omega_1)$ the corresponding vector of orthonormalized eigenfunctions, that is, $K\gamma_i(\omega_1) = \lambda_i \gamma_i(\omega_1)$ for $i = 1, \ldots, q$.[14] It follows that

$$(Kf)(\omega_1) = \sum_{i=1}^{q} \lambda_i \gamma_i(\omega_1) \langle f, \gamma_i \rangle .$$

The choice of the optimal C-ALS estimator is related to the inverse of the covariance operator K. Inverting K is equivalent to find the solution ϕ to a Fredholm equation of the first kind $K\phi = f$ for a given $f \in H$. With $\mathcal{R}(K)$ of finite dimension and thereby closed, the uniqueness of the inverse of K is achieved by the solution of minimal norm for the equation of the first kind.[15] In addition to the assumptions **A.2** and **A.3**, the identification of α thus requires the following assumption in the case of the optimal C-ALS estimator.

A.5 i) $g(\alpha, \widehat{\beta}_T, \omega) \in \mathcal{R}(K)$, $\forall \alpha \in \mathcal{A}$ under P_0, ii) $Prob(rank(K_T) = q) \to 1$ as $T \to \infty$ where K_T is a consistent estimator of K and iii) $\partial g(\alpha, \widehat{\beta}_T, \omega)/\partial \alpha \in \mathcal{H}(K)$ for all $\alpha \in \mathcal{A}_0$ under P_0.

Assumption **A.5** i) means that $g(\cdot, \cdot, \omega)$ is well-specified on the closed range of $K^* = K$ and $\mathcal{R}(K) = N(K)^{\perp}$.[16] This implies that a unique minimal norm solution exists for the equation of the first kind and the solution is given by the following Moore-Penrose inverse denoted $K^{\dagger}$:

$$\left(K^{\dagger} f\right)(\omega_1) = \sum_{i=1}^{q} \frac{1}{\lambda_i} \gamma_i(\omega_1) \langle f, \gamma_i \rangle . \tag{3.12}$$

Thus, by Assumption **A.5** i), we have $\|(K^{\dagger})^{1/2} g(\alpha, \beta, \omega)\| = 0$ under P_0 implies $g(\alpha_0, \beta_0, \omega) = 0$ $\forall \omega \in \mathcal{I}$ has a unique solution $\alpha = \alpha_0$. Assumption **A.5** ii) implies that the asymptotic rank of a consistent estimator K_T of K is the same as K with probability one. This is a sufficient condition to ensure that the Moore-Penrose

[13] For a general discussion of linear inverse problems in econometrics, see Carrasco et al. (2007b).

[14] The derivation of the explicit expressions for the eigenvalues λ_i and the eigenfunctions $\gamma_i(\omega)$ is given in the online Appendix.

[15] See Proposition 3.3 in Carrasco et al. (2007b).

[16] See Luenberger, 1997, p.156, Theorem 2 and Babii and Florens (2021).

inverse estimator $K_T^\dagger$ is consistent for $K^\dagger$ for a consistent estimator K_T of K.[17] However, the stability of the Moore-Penrose inverse depends on the rate of decay of the eigenvalues and thus a regularization method cannot be precluded in finite samples in the presence of tiny eigenvalues. In particular, the severity of rank deficiency in finite samples can give a useful indication as to whether the information contained in a certain interval is sufficient to implement the optimal estimation and the corresponding testing procedure presented below. Assumptions **A.5** iii) guarantees that the partial derivatives of the function $g(\alpha, \widehat{\beta}_T, \omega)$ with respect to α are in the RKHS of the covariance operator K. Finally, the following full-rank assumptions are also needed to derive the asymptotic distribution.

A.6 The matrices $\left\langle S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle$ and $\left\langle (K^\dagger)^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), (K^\dagger)^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle$ are positive definite and symmetric, which implies that $dim(\alpha) \leq dim(\beta)$ $(r \leq q)$ where $K^{-1/2} = (K^\dagger)^{1/2}$.

The following proposition shows that the C-ALS estimator is consistent and asymptotically normally distributed.

Proposition 3.1. *Suppose that Assumptions **A.1** to **A.6** are satisfied, that S_T denotes a sequence of random bounded linear operators converging to S, and that the C-ALS estimator associated with S_T is a solution $\widehat{\alpha}_T(S_T)$ to the problem:*

$$\widehat{\alpha}_T(S_T) = \arg\min_{\alpha \in \mathcal{A}} \|S_T g(\alpha, \widehat{\beta}_T, \omega)\|^2. \tag{3.13}$$

The C-ALS estimator exists, and $\widehat{\alpha}_T \xrightarrow{p} \alpha_0$. Moreover, it is asymptotically normally distributed:

$$\sqrt{T}\left(\widehat{\alpha}_T(S_T) - \alpha_0\right) \xrightarrow{d} N(0, Q(S))$$

with

$$\begin{aligned} Q(S) \quad = \quad & \left\langle S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle^{-1} \left\langle S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), (SK(S^\star)\, S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle \\ & \times \left\langle S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle^{-1} \end{aligned}$$

and among all admissible operators S, $S = K^{-1/2}$ yields the optimal estimator, i.e. with minimal variance, where $K^{-1/2} = (K^\dagger)^{-1/2}$.

The results follow from Carrasco and Florens (2000) but are applied to a continuum of estimating equations.[18] Proposition 3.1 implies that the asymptotic variance of an alternative estimator of α based on a discretized ALS is necessarily greater or equal to the lower bound achieved with the optimal C-ALS estimator. Consequently, the optimal C-ALS achieves asymptotic efficiency by exploiting all the information available in the frequency interval. Moreover, the integrals appearing in the asymptotic expressions can be solved analytically without carrying out numerical integration, thereby avoiding approximation errors.

[17] See Andrews (1987) and Tyler (1981).

[18] A detailed derivation of the proof is given in the online Appendix.

As in Carrasco et al. (2007a), we propose a simple expression of the objective function, which permits to write the objective function in terms of finite dimensional vectors and matrices. This requires a first-step consistent C-ALS estimator, denoted by $\widehat{\alpha}_T^1$, defined by (using the identity operator as a kernel operator):

$$\widehat{\alpha}_T^1 = \arg\min_{\alpha \in \mathcal{A}} \int_{\mathcal{I}} g(\alpha, \widehat{\beta}_T, \omega)' \overline{g(\alpha, \widehat{\beta}_T, \omega)} d\omega.$$

Proposition 3.2. *A simplified expression for the objective function of the optimal C-ALS problem is given by:*

$$\widehat{\alpha}_T = \arg\min_{\alpha \in \mathcal{A}} \underline{s}(\alpha, \widehat{\beta}_T)' \widetilde{W}_T^2 \overline{\underline{s}(\alpha, \widehat{\beta}_T)},$$

where $\widetilde{W}_T$ is a generalized inverse of W_T and

$$W_T = \int_{\mathcal{I}} \Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega)} \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega) \Omega_T^{1/2} d\omega$$

is a $q \times q$-matrix and

$$\underline{s}(\alpha, \hat{\beta}_T) = \int_{\mathcal{I}} \Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega)} g(\alpha, \widehat{\beta}_T, \omega) d\omega$$

is a q-vector. When the matrix W_T is of full rank, then $\widetilde{W}_T = W_T^{-1}$.

Proof: See the Appendix.

The transformed system of estimating equations, $\underline{s}(\alpha, \hat{\beta}_T)$, is thus defined by an inner product condition in a Hilbert space, which is the projection of $g(\alpha, \widehat{\beta}_T, \omega)$ on the subspace spanned by $\widehat{\beta}_T$ given by $\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega)\widehat{\Omega}_T^{1/2}$.

As explained before, the matrix W_T might not be of full rank q as the frequency interval shrinks toward a point (e.g., the zero frequency). As proposed by Carrasco and Florens (2000) in a C-GMM context, a generalized inverse of W_T might be obtained through a Tikhonov's regularization. Thus, a simplified expression for the regularized objective function of the second-step C-ALS problem is given by:

$$\widehat{\alpha}_T = \arg\min_{\alpha \in \mathcal{A}} \underline{s}(\alpha, \widehat{\beta}_T)' [\eta_T I_q + W_T^2]^{-1} \overline{\underline{s}(\alpha, \widehat{\beta}_T)},$$

where the regularization parameter η_T goes to zero at a suitable rate (see Carrasco et al., 2007a; Carrasco, 2012).

3.2 Test of overidentification

A specification test can be performed using the $J-$test of overidentifying restrictions (Hansen, 1982).

Proposition 3.3. *Under Assumptions **A.1** to **A.6**, the following J_T-statistic is asymptotically distributed as a sum of weighted χ^2-distributed random variables with 1 degree of freedom:*

$$J_T = \|\sqrt{T} S_T g(\widehat{\alpha}, \widehat{\beta}_T, \omega)\|^2 \to \sum_{j=1}^{m} \lambda_j \chi_j^2(1), \tag{3.14}$$

where m is the dimension of the range of $(S\tilde{K}S^*)$, $\tilde{K}$ is a covariance operator defined in the Appendix, and $\lambda_j, j = 1, \dots m$ are the eigenvalues of $S\tilde{K}S^*$ ordered in decreasing order and $m \leq q - r$.

Proof: See the Appendix.

When $S_T = I$, the λ_j terms are the eigenvalues of the covariance operator $\tilde{K}$ with a range of dimension $q-r$. For the optimal C-ALS obtained with a consistent estimator of $S = (K^\dagger)^{1/2}$, the J_T-statistic is Chi-square distributed with $q - r$ degrees of freedom. In contrast to Carrasco and Florens (2000) and Amengual et al. (2020), we obtain a finite Chi-square as limiting distribution since K is degenerate and the number of eigenvalues of K is finite.

In the presence of a rank-order deficiency, Proposition 3.2 can be used to define a regularized version of the test statistic. More specifically, for a fixed value of η, the J_T statistic converges to the following distribution:

$$T\underline{s}(\widehat{\alpha}_T, \widehat{\beta}_T)'[\eta I_q + W_T^2]^{-1}\overline{\underline{s}(\widehat{\alpha}_T, \widehat{\beta}_T)} \xrightarrow{d} \sum_{j=1}^{q-r} \frac{\lambda_j^2}{\lambda_j^2 + \eta} \chi_j^2(1). \tag{3.15}$$

Critical values of the limiting distribution (3.14) and (3.15) can be obtained either by implementing the numerical inversion of the characteristic function proposed by Imhof (1961) or by simulating independent Chi-squared distributions (Robin and Smith, 2000) using consistent estimators of the eigenvalues.

3.3 Data-driven procedure for the frequency interval

The next question to address is the determination of the interval $I_\omega = [\underline{\omega}, \overline{\omega}]$ on which one might impose and assess the reliability of the identifying restrictions. For the sake of simplicity, we consider the class of symmetric intervals of ω around zero, that is, $I_\omega = [-\omega, \omega]$. Then we use the information criteria-based methodology of Hall et al. (2012) to select the largest interval I_ω that might guarantee consistent estimation of $\widehat{\alpha}_T$. In so doing, $\hat{\omega}_T$ is obtained by minimizing the valid interval selection criterion (VISC) defined by:

$$\hat{\omega}_T = \underset{\omega \in \mathcal{C}(\omega)}{\text{argmin}} \quad \mathtt{VISC}_T(\omega),$$

where $\mathcal{C}(\omega)$ is the class of symmetric intervals around zero and

$$\mathtt{VISC}_T(\omega) = J_T(\omega) - h(|w|)\kappa_T, \tag{3.16}$$

where $h(|w|)\kappa_T$ is a deterministic penalty, which is an increasing function of the length of the interval. Proposition 3.4 shows that $\hat{\omega}_T$ converges in probability to the unique ω_0 under P_0 that chooses the maximal bound for a valid consistent estimation of $\widehat{\alpha}_T$.

Proposition 3.4. *Suppose that (1) there exists a lower bound ω_{lb} such that the restrictions are respected for the interval $[-\omega_{lb}, \omega_{lb}]$, (2) there exists a maximal interval given by $I_{\omega_0} = [-\omega_0, \omega_0]$, and (3) $h(\cdot)$ is strictly increasing and $\kappa_T \to \infty$ as $T \to \infty$ with $\kappa_T = o(T)$. Then the estimator $\hat{\omega}_T$ defined as the solution of the criterion (3.16) converges in probability to ω_0.*

Note that the first assumption imposes that the restrictions are valid for at least an interval with minimal length characterized by the lower bound ω_{lb}. The second assumption ensures that the interval $[-\omega, \omega]$ is uniquely identified. The last one imposes restrictions on the penalty terms that guarantee the validity of the criterion. The SIC-type penalty term ($(h|\omega|) = 2\omega$ and $\kappa_T = \ln(T)$) and the Hannan–Quinn-type penalty term ($(h|\omega|) = 2\omega$ and $\kappa_T = ln(ln(T))$) satisfy this assumption, while the AIC-type penalty term ($(h|\omega|) = 2\omega$ and $\kappa_T = 2$), does not.

4 The C-ALS estimator for structural VAR models

In this section, we apply the general results of Section 3 to identify structural VAR models with frequency-based restrictions.[19] Suppose an N-dimensional vector $X_t = (X_{1t}, \cdots, X_{Nt})'$ follows a (reduced-form) VAR(p) specification (2.1). Without loss of generality, we assume that the (functional) system of estimating equations can be decomposed as follows:

$$g(a(0), \beta_0, \omega) = \begin{bmatrix} g_1(a(0), \beta_0) \\ g_2(a(0), \beta_0, \omega) \end{bmatrix}, \tag{4.17}$$

where the estimating equations $g_1(a(0), \beta_0) = 0$ result from $\text{vech}\left(\Sigma - A(0)A(0)'\right)$ and the q_2 (functional) estimating equations $g_2(a(0), \beta_0, \omega) = 0$, which stem from the frequency identifying restrictions, can be written as:

$$\text{Hvec}\left(A(e^{-i\omega}) - C(e^{-i\omega})A(0)\right) = \mathbf{0_{q_2 \times 1}} \tag{4.18}$$

or

$$\text{H}\left(I_N \otimes C(e^{-i\omega})\right) a(0) = b(e^{-i\omega}),$$

where H is an $q_2 \times N^2$ selection matrix, $a(0) = \text{vec}(A(0))$, and $b(e^{-i\omega}) = \text{Hvec}(A(e^{-i\omega}))$.

Starting from the (functional) system of estimating equations (4.17), the first-step C-ALS estimator with the identity operator is given by the following minimization problem:

$$\widehat{a}_T^1 \quad = \quad \arg\min_a \left[g_1(a(0), \widehat{\beta}_T)' g_1(a(0), \widehat{\beta}_T) + \int_{\underline{\omega}}^{\overline{\omega}} g_2(a(0), \widehat{\beta}_T, \omega)' \overline{g_2(a(0), \widehat{\beta}_T, \omega)} d\omega \right].$$

Proposition 4.1 establishes the main results irrespective of the selection of the estimating equations, H.

[19] An online technical appendix shows how these results can be applied in the case of the identification of a single structural shock in any N-variate VAR model. We also provide more detailed results in the bivariate case.

Proposition 4.1. *Consider the reduced-form VAR(p) given by equation (2.1) and the vector of estimating equations defined by equation (4.17) for a given frequency interval $\omega \in [\underline{\omega}, \overline{\omega}]$. Suppose that the moments of order three of u_t are zero. Let $\beta = (vec(\Phi_p)', vech(\Sigma_u)')' \equiv (\Phi', \sigma')'$ denote the vector of reduced-form parameters, and $\Omega_T = \begin{pmatrix} \Omega_\Phi & \mathbf{0} \\ \mathbf{0} & \Omega_\sigma \end{pmatrix}$ the corresponding partitioning of the asymptotic variance–covariance matrix of the OLS estimator of β. Then,*

- *The first-step C-ALS estimator of $a(0)$, denoted by $\widehat{a}_T^1$, solves:*

$$\int_{\underline{\omega}}^{\overline{\omega}} \left(I_N \otimes C^\star(e^{-i\omega})\right) \mathrm{H}'\mathrm{H} \left(I_N \otimes C(e^{-i\omega})\right) d\omega vec(A(0)) - 2\left(D_N^+ \left(A(0) \otimes I_N\right)\right)' vech\left(\widehat{\Sigma}_T - A(0)A(0)'\right) = \mathbf{0},$$

  where $D_N^+ = (D_N' D_N)^{-1} D_N'$ and D_N is the $N^2 \times \frac{1}{2}N(N+1)$ duplication matrix such that $vec(X) = D_N vech(X)$.

- *The vector of estimating equations in the second step is given by:*

$$\underline{s}(a(0), \widehat{\beta}_T) = \begin{pmatrix} g_1(a(0), \widehat{\beta}_T) \\ \Omega_\Phi^{1/2} \int_{\underline{\omega}}^{\overline{\omega}} \overline{\frac{\partial g_2'}{\partial \Phi}\left(\widehat{a}_T^1, \widehat{\beta}_T, \omega\right)} g_2(a(0), \widehat{\beta}_T, \omega) d\omega \end{pmatrix}.$$

- *The second-step C-ALS estimator, denoted by $\widehat{a}_T$, solves:*

$$\widehat{a}_T = \arg\min_a \left[g_1(a(0), \widehat{\beta}_T)' W_{1T} g_1(a(0), \widehat{\beta}_T) + \underline{s_2}(a(0), \widehat{\beta}_T)' \tilde{W}_{2T}^2 \underline{s_2}(a(0), \widehat{\beta}_T)\right],$$

  where W_{1T} is the inverse of $2D_N^+(\widehat{\Sigma} \otimes \widehat{\Sigma}) D_N^{+\prime}$, $\underline{s_2}\left(a(0), \widehat{\beta}_T\right)$ is the second set of transformed estimating equations, and $\tilde{W}_{2T}$ is the generalized inverse of W_{2T}:

$$W_{2T} = \Omega_\Phi^{1/2} \int_{\underline{\omega}}^{\overline{\omega}} \overline{\frac{\partial g_2'}{\partial \Phi}\left(\widehat{a}_T^1, \widehat{\beta}_T, \omega\right)} \frac{\partial g_2}{\partial \Phi'}\left(\widehat{a}_T^1, \widehat{\beta}_T, \omega\right) d\omega \Omega_\Phi^{1/2}.$$

Proof: See the Appendix.

The first-step C-ALS estimator can be easily obtained from the first-order conditions and only requires to solve N^2 nonlinear equations with respect to the parameters of interest $a(0) = \mathrm{vec}(A(0))$. Interestingly, for a given ω, the two terms $(I_N \otimes C(z)')\mathrm{H}'$ and $D_N^+ \left(A(0) \otimes I_N\right)$ are those that provide the rank condition and thus the local identification of structural VARs.[20] Meanwhile, taking the block diagonal structure of the optimal variance–covariance matrix when the third-order moments of u_t are zero, the second-step C-ALS estimator can also be derived in a straightforward way by numerically solving either the minimization problem or the first-order conditions. In both cases, there is no need for numerical integration. Finally, the vector of estimating equations in the second step involves the optimal projection of $g_2(\widehat{a}_T^1, \widehat{\beta}_T, \omega)$ onto the subspace spanned by the reduced-form estimates of the autoregressive parameters and the standard $N(N+1)/2$ estimating equations, $g_1(a(0), \widehat{\beta}_T)$, that defines the mapping between the variance–covariance

[20] See Proposition 9.4. in Lütkepohl (2007).

of the innovations and the structural shocks.

As an application of Proposition 4.1, suppose that we consider a bivariate model and that the second structural shock has no impact on the first variable on a frequency band $[\underline{\omega}, \overline{\omega}]$ with $\underline{\omega} = -\overline{\omega}$. In particular, the matrix $A(0)$ is locally identified (up to a sign restriction) when

$$\begin{aligned} g_2(\tilde{a}_{12}, \widehat{\beta}_T, \omega) &= \widehat{c}_{11}(e^{-i\omega})\tilde{a}_{12} + \widehat{c}_{12}(e^{-i\omega}) = 0 \\ &= \sum_{j=0}^{\infty} \left[\widehat{c}_{11,j} e^{-i\omega j} \tilde{a}_{12} + \widehat{c}_{12,j} e^{-i\omega j}\right] = 0, \end{aligned}$$

where $\tilde{a}_{12}(0) = a_{12}(0)/a_{22}(0)$. In this case, one can proceed sequentially, that is, one first determines an efficient two-step ALS estimate of $\tilde{a}_{12}(0)$, denoted by $\widehat{\tilde{a}}_{12,T}$, using only the previous functional estimating equation and then obtains an estimate of $a(0)$ using the first set of just-identified estimating equations $g_1(a(0), \beta_0) \equiv \text{vech}\left(\widehat{\Sigma}_T - A(0)A(0)'\right) = 0$ after replacing $a_{12}(0)$ with $\widehat{\tilde{a}}_{12,T} a_{22}(0)$. More specifically, using the identity operator, the first-step C-ALS estimator of $\tilde{a}_{12}(0)$ is given by:

$$\widehat{\tilde{a}}^1_{12,T} = -\frac{\sum\limits_{j=0}^{\infty}\sum\limits_{l=0}^{\infty} \widehat{c}_{11,j}\widehat{c}_{12,l} \int_{\underline{\omega}}^{\overline{\omega}} \cos((l-j)\omega) d\omega}{\sum\limits_{j=0}^{\infty}\sum\limits_{l=0}^{\infty} \widehat{c}_{11,j}\widehat{c}_{11,l} \int_{\underline{\omega}}^{\overline{\omega}} \cos((l-j)\omega) d\omega}. \tag{4.19}$$

This first-step estimator corresponds to the Min-effect/Max-effect frequency estimator proposed by Wen (2001, 2002). Meanwhile, imposing the identifying restriction at $\omega = 0$ (Blanchard and Quah, 1989) yields: $\widehat{a}_{12,T} = -\frac{\widehat{c}_{12}(1)}{\widehat{c}_{11}(1)}$. The one-step C-ALS estimator can be seen as a generalized least squares estimator in which the weights (i.e., the cosine terms) capture not only the information at the zero frequency but also its neighborhood. Finally, using the simplified expression of the objective function and a consistent estimate of W_{2T}, the efficient second-step C-ALS estimator is:[21]

$$\widehat{\tilde{a}}_{12,T} = -\frac{\widehat{s}'_{11,T}(\widehat{W}^2_{2T})^{-1}\widehat{s}_{12,T}}{\widehat{s}'_{11,T}(\widehat{W}^2_{2T})^{-1}\widehat{s}_{11,T}},$$

where $\widehat{s}_{1k,T}$ for $k = 1, 2$ and $\widehat{W}_{2T}$ are given by:

$$\widehat{s}_{1k,T} = \widehat{\Omega_{\Phi}}'^{1/2} \sum_{j=0}^{\infty}\sum_{l=0}^{\infty} \left[\left(\frac{\partial \widehat{c}_{11,j}}{\partial \Phi}\widehat{\tilde{a}}^1_{12,T} + \frac{\partial \widehat{c}_{12,j}}{\partial \Phi}\right)\widehat{c}_{1k,l}\right] \int_{\underline{\omega}}^{\overline{\omega}} \cos((l-j)\omega) d\omega,$$

and

$$\widehat{W}_{2T} = \widehat{\Omega}'^{1/2}_{\Phi} \left[\sum_{j=0}^{\infty}\sum_{l=0}^{\infty} \left(\frac{\partial \widehat{c}_{11,j}}{\partial \Phi}\widehat{\tilde{a}}^1_{12,T} + \frac{\partial \widehat{c}_{12,j}}{\partial \Phi}\right)\left(\frac{\partial \widehat{c}_{11,l}}{\partial \Phi'}\widehat{\tilde{a}}^1_{12,T} + \frac{\partial \widehat{c}_{12,l}}{\partial \Phi'}\right) \int_{\underline{\omega}}^{\overline{\omega}} \cos((l-j)\omega) d\omega\right] \widehat{\Omega}^{1/2}_{\Phi},$$

where $\Phi = vec(\Phi_p)$.

[21] The proof is available on the online Appendix.

5 Monte Carlo simulations

In this section, we provide some Monte Carlo simulations to study the finite sample performances of the C-ALS estimator. We assume that the data generating process (DGP) is a bivariate VAR model in which the first variable, $X_{1,t}$, is nonstationary and thus written in first-difference and the second variable, $X_{2,t}$, is a weakly stationary process:

$$\Delta X_{1,t} = \vartheta_1 + \rho_{11,1}\Delta X_{1,t-1} + (\rho_{12,1} + \delta)X_{2,t-1} - \rho_{12,1}X_{2,t-2} + \epsilon_{1,t} \tag{5.20}$$

$$X_{2,t} = \vartheta_2 + \rho_{21,1}\Delta X_{1,t-1} + \rho_{22,1}X_{2,t-1} + \rho_{22,2}X_{2,t-2} + b_{21}\epsilon_{1,t} + \epsilon_{2,t}, \tag{5.21}$$

where the vector $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ represents some structural shocks, with $\epsilon_t \sim N(0, I_2)$. Since the presence of deterministic terms might contaminate the estimation at $\omega = 0$ and thus the finite sample properties, we include one intercept term in each equation and calibrate them so that to match the unconditional means of the variables of interest in our application (Section 6). The parameter δ controls the magnitude of the long-run effect of the second shock $\epsilon_{2,t}$ on the first variable $X_{1,t}$. When $\delta = 0$, only the first shock has a long-run impact on the first variable. To some extent, the corresponding specification can be viewed as the one often encountered in the macro literature in order to identify a permanent shock, for example, the identification of a technology shock with some measures of (labor or total) productivity and hours worked (see Section 6). It is worth emphasizing that the VAR(1) specification (the first set of experiments) is the DGP of Gospodinov et al. (2013) and Chevillon et al. (2020).

Using equations (5.20) and (5.21), we generate 10,000 samples of size $T = 200$ observations—a sample size often encountered in applied macro works—and the effect of initial conditions is controlled by including 200 pre-sampled observations that are subsequently discarded in the estimation. For each repetition, the lag order is set to its true value so that results are interpreted free of any lag order misspecification issue.[22] Our method, denoted by `C-ALS` is compared with three approaches. The first one, denoted by `LR`, is a standard long-run identification scheme à la Blanchard–Quah, that is, we only impose the identification constraint at $\omega = 0$.[23] The second alternative is the first-step C-ALS estimator defined in equation (4.19) when the kernel operator is the identity operator. The last alternative is the max-share procedure of DiCeccio and Owyang (2010), Dieppe et al. (2021) and Francis et al. (2014), denoted by `MS`, which has been used in recent contributions (e.g., Angeletos et al., 2020).

With the exception of the `LR` method, we consider four fixed symmetric frequency intervals $\omega_n = [-\frac{2\pi}{n}, \frac{2\pi}{n}]$ with $n = 30, 60, 90$, and 120 quarters. The results are then assessed along three dimensions. First, we compute the contemporaneous effect of each structural shock on each variable (that is, the impulse response function at $h = 0$) and determine the corresponding mean absolute bias and root mean squared

[22]Several robustness exercises, which are available upon request, have been experimented with to control for the lag order selection. Overall, our results remain unchanged, and our estimator performs better than competing estimators.

[23]We also implement the methodology of Christiano et al. (2006b), that is, a nonparametric approach to estimate the zero-frequency spectral density (with a Bartlett or Andrews-Monahan kernel). However, our Monte Carlo results show that their approach underperforms with respect to the max-share approach and our C-ALS procedure.

error (RMSE). Second, we provide the cumulative mean absolute bias and RMSE for $h \in [0, H]$, with $H = 4$, 8, and 12 by using the impulse response functions.[24] More specifically, the cumulative mean absolute bias is defined as $cmd(H) = \sum_{h=0}^{H} |\texttt{irf}_h(\texttt{model}) - \texttt{irf}_h(\texttt{svar})|$ where H denotes the selected horizon, $\texttt{irf}_h(\texttt{model})$ the impulse response at horizon h from the model defined by equations (5.20) and (5.21), and $\texttt{irf}_h(\texttt{svar}) = (1/N) \sum_{j=1}^{N} irf_{h,j}(\texttt{svar})$ the average impulse response function over the N simulation experiments. In other words, the cumulative mean absolute bias is a measure of the area between the impulse response function up to a given horizon H and the horizontal axis. Third, we contrast the true impulse response function of the second variable relative to the first structural shock with the impulse response function estimates of the competing methods.

We first consider a VAR(1) specification with $(\rho_{11,1}, \rho_{12,1}, \rho_{21,1}, \rho_{22,1}, \rho_{22,2}, b_{21}, \delta) = (0, 0, 0.2, \rho, 0, 0.2, 0)$ where $\rho = 0.9$, 0.95, or 0.98. When $\rho_{11,1} = \rho_{12,1} = \delta = 0$, the first variable, $X_{1,t}$, is a random walk, and the second variable is a persistent stationary process driven by ρ. Moreover, $\delta = 0$ implies that only the first structural shock has a permanent effect on the first variable and that the identifying restriction is correctly specified (null hypothesis). In particular, the variance contribution of the second structural shock to the first variable is equal to zero irrespective of the frequency interval under consideration, and the long-run restriction is always satisfied irrespective of the frequency interval ω_n. Figure 1 reports the mean absolute bias (left panel) and RMSE (right panel) of the contemporaneous effect of each structural shock on each variable for different frequency intervals ω_n.

Three points are worth commenting. First, inspecting Figure 1 shows that the mean absolute bias and RMSE curves of the frequency-based approaches are below the solid line that represents the results of the `LR` approach. This also holds true when compared to the `MS` approach. Second, the second-step `C-ALS` approach outperforms other methods for both statistical criteria, and, in particular, the mean absolute bias differences between the second-step `C-ALS` estimator and other alternatives are substantial irrespective of the frequency. The performance of the second-step C-ALS estimator in terms of bias and RMSE improves monotonically and impressively in the range $[0; 0.05]$, whereas gains are mostly negligible for higher frequencies. This is further illustrated by the discrepancy between the first-step C-ALS and the second-step C-ALS: there is an effective gain to exploit the relevant information of the reduced-form VAR estimation through the optimal weighting matrix. Notably, the second-step C-ALS estimator captures the persistence of the transitory shock, driven by ρ, through the system of transformed estimating equations, whereas there is a confounding effect between the permanent shock and the persistent transitory shock in the absence of the optimal projection into the space spanned by the reduced-form estimates, hence leading to a less substantial bias and RMSE reduction for the first-step C-ALS estimator. Overall, the `C-ALS` estimator leads to a significant bias reduction while being more efficient. Third, as to be expected, the first-step C-ALS and max-share estimators display roughly the same finite sample properties under the null hypothesis of a well-specified identifying restriction. Indeed, the objective function of the first-step estimator is equivalent to minimize (respectively,

[24]As our results are qualitatively the same irrespective of the horizon, we only report those at the impact and those at $H = 12$. Other results are available upon request.

to maximize) the contribution of the second (respectively, the first) structural shock to the (marginal) spectrum of the first variable. Under the null hypothesis of a well-specified identifying restriction, it amounts to finding the largest eigenvalue of the forecast error variance decomposition, which is the purpose of the max-share approach.

Regarding the cumulative absolute bias between the average response in SVARs and the true response, and the cumulative RMSE up to 12 periods for $n = 30, 60, 90$, and 120, Figure 2 provides very supportive evidence for the `C-ALS` approach. In particular, the cumulative bias and RMSE performances of our two-step procedure are better than those of the competing approaches when studying the effect of each structural shock on each variable of interest. Notably, the `C-ALS` estimator displays less cumulative bias and RMSE for tiny intervals around $\omega = 0$. Unreported results for $\rho = 0.9$ and $\rho = 0.98$ lead to the same conclusions.

To further contrast the different approaches, Figure 3 displays the true and estimated impulse response function of the second variable relative to the first structural shock as well as the confidence intervals in the case of the frequency interval $\omega_{60} = [-\frac{2\pi}{60}, \frac{2\pi}{60}]$.[25] Interestingly, the impulse responses for the LR restriction mimics the empirical results for the impact of a technology shock on hours worked when the hours series is specified in level in a VAR (see Christiano et al., 2006a); the response is positive at the impact and declines toward zero, and the confidence interval contains zero at all horizons. This is also the case for the first-step and `MS` estimators that display narrower confidence intervals than the one of the `LR` method. However, the `C-ALS`-based impulse response function is more precise, and one can reject the hypothesis that the effect of the first shock on the second variable is equal to zero up to $H = 20$. Therefore, by computing the optimal weighting matrix, there is a huge gain of efficiency relative to the first-step C-ALS and the max-share approach. Finally, we implement (the regularized version of) the overidentification test (Proposition 3.3).[26] As reported in panels a, b, and c of Table 1, when $\delta = 0$, the test is conservative under the null hypothesis irrespective of the frequency interval.

To evaluate the robustness of the previous results, we now proceed with a misspecified exclusion restriction (alternative hypothesis) in the sense that both shocks have a permanent effect on the first variable ($\delta \neq 0$) whereas we only impose that the first structural shock matters permanently for the first variable. As reported in Table 1, when $\delta = 0.05$ or 0.1 and n increases and thus the length of the frequency band decreases, the proportion of the variance explained by the second structural shock for the first variable increases. This can be observed in Figure 4a), which displays the (marginal) spectral densities of the first variable resulting from the first and the second structural shocks. In this respect, Figure 5 supports that our

[25]The results are qualitatively the same for ω_{30}, ω_{90}, and ω_{120}.

[26]One issue is the regularization of the W matrix and the fact that the test cannot be implemented if the original weighting matrix displays a large rank deficiency (i.e., the proportion of smaller eigenvalues is too large). More specifically, as the length of the symmetric frequency interval decreases, the matrix W might not be of full rank so that we make use of a generalized inverse through a regularization method. To circumvent this issue, we determine the rank of the matrix, say k, and take as the regularization parameter the kth (ordered) eigenvalue. All in all, our worst case leads us to reject 4 simulations out of 10,000 for the frequency interval ω_{120}. We apply the same procedure for the application in the next section.

methodology clearly outperforms other methods in terms of mean absolute bias and RMSE.[27] Here again, the reduction of both the bias and the RMSE is quite substantial, and it is worth emphasizing that the first-step C-ALS estimator has better finite samples properties than the max-share estimator in the presence of misspecification, irrespective of ρ. Moreover, as already observed in the benchmark parameter vector, Figure 6 shows that the discrepancy between the true impulse response function and the one obtained from the `C-ALS` estimator is rather small, whereas those of other methods display a significant and substantial bias at very short horizons—the IRF estimates being even below the lower bound of the confidence band for the first five quarters —and also in the medium term. This relatively good performance of the `C-ALS` estimator can be explained by the fact that the minimization solution yields a pseudo-true value getting closer to the true value when the frequency interval becomes wider and thus there is a decreasing relative contribution of the second shock to the spectral density of the first variable as observed in Figure 4a). Regarding the overidentification test, panels a,b, and c of Table 1 show that it performs well. As δ increases, the contribution of the second structural shock to the first structural variable increases and so the rejection rate, the probability of rejecting the null hypothesis being between 90% and 99% for $\delta = 0.1$.

As a second set of experiments, we consider other parameter configurations. For instance, as shown in the online Appendix, using a VAR(2) specification with $(\rho_{11,1}, \rho_{12,1}, \rho_{21,1}, \rho_{22,1}, \rho_{22,2}, b_{21}) = (0, -0.08, 0.2, \rho + 0.55, -0.55\rho, 0.2)$ under the null ($\delta = 0$) and alternative ($\delta > 0$) hypothesis where $\rho = 0.9$, 0.95, or 0.98, our results are qualitatively similar. We also compare the discretized ALS estimators with the corresponding C-ALS estimators. The simulation results (not reported here) show that the second-step discretized ALS estimator performs better than the `LR` and `MS` estimators in terms of bias and RMSE but is very slightly dominated by the optimal two-step C-ALS estimator. This is probably due to the excessive smoothness of the marginal spectral densities and of the cross-spectrum, which is implied by our DGP and can be easily replicated using a discretization procedure.[28]

Finally, we evaluate the small sample performances of the data-driven procedure for the selection of the frequency interval (Section 4.4). In so doing, we first consider a DGP where the first structural shock does not contribute to the spectral density of the second variable at any frequency. In this case, the $VICS_T(\omega)$ criterion should select the widest possible interval. Then, as a local alternative, the first structural shock is assumed to contribute to the marginal spectral density of the second variable only in an interval strongly concentrated around the zero frequency (Figure 4b). In this case, the $VICS_T(\omega)$ criterion should select the narrowest possible interval around zero. In both cases, we make use of symmetric intervals $\omega_n = [-\frac{2\pi}{n}, \frac{2\pi}{n}]$ where $n = 30, 40, 50, \ldots, 240$ and of a narrow interval with $n = 2000$ to mimic a long-run restriction. Our results show that the $VICS_T(\omega)$ test statistic based on a SIC penalty term selects the largest interval with a probability of 0.757 in the first case whereas the narrowest interval is recovered with a probability of 0.668

[27]The results for the cumulative mean absolute bias and cumulative RMSE are very similar but not reported here for sake of space.

[28]In the online Appendix, we provide theoretical results about the asymptotic distribution of C-ALS estimators and overidentification tests in the presence of local misspecification.

in the second case. It turns out that our data-driven interval selection procedure performs well for (finite) sample sizes encountered in practice and even better when the sample size gets larger and larger.

To summarize, our Monte Carlo simulations provide evidence that the two-step C-ALS estimator outperforms other methods in terms of both (cumulative) mean absolute bias and RMSE. Contrasting the true impulse responses with those of the competing methods shows that the two-step `C-ALS` estimator is more reliable and precise. At the same time, the proposed J-test behaves nicely in the presence of local alternatives and misspecified identifying restrictions.

6 Application: The hours–productivity debate using bivariate SVAR models

We now discuss an application regarding the technology–hours debate in light of the contribution of Francis and Ramey (2009). Indeed, structural VAR models yield conflicting results regarding the effect of technology shocks on hours. Consequently, the predominant role of technology shocks as the main source behind movements in macro data has been sharply challenged since the appraisal of Galí (1999). One key issue of the technology–hours debate is the assumed DGP for the hours worked (per capita) measures. Using a first difference specification of hours, structural VAR models predict a decline in hours in response to a positive technological shock (e.g., Galí, 1999, or Francis and Ramey, 2005), opposite of what is implied by Real Business Cycles models.[29] However, entered in level, hours increase in response to a positive technological shock, and the standard result at the core of the long-standing RBC model emerges (Christiano et al, 2006a). To go one step further, Francis and Ramey (2009) argue that one potential explanation for these conflicting results is that the standard measure of hours per capita and productivity have significant low-frequency movements, and these movements can lead to misleading results in the level-based specification of a structural VAR model.

More specifically, Francis and Ramey (2009) show that demographic trends and sectoral allocation are important sources of low-frequency movements in hours worked and labor productivity.[30] Accordingly, labor productivity might be driven by two permanent shocks, the technology shock and the demographic shock, and thus the usual long-run restriction of hours-productivity VAR models might be violated. To circumvent this issue, Francis and Ramey (2009) propose new measures of hours worked per capita and labor productivity that are more comfortable with the imposed long-run restriction(s). Using the adjusted series, it turns out that the response of hours worked is now negative in the short run, and then becomes slightly positive after a year for a structural VAR model in which the adjusted hours worked per capita variable is specified in level. In this respect, a more complete test of their results begs the following questions: Is there any

[29] While standard unit root tests cannot reject the presence of a unit root for hours worked series, most dynamic macroeconomic models with standard preference specifications imply that the hours worked per capita should be stationary in the absence of permanent structural changes in government spending, labor income taxes, and preferences (see Francis and Ramey, 2009).

[30] Several strands of research have discussed the existence of alternative shocks that can result in permanent effects on labor productivity (e.g., Mertens and Ravn, 2013; Fisher, 2006; Ben Zeev and Kahn, 2015).

evidence that only technology shocks have a long-run effect on labor productivity using unadjusted hours and productivity measures? If not, how effective is the technology shock identified with the adjusted series?

To this end, we conduct structural bivariate VAR analysis in which the first variable is labor productivity and the second variable (in level) is subsequently the standard hours per capita measure (private business hours per capita) and the adjusted hours series constructed by Francis and Ramey (2009). Starting from the two-step procedure in the bivariate case presented in Section 4, we implement the overidentification test to assess the reliability of the identifying restrictions, and thus proceed as follows. We estimate the two reduced-form VAR models in which the hour series is in level. As in Francis and Ramey (2009), the sample period is 1948Q1–2007Q4 and the lag order is set to 4. To compare the results of our approach with those of Francis and Ramey (2009), the identifying constraints are imposed over the frequency interval $\omega_{120} = [-\frac{2\pi}{120}, \frac{2\pi}{120}]$. Finally, confidence intervals at 95% are built from a bootstrap procedure with 1,000 replications.

As reported in Figure 7, using the standard `LR` restriction, the two graphs show that (unadjusted) private business hours per capita respond significantly, with the exception of the initial period, and positively in the short run to a positive technological shock and then decreases at intermediate to long horizons. In contrast, using our approach, (unadjusted) private business hours per capita initially decrease, and then respond positively in the short run (after one year) before gradually decreasing toward zero in the medium-to-long term. Moreover, none of the effect of the technological shock is statistically different from zero. As pointed out by Francis and Ramey (2009), one explanation for this apparent discrepancy is that the identifying assumption, namely that the technological shock alone explains the long-run effect on labor productivity for the unadjusted hours series, is misspecified. To shed some light on this issue, we perform our identification test and find that the J_T statistic has a p-value of 0.0005. Consequently, our proposed overidentification test clearly rejects the hypothesis that only one shock has a permanent effect on the labor productivity when using the unadjusted series of hours.

As reported in the two graphs at the bottom of Figure 7, both methods lead to the same shape of the impulse response function, with the exception of the initial effect, when the VAR specification contains the adjusted series of hours. More specifically, there is a statistically significant negative effect of the technological shock on (adjusted) hours worked over the first periods in the case of our methodology, whereas those effects are not statistically different from zero using the standard `LR` method. Note that the `LR` results are consistent with those of Francis and Ramey (2009). Interestingly, the J_T statistic now has a p-value of 0.6512. In other words, this provides some support of the argument of Francis and Ramey (2009); the adjusted hours worked series for demographic and sectoral changes is now compatible with the hypothesis that only the technology shock has a long-run effect on labor productivity. We also conduct our data-driven procedure to select the optimal frequency interval such that the imposed restrictions do hold. We find significant evidence for $\hat{\omega}_T = 80$ quarters (with a p-value of 0.4307), and this provides additional support for the previous results with $\hat{\omega}_T = 120$.[31] Finally, the right panels of Figure 7 also show that the impulse response functions derived from the discretized ALS estimator are close to but slightly different from the ones of the C-ALS estimator.

Therefore, to answer our two questions, the evidence that only the technological shock has a long-run effect on labor productivity is weak, and correcting the hours series for demographic and sectoral changes is more consistent with the Blanchard–Quah long-run restriction and leads to a negative effect of a technological shock in the short run.

7 Conclusion

In this paper, we propose a joint methodology for the identification and inference of structural VAR models in the presence of frequency identifying restrictions. Using the methodology of Carrasco and Florens (2000) and Carrasco et al. (2007a), and the ALS procedure of Gouriéroux et al. (1985), we derive a C-ALS estimator that allows obtaining reliable estimates of the dynamic responses of macroeconomic variables to structural shocks and formally assessing the relevance of the imposed restrictions over either a given set of frequencies or a data-driven selected interval. Monte Carlo simulations argue in favor of our approach with respect to competing methods. Finally, our application regarding the hours–productivity debate provides some new insights and highlights the relevant argument of Francis and Ramey (2009).

From an empirical point of view, our methodology and the associated testing procedure (overidentification, interval selection) can be used to reassess several debates, such as the identification and reliability of news' shocks (Beaudry and Portier, 2006; Barsky and Sims, 2011; Kurmann and Sims, 2021), the assessment of the long-run neutrality (super-neutrality) of money or the long-run Fisher relation, or the identification and estimation of the main driver (Angeletos et al., 2020). However, the derivation of optimal rules for the choice of the regularization parameter for testing procedures, the existence of nonfundamental representations (Gouriéroux et al., 2020) and the recoverability condition (Chahrour and Jurado, 2022) deserve further research.

[31] In contrast, the p-value of the J-stat is 0.0433 when $\hat{\omega}_T = 60$ quarters.

References

[1] Amengual, D., Carrasco, M., and E. Sentana, 2020, Testing distributional assumptions using a continuum of moments. Journal of Econometrics vol. 218(2), 655-689.

[2] Andrews, D.W.K., 1987, Asymptotic Results for Generalized Wald Test. Econometric Theory, vol. 3, 348-358.

[3] Angeletos, G-M., Collard, F., Dellas, H., 2020. Business-Cycle Anatomy. American Economic Review vol. 110(10), 3030-3070.

[4] Babii, A., Florens, J.P., 2021. Is Completeness Necessary? Estimation in Nonidentified Linear Models. Toulouse School of Economics.

[5] Barsky, R.B., Sims, E.R., 2011. News Shocks and Business Cycles. Journal of Monetary Economics vol. 58(3), 273-289.

[6] Beaudry, P., Portier, F., 2006. Stock Prices, News, and Economic Fluctuations. American Economic Review vol. 96(4), 1293-1307.

[7] Ben Zeev, N., Khan, H., 2015. Investment-specific News Shocks and U.S. Business Cycles. Journal of Money, Credit, and Banking vol. 47(7), 1443-1464.

[8] Bernanke, B., 1986. Alternative Explanations of the Money-Income Correlation. Carnegie Rochester Conference Series on Public Policy vol. 25(0), 49-99.

[9] Bernanke, B., Mihov, I., 1998. Measuring Monetary Policy. Quarterly Journal of Economics vol. 113(3), 869-902.

[10] Blanchard, O.J., Quah, D., 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances. American Economic Review vol. 79, 655-673.

[11] Carrasco, M., 2012. A Regularization Approach to the Many Instruments Problem. Journal of Econometrics vol. 170(2), 383-398

[12] Carrasco, M., Chernov, M., Florens, J.P., Ghysels, E. 2007a. Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions. Journal of Econometrics vol. 140(2), 529-573.

[13] Carrasco, M., Florens, J.P., Renault, E., 2007b. Linear Inverse Problems in Structural Econometrics: Estimations based on Spectral Decomposition and Regularization, edited in Handbook of Econometrics, vol. 6 (chapter 77), part B, 5633-5751.

[14] Carrasco, M., Florens, J.P., 2000. Generalization of GMM to a continuum of moment conditions. Econometric Theory vol. 16, 797-834.

[15] Chahrour, R., Jurado, K., 2022. Recoverability and Expectations-Driven Fluctuations. The Review of Economic Studies vol. 89(1), 214239

[16] Chevillon, G., Mavroeidis, S., Zhan, Z., 2020. Robust Inference in Structural Vector Autoregressions with Long-Run Restrictions. Econometric Theory vol. 36, 86-121.

[17] Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2006a. Assessing Structural VARs. NBER Macroeconomics Annual vol. 21, 1-106.

[18] Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2006b. Alternative Procedures for Estimating Vector Autoregressions Identified with Long-run Restrictions. Journal of the European Economic Association vol. 4(2-3), 475-483.

[19] Cooley, T., LeRoy, S., 1985. Atheoretical Macroeconomics: A Critique. Journal of Monetary Economics vol. 16(3).

[20] DiCecio, R., Owyang, M.T., 2010. Identifying technology shocks in the frequency domain. Working Papers 2010-025, Federal Reserve Bank of St. Louis.

[21] Dieppe, A., Neville, F., Kindberg-Hanlon, G., 2021. The identification of dominant macroeconomic drivers: coping with confounding shocks. ECB Working Paper Series No 2534(April).

[22] Faust, J., 1996. Near Observational Equivalence and Theoretical Size Problems with Unit Root Tests. Econometric Theory vol. 12, 724-731.

[23] Faust, J., 1998. The Robustness of Identified VAR Conclusions about Money. Carnegie-Rochester Conference Series ob Public Policy vol. 49(0), 207-44.

[24] Faust, J., 1999. Conventional Confidence Intervals for Points on Spectrum Have Confidence Level Zero. Econometrica vol. 67, 629-637.

[25] Faust, J., Leeper, E.M., 1997. When Do Long-Run Identifying Restrictions Give Reliable Results? Journal of Business & Economic Statistics vol. 15, 345-353.

[26] Fisher, J.D.M., 2006. The Dynamic Effects of Neutral and Investment-specific Technology Shocks. Journal of Political Economy vol. 114(3), 413-451.

[27] Francis, N., Owyang, M.T., Roush, J.E., DiCecio, R., 2014. A Flexible Finite-Horizon Alternative to Long-Run Restrictions with an Application to Technology Shocks. The Review of Economics and Statistics vol. 96(3), 638-647.

[28] Francis, N., Ramey, V., 2009. Measures of Per Capita Hours and Their Implications for the Technology-Hours Debate. Journal of Money, Credit and Banking vol. 41(6), 1071-1097.

[29] Francis, N., Ramey, V., 2005. Is the Technology-driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited. Journal of Monetary Economics vol. 52, 1379-1399.

[30] Galí, J., 1999. Technology, Employment and the Business Cycle: So Technology Shocks Explain Aggregate Productivity? American Economic Review vol. 41, 1201-1249.

[31] Gospodinov, N., Herrera, A.M., Pesavento, E., 2013. Unit Roots, Cointegration and Pretesting in VAR Models. Advances in Econometrics vol. 32, 81115

[32] Gouriéroux, C., Monfort, A., 1995. Statistics and Econometric Models: Volume 1, Cambridge University Press.

[33] Gouriéroux, C., Monfort, A., Renne, J.P., 2020. Identification and Estimation in Non-fundamental Structural VARMA models. Review of Economic Studies vol. 87, 1915-1953.

[34] Gouriéroux, C., Monfort, A., Trognon, A., 1985. Moindres Carrés Asymptotiques. Annales de l'INSEE vol. 58, 91-121.

[35] Groetsch, C., 1984, The Theory of Tikkonov Regularization for Fredholm Equations of the First Kind, Pitman, London.

[36] Hall, A.R., Inoue, A., Nason, J.M., Rossi, B., 2012. Information criteria for impulse response function matching estimation of DSGE models. Journal of Econometrics vol. 170(2), 499-518.

[37] Hamilton, J.D. 1994. Time Series Analysis. Princeton University Press.

[38] Hansen, L.P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica vol. 50, 1029-1054.

[39] Hauser, M.A., Pötscher, B.M., Reschenhofer, E., 1999. Measuring Persistence in Aggregate Output: ARMA Models, Fractionally integrated ARMA Models and Nonparametric Procedures. Empirical Economics vol. 24, 243-269.

[40] Imhof, J.P., 1961. Computing the Distribution of Quadratic Forms in Normal Variables. Biometrika vol. 48(3-4), 419-426.

[41] Kailath, T., 1971. An RKHS Approach to Detection and Estimation Problems, Pt. I: Deterministic Signals in Gaussian Noise. IEEE Trans. Information Theory vol. 17(5), 530-549.

[42] Kilian, L., 2013. Structural Vector Autoregressions, in N. Hashimzade and M.A. Thornton (eds.), Handbook of Research Methods and Applications in Empirical Macroeconomics, Cheltenham, UK Edward Elgar, 2013, pp. 515-554.

[43] Kilian,L., Lütkepohl, H., 2017. Structural Vector Autoregressive Analysis, Cambridge University Press.

[44] King, R. G., Watson, M.W., 1996. Money, Prices, Interest Rates and the Business Cycle. The Review of Economics and Statistics vol. 78, 35-53.

[45] Kurmann, A., Sims, E., 2021. Revisions in Utilization-Adjusted TFP and Robust Identification of News Shocks. The Review of Economics and Statistics vol. 103(2), 216235

[46] Luenberger, D.G., 1997, Optimization by Vector Space Methods, Wiley professional paperback series.

[47] Lütkepohl, H., 2007. New Introduction to Multiple Time Series Analysis, Springer-Verlag, Berlin.

[48] Mertens, K., Ravn, M.O., 2013. The Dynamics Effects of Personal and Corporate Income Tax Changes in the United States. American Economic Review vol. 103(4), 1212-1247.

[49] Pesendorfer, M., Schmidt-Dengler, P., 2008. Asymptotic Least Squares Estimators for Dynamic Games. The Review of Economic Studies vol. 75, 901-928.

[50] Phillips, P.C.B. (1998), "Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs", *Journal of Econometrics*, vol. 83, 21-56.

[51] Pötscher, B.M., 2002. Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots, and Estimation of Long Memory Parameters. Econometrica vol. 70, 1035-1065.

[52] Robin, J.M., Smith, R.J., 2000. Tests of Rank. Econometric Theory vol. 16(2), 151-175.

[53] Sims, C. (1971), "Distributed Lag Estimation When the Parameter Space is Explicitly Infinite-Dimensional", *Annals of Mathematical Statistics*, vol. 42, 1622-1636.

[54] Sims, C. (1972), "The Role of Approximation Prior Restrictions in Distributed Lag Estimation", *Journal of the American Statistical Association*, vol. 67, 169-175.

[55] Sims, C., 1980a. Macroeconomics and Reality. Econometrica vol. 48(1), 1-48.

[56] Sims, C., 1980b. Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered. American Economic Review vol. 70(2), 250-57.

[57] Szroeter, J., 1983. Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions. Econometrica vol. 51(2), 335-353.

[58] Tyler, D. E., 1981. Asymptotic Inference for Eigenvectors. Annals of Statistics vol. 9(4), 725-736.

[59] Wahba, G., 1990. Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, Philadelphia: SIAM.

[60] Watson, M. W., 1993. Measures of Fit for Calibrated Models. Journal of Political Economy vol. 101, 1011-1041.

[61] Wen, Y., 2001. A Generalized Method of Impulse Identification. Economic Letters vol. 73, 367-374.

[62] Wen, Y., 2002. The Business Cycle Effects of Christmas. Journal of Monetary Economics vol. 49, 1289-1314.

Appendix: Proofs

Proof of Proposition 3.2

The optimal C-ALS estimator is defined as the solution of the following problem:

$$\widehat{\alpha}_T = \arg\min_{\alpha \in \mathcal{A}} \|K_T^{-1/2} g(\alpha, \widehat{\beta}_T, \omega)\|^2$$
$$\iff \widehat{\alpha}_T = \arg\min_{\alpha \in \mathcal{A}} \left\langle K_T^{-1} g(\alpha, \widehat{\beta}_T, \omega), g(\alpha, \widehat{\beta}_T, \omega) \right\rangle.$$

where $K^{-1/2} = (K^{\dagger})^{-1/2}$. We can rewrite this objective function as:

$$\widehat{\alpha}_T = \arg\min_{\alpha \in \mathcal{A}} \left\langle K_T^{-1} g(\alpha, \widehat{\beta}_T, \omega), K_T K_T^{-1} g(\alpha, \widehat{\beta}_T, \omega) \right\rangle.$$

For sake of notation, $g(\alpha, \widehat{\beta}_T, \omega) \equiv g(\omega)$. Let h_T denote $h_T(\omega) = K_T^{-1} g(\omega)$, the objective function is thus given by:

$$\langle h(\omega), K_T h(\omega) \rangle$$

where

$$K_T h(\omega) = \frac{\partial g}{\partial \beta'}(\omega) \Omega_T^{1/2} \int_{\mathcal{I}} \Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\omega_1)} h(\omega_1)) d\omega_1.$$

This yields

$$\langle h(\omega), K_T h(\omega) \rangle = \int_{\mathcal{I}} h(\omega_1)' \overline{\frac{\partial g}{\partial \beta'}(\omega_1) \Omega_T^{1/2}}) d\omega_1 \int_{\mathcal{I}} \Omega_T^{1/2} \frac{\partial g'}{\partial \beta}(\omega_2) \overline{h(\omega_2)}) d\omega_2.$$

Using the notation,

$$b = \int_{\mathcal{I}} \Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\omega)} h(\omega) d\omega,$$

the objective function is then defined by $b'\overline{b}$.

After multiplying $K_T h(\omega)$ by $\Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\omega)}$ and integrating, one obtains:

$$\int_{\mathcal{I}} \Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\omega)} \frac{\partial g}{\partial \beta'}(\omega) \Omega_T^{1/2} d\omega \int_{\mathcal{I}} \Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\omega_1)} h(\omega_1) d\omega_1$$
$$= \int_{\mathcal{I}} \Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\omega_1)} g(\omega_1) d\omega_1 = s.$$

using $K_T h(\omega) = g(\omega)$. Denoting $W = \int_{\mathcal{I}} \Omega_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\omega)} \frac{\partial g}{\partial \beta'}(\omega) \Omega_T^{1/2} d\omega$, we obtain: $Wb = s$. Now suppose that there exists a generalized inverse of the matrix W denoted $\widetilde{W}$. Then $b = \widetilde{W}s$ and the objective function can be rewritten as $s'\widetilde{W}^2\overline{s}$. This provides the result. When W is of rank equal to q, then $\widetilde{W} = W^{-1}$.

Proof of Proposition 3.3

To derive the limiting distribution of the J-statistic, we need the following Lemma:

Lemma 7.1. *Suppose that $\xi \sim N(0, K)$ is a Gaussian random vector with the covariance operator K defined as the operator from H to H with a finite range of dimension m . Thus,*

$$\xi = \sum_{j=1}^{m} \langle \xi, \phi_j \rangle \phi_j$$

where $\lambda_j, j = 1, \ldots, m$ are the eigenvalues (arranged in decreasing order) of K and ϕ_j the corresponding eigenfunctions of K. The inner product $\langle \xi, \xi \rangle$ is distributed as $\sum_{j=1}^{m} \lambda_j \chi_j^2(1)$ where $\chi_j^2(1)$ are independent central chi-square random variables with 1 degree of freedom.

Proof of Lemma 7.1: Consider the spectral decomposition of the compact linear self-adjoint operator K with a finite range of dimension m such that:

$$K\phi_j = \lambda_j \phi_j$$

where λ_j are the eigenvalues of K for $j = 1, \ldots, m$ and ϕ_j the corresponding eigenfunctions of K such that for $f \in H$

$$Kf = \sum_{j=1}^{m} \lambda_j \langle f, \phi_j \rangle \phi_j$$

with the eigenvalues $\lambda_j, j = 1, \ldots, m$ arranged in decreasing order. For $\xi \sim N(0, K)$, this gives

$$\xi = \sum_{j=1}^{m} \langle \xi, \phi_j \rangle \phi_j = \sum_{j=1}^{m} \sqrt{\lambda_j} \frac{\langle \xi, \phi_j \rangle}{\sqrt{\lambda_j}} \phi_j$$

and $\frac{\langle \xi, \phi_j \rangle}{\sqrt{\lambda_j}}$ are $i.i.d. N(0,1)$. This implies that $\langle \xi, \xi \rangle$ is distributed as $\sum_{j=1}^{m} \lambda_j \chi_j^2(1)$ where $\chi_j^2(1)$ are independent central chi-square random variables with 1 degree of freedom.

Using the mean value expansion, one has:

$$g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) = g(\alpha_0, \beta_0, \omega) + \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T, \bar{\beta}, \omega)(\widehat{\beta}_T - \beta_0) + \frac{\partial g}{\partial \alpha'}(\bar{a}, \widehat{\beta}_T, \omega)(\widehat{\alpha}_T - \alpha_0)$$

where $\bar{\alpha}$ is on the line segment joining $\widehat{\alpha}_T$ and α_0 and $\bar{\beta}$ is on the line segment joining $\widehat{\beta}_T$ and β_0. Taking the asymptotic distribution of $\sqrt{T}(\widehat{\alpha}_T - \alpha_0)$ in Proposition 3.1 and $g(\alpha_0, \beta_0, \omega) = 0$, it follows that:

$$\begin{aligned}
\sqrt{T} g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) &= \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T, \bar{\beta}, \omega)\sqrt{T}(\widehat{\beta}_T - \beta_0) \\
&\quad - \frac{\partial g}{\partial \alpha'}(\bar{a}, \widehat{\beta}_T, \omega) \left\langle S_T \frac{\partial g}{\partial \alpha'}(\widehat{\alpha}_T, \widehat{\beta}_T, \omega), S_T \frac{\partial g}{\partial \alpha'}(\bar{a}, \widehat{\beta}_T, \omega) \right\rangle^{-1} \left\langle S_T^* S_T \frac{\partial g}{\partial \alpha'}(\widehat{\alpha}_T, \widehat{\beta}_T, \omega), \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T, \bar{\beta}, \omega)\sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle.
\end{aligned}$$

where $S_T^* = \overline{S}'$. Since $\widehat{\alpha}_T \xrightarrow{p} \alpha_0$, $\widehat{\beta}_T \xrightarrow{p} \beta_0$ and under the assumption that $\|S_T - S\| \to 0$ in probability:

$$\begin{aligned}
\sqrt{T} g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) &= \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega)\sqrt{T}(\widehat{\beta}_T - \beta_0) \\
&\quad - \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \left\langle S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle^{-1} \left\langle S^* S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega)\sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle \\
&\quad + o_p(1).
\end{aligned}$$

The covariance operator of the right-hand term of the equation above denoted $\tilde{K}$ as a kernel $\tilde{k}(\omega_1, \omega_2)$ given by

$$\tilde{k}(\omega_1, \omega_2) = k(\omega_1, \omega_2) - k_1(\omega_1, \omega_2) - k_1(\omega_1, \omega_2)^* + k_2(\omega_1, \omega_2)$$

where

$$\begin{aligned}
k_1(\omega_1, \omega_2) &= E\left[\frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega_1)\sqrt{T}(\widehat{\beta}_T - \beta_0) \overline{\left\langle S^* S \frac{\partial g}{\partial \alpha'}(\alpha_0, b_0, \omega_1), \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega_2)\sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle}' \overline{A'(\omega_2)} \right] \\
&= E\left[\frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega_1)\sqrt{T}(\widehat{\beta}_T - \beta_0) \int_{\mathcal{I}} \sqrt{T}(\widehat{\beta}_T - \beta_0)' \overline{\frac{\partial g'}{\partial \beta}(\alpha_0, b_0, \omega_2)} S^* S \frac{\partial g}{\partial \alpha'}(\alpha_0, b_0, \omega_1) d\omega \overline{A'(\omega_2)} \right] \\
&= \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega_1)\Omega \int_{\mathcal{I}} \overline{\frac{\partial g'}{\partial \beta}(\alpha_0, b_0, \omega_2)} S^* S \frac{\partial g}{\partial \alpha'}(\alpha_0, b_0, \omega) d\omega_1 \overline{A'(\omega_2)} \\
&= (KS^* S) \frac{\partial g}{\partial \alpha'}(\alpha_0, b_0, \omega_1) \overline{A'(\omega_2)}
\end{aligned}$$

with $A(\omega) = \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \left\langle S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle^{-1}$ and

$$\begin{aligned}
k_2(\omega_1, \omega_2) &= \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega_1) \left\langle S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle^{-1} \left\langle S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), (SKS^*) S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle \\
&\quad \times \left\langle S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle^{-1} \overline{\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega_2)'}.
\end{aligned}$$

Collecting the previous results, under the assumption that $\|S_T - S\| \to 0$ in probability and by Lemma 3.1, one gets:

$$\sqrt{T}S_T g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) \Rightarrow \mathcal{N}\left(0, S\tilde{K}S^*\right)$$

where $\tilde{K} = (I - P(\omega_1))' K \overline{(I - P(\omega_2))'}$ with $P(\omega) = \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \left\langle S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, , \omega), S\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle^{-1} \overline{\frac{\partial g'}{\partial \alpha}(\alpha_0, \beta_0, \omega)SS^*}$. Since the $\mathcal{R}(K)$ is of dimension q and the operator $I - P(\omega)$ from H to H is the projection orthogonal to the subspace spanned by $\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)$ which is of dimension r using Assumption **A.6**, the range of the operator $\tilde{K}$, denoted $\mathcal{R}(\tilde{K})$, is of dimension $q - r$.

Let denote λ_j the eigenvalues ordered in decreasing order of $S\tilde{K}S^*$, by Lemma 7.1, we get

$$\left\langle S_T\sqrt{T}g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega), S_T\sqrt{T}g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega)\right\rangle \xrightarrow{d} \sum_{j=1}^{m} \lambda_j \chi_j^2(1)$$

where m is the dimension of the range of $S\tilde{K}S^*$ with $m \leq q - r$ since the range of $\tilde{K}$ is of dimension $q - r$. When $S_T = I$, λ_i are the eigenvalues of the covariance operator $\tilde{K}$ with a range of dimension $q - r$. For the optimal C-ALS obtained with a consistent estimator $K_T^{-1/2}$ of $K^{-1/2} = (K^\dagger)^{-1/2}$ evaluated at a first-step consistent estimator of α, the asymptotic distribution of the statistic is a chi-square distribution with $q - r$ degrees of freedom since

$$\sqrt{T}K_T^{-1/2} g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) \Rightarrow \mathcal{N}(0, M)$$

where $M = I - K^{-1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega_1)\left\langle K^{-1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, , \omega), K^{-1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle^{-1} \overline{\frac{\partial g'}{\partial \alpha}(\alpha_0, \beta_0, \omega_2)K^{-1/2}}$ is a bounded idempotent operator with $q - r$ eigenvalues equal to one and r eigenvalues equal to zero.

Proof of Proposition 3.4 Suppose there exists a lower bound ω_{lb} such that for this lower bound $J_T(\omega_{lb}) = O_p(1)$. The restrictions are then asymptotically valid for the interval $(-\omega_{lb}, \omega_{lb})$. Now, there exists two possible cases for which $|\omega| \neq |\omega_0|$. First, consider the case where $|\omega| > |\omega_0|$. For this case, $J_T(\omega) \to \infty$ while $J_T(\omega_0) = O_p(1)$. Thus $VISC_T(\omega_0) - VISC_T(\omega) \xrightarrow{p} -\infty$. The criterion selects the interval $(-\omega_0, \omega_0)$ with a probability going to one when T is going to ∞. For the second case, $|\omega| < |\omega_0|$ which implies that both $J_T(\omega)$ and $J_T(\omega_0)$ are $O_p(1)$. Since $|\omega| < |\omega_0|$, $-h(|\omega_0|)\kappa_T + h(|\omega|)\kappa_T \to -\infty$ which implies $VISC_T(\omega_0) - VISC_T(\omega) \xrightarrow{p} -\infty$. By combining the two results, the criterion selects ω_0 with a probability going to one when T diverges toward ∞ for all $\omega \neq \omega_0$.

Proof of Proposition 4.1

Consider the vector of just- or over-identified estimating equations defined by Eq. 3.11. Let $\beta = (\text{vec}(\Phi_p)', \text{vec}(\Sigma)')' \equiv (\Phi', \sigma')'$ denote the vector of reduced-form parameters, and $\Omega_T = \begin{pmatrix} \Omega_\Phi & \mathbf{0} \\ \mathbf{0} & \Omega_\sigma \end{pmatrix}$ the corresponding partitioning of the asymptotic variance-covariance matrix of the OLS estimator of β. The first-order conditions of the first-step objective function with respect to a are given by:

$$\int_{\underline{\omega}}^{\overline{\omega}} \frac{\partial}{\partial a'}\left\{\text{vec}\left(C(e^{-i\omega})A(0)\right)' \mathrm{H}'\mathrm{H}\overline{\text{vec}\left(C(e^{-i\omega})A(0)\right)}\right\} d\omega + \frac{\partial}{\partial a'}\left\{\text{vech}\left(\Sigma - A(0)A(0)'\right)' \text{vech}\left(\Sigma - A(0)A(0)'\right)\right\} = \mathbf{0},$$

where

$$\begin{aligned}
\frac{\partial}{\partial a'}\left\{\text{vech}\left(\Sigma - A(0)A(0)'\right)' \text{vech}\left(\Sigma - A(0)A(0)'\right)\right\} &= -2\text{vech}\left(\Sigma - A(0)A(0)'\right)' D_N^+ \left[K_{NN}\left(A(0) \otimes I_N\right) + \left(A(0) \otimes I_N\right)\right] \\
&= -2\text{vech}\left(\Sigma - A(0)A(0)'\right)' D_N^+ \left[\left(A(0) \otimes I_N\right) + \left(A(0) \otimes I_N\right)\right] \\
&= -4\text{vech}\left(\Sigma - A(0)A(0)'\right)' D_N^+ \left(A(0) \otimes I_N\right)
\end{aligned}$$

using the results in Lütkepohl, (2007, p. 363) and $D_N^+ K_{NN} = D_N^+$ with K_{NN} the commutator matrix for which $K_{NN}\text{vec}(X) = \text{vec}(X')$ for any $N \times N$ matrix X, and

$$\begin{aligned}
\frac{\partial}{\partial a'}\left\{\text{vec}\left(C(e^{-i\omega})A(0)\right)' \mathrm{H}'\mathrm{H}\overline{\text{vec}\left(C(e^{-i\omega})A(0)\right)}\right\} &= \frac{\partial}{\partial a'}\left[\text{vec}(A(0))' \left(I_N \otimes C(e^{-i\omega})\right)' \mathrm{H}'\mathrm{H}\left(I_N \otimes \overline{C(e^{-i\omega})}\right) \text{vec}(A(0))\right] \\
&= 2\text{vec}(A(0))' \left(I_N \otimes C(e^{-i\omega})\right)' \mathrm{H}'\mathrm{H}\left(I_N \otimes \overline{C(e^{-i\omega})}\right) \frac{\partial \text{vec}(A(0))}{\partial a'} \\
&= 2\text{vec}(A(0))' \left(I_N \otimes C(e^{-i\omega})\right)' \mathrm{H}'\mathrm{H}\left(I_N \otimes \overline{C(e^{-i\omega})}\right)
\end{aligned}$$

Finally the first result follows by taking the transpose of the two previous expressions. For the second-step estimator, using

$$\frac{\partial \text{vec}(\Sigma)}{\partial \Phi'} = \mathbf{0} \qquad \text{and} \qquad \frac{\partial \text{vec}\left(\mathbf{A}(\mathbf{0})\right)}{\partial \sigma'} = \mathbf{0},$$

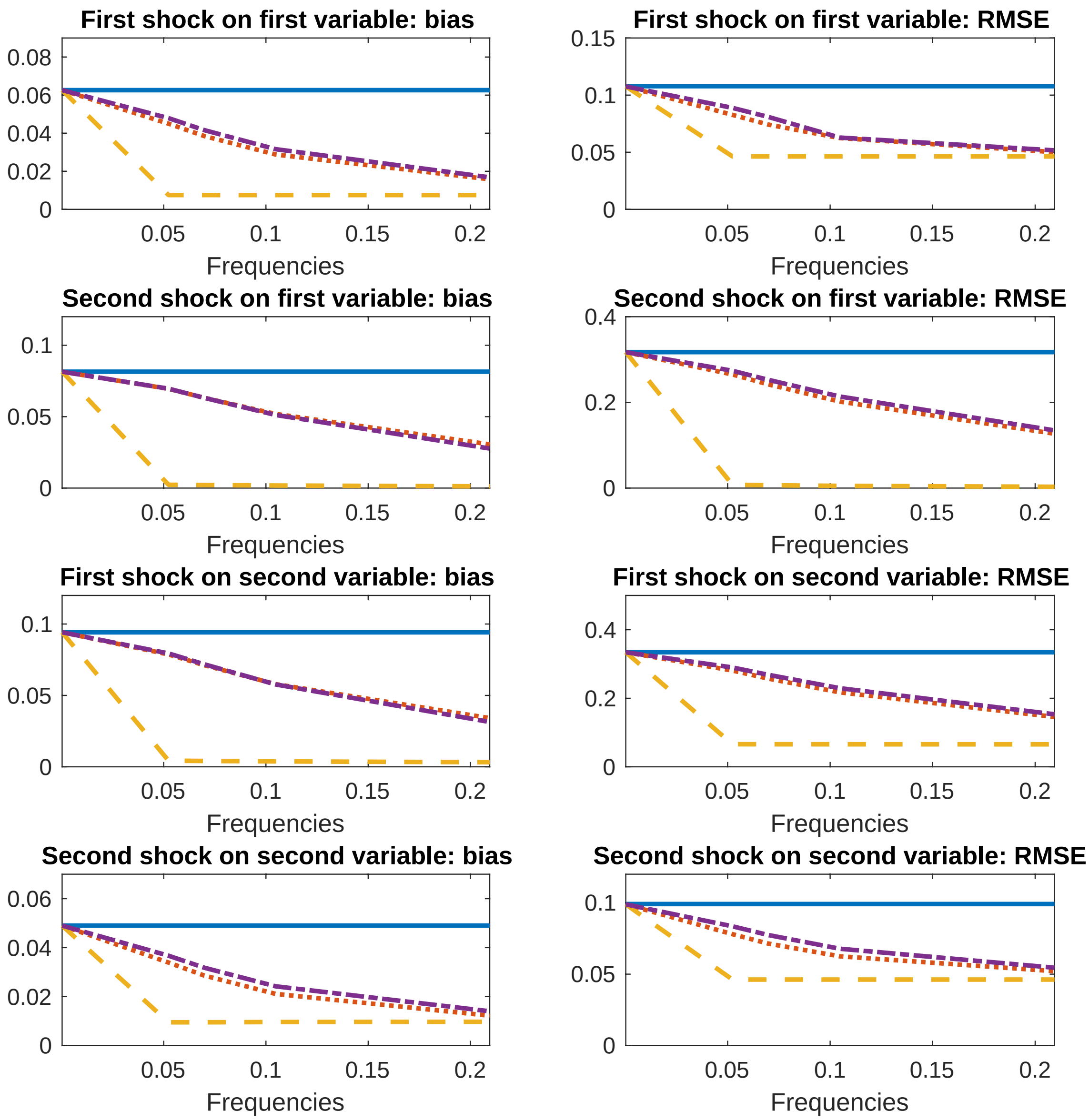
imply that the weighting matrix is block diagonal. The optimal weighting matrix for the first set of estimating equations is given by Proposition 3.2 and $\frac{\partial \text{vec}(C(e^{-i\omega}))A(0))}{\partial \Phi'} = (A(0)' \otimes I)\frac{\partial \text{vec}(C(e^{-i\omega}))}{\partial \Phi'}$ where $\frac{\partial \text{vec}(C(e^{-i\omega}))}{\partial \Phi'}$ can be easily derived from Lütkepohl (2007, p. 111). The optimal weighting matrix for the second set of estimating equations is given by $2D_N^+(\widehat{\Sigma} \otimes \widehat{\Sigma})D_N^{+\prime}$ (see Lütkepohl, 2007, p. 93).

Table 1: J-test

	Quarters											
n	30	60	90	120	30	60	90	120	30	60	90	120
	a. VAR(1) : $\rho = .90$											
	$\delta = 0$				$\delta = .05$				$\delta = .1$			
% 2nd struct. shock	0	0	0	0	.1139	.1456	.1575	.1628	.3218	.3730	.3903	.3979
.05	.0274	.0318	.0332	.0324	.4039	.4165	.4374	.4546	.9034	.9139	.9218	.9226
.10	.0457	.0494	.0547	.0508	.4705	.4770	.4977	.5137	.9273	.9349	.9403	.9424
	b. VAR(1) : $\rho = .95$											
	$\delta = 0$				$\delta = .05$				$\delta = .1$			
% 2nd struct. shock	0	0	0	0	.2421	.3176	.3521	.3708	.5513	.6128	.6356	.6472
.05	.0259	.0306	.0358	.0393	.5875	.6336	.6506	.6628	.9779	.9818	.9831	.9837
.10	.0448	.0497	.0553	.0581	.6464	.6826	.6993	.7070	.9847	.9864	.9874	.9878
	c. VAR(1) : $\rho = .98$											
	$\delta = 0$				$\delta = .05$				$\delta = .1$			
% 2nd struct. shock	0	0	0	0	.5358	.6229	.6586	.6778	.8245	.8357	.8389	.8401
.05	.0242	.0278	.0341	.0369	.7880	.8115	.8208	.8306	.9924	.9942	.9946	.9956
.10	.0442	.0457	.0530	.0563	.8236	.8354	.8489	.8590	.9942	.9960	.9962	.9965

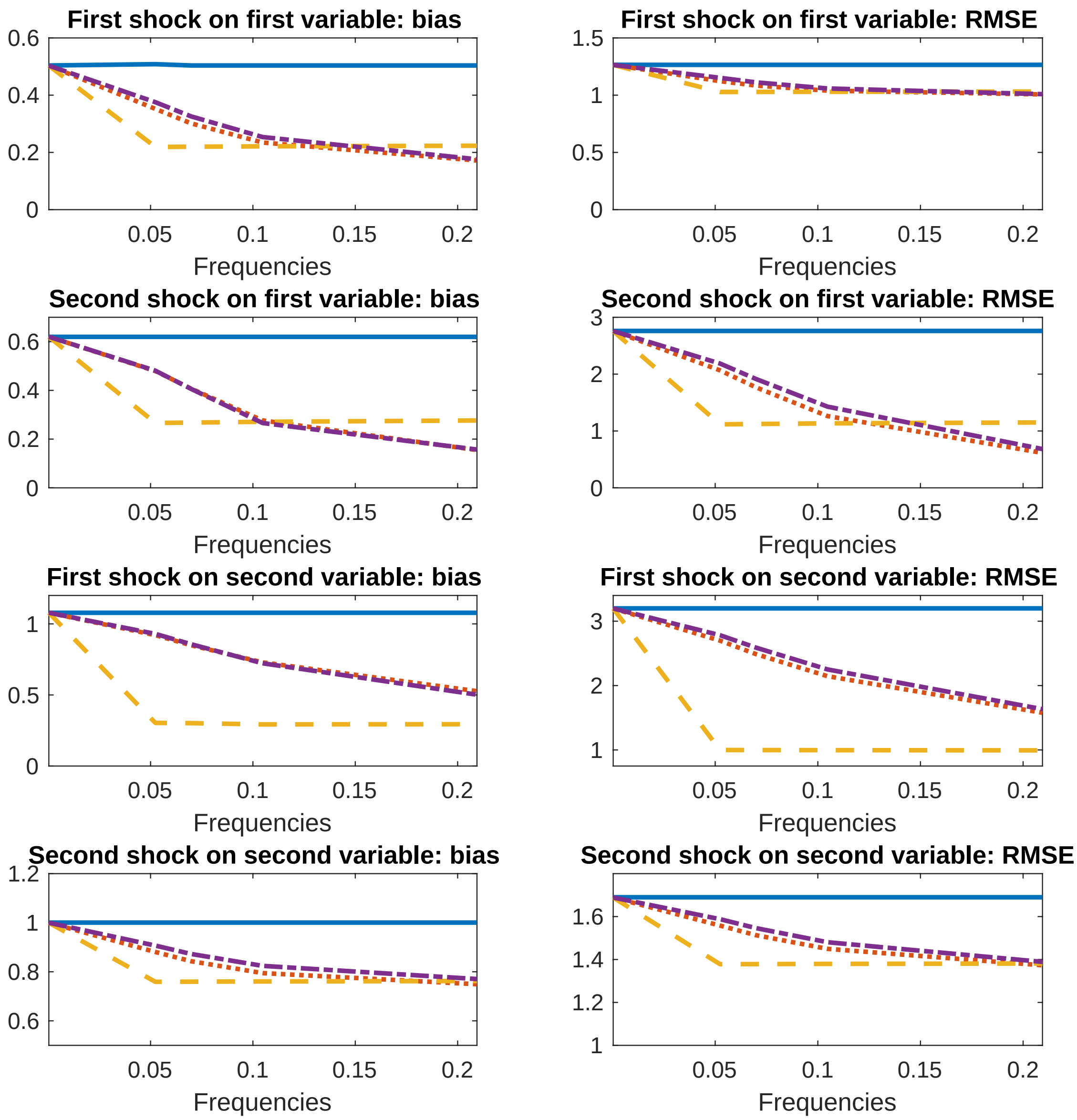
Note: The frequency intervals under investigation are: $\omega_n = (-\frac{2\pi}{n}, \frac{2\pi}{n})$ for $n = 30, 60, 90, 120$ quarters. The percentage of the second structural shock represents the proportion of the variance explained by the second shock for the first variable in the frequency interval of interest.

Figure 1: Contemporaneous bias and RMSE using a VAR(1) model with $\rho = .95$ and $\delta = 0$



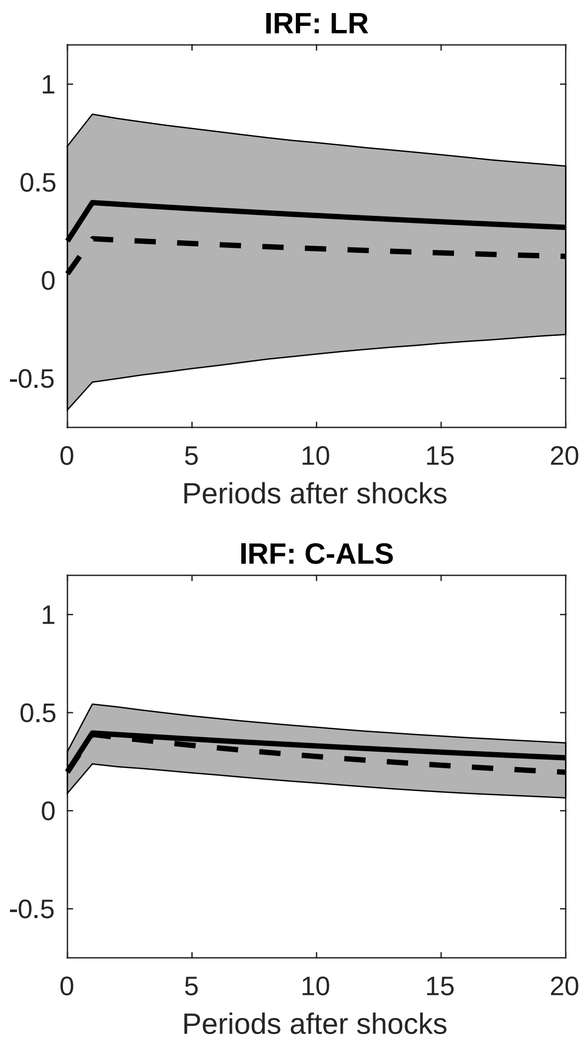
Note: The left panel displays the (average) bias of the contemporaneous effect of shock i on variable j $(i, j = 1, 2)$, i.e. the average impulse response function minus the true impulse response function for $h = 0$, using subsequently the frequency intervals $\omega_n = \left[-\frac{2\pi}{n}, \frac{2\pi}{n}\right]$ for $n = 30, 60, 90$, and 120 quarters. The right panel displays the corresponding RMSE. The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively. Results are obtained from 10,000 samples of size $T = 200$ using equations (5.20) and (5.21).

Figure 2: Cumulative Bias and RMSE up to 12 quarters using a VAR(1) model with $\rho = .95$ and $\delta = 0$



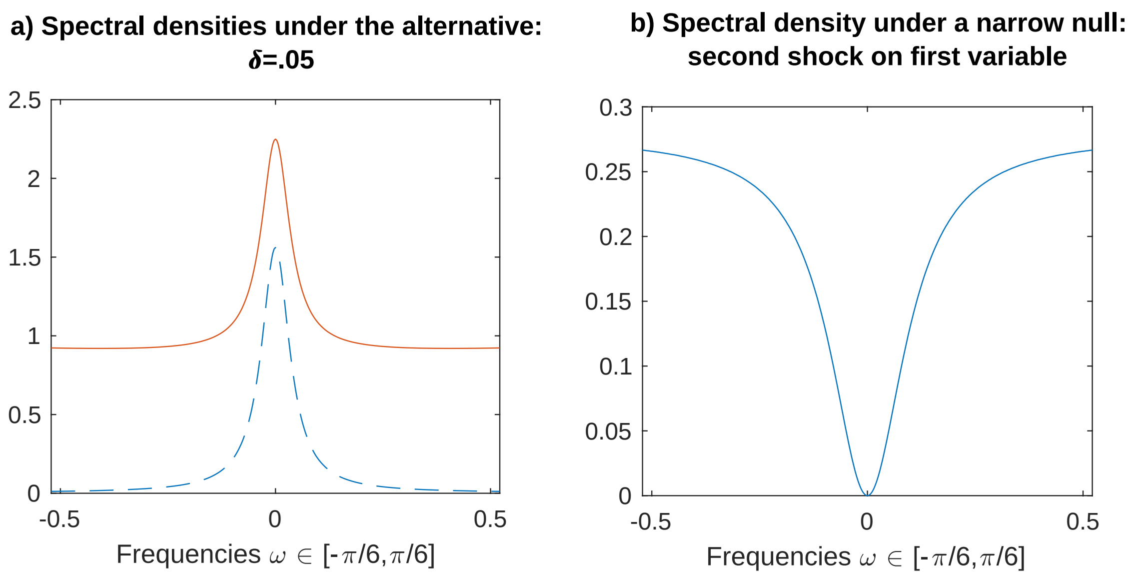
Note: The left panel displays the cumulative mean absolute bias of the effect of shock i on variable j $(i, j = 1, 2)$ up to $H = 12$ quarters using subsequently the frequency intervals $\omega_n = \left[-\frac{2\pi}{n}, \frac{2\pi}{n}\right]$ for $n = 30, 60, 90,$ and 120 quarters. The right panel displays the corresponding RMSE. The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively. Results are obtained from 10,000 samples of size $T = 200$ using equations (5.20) and (5.21).

Figure 3: Impulse Responses for the first shock on second variable when $n = 60$ quarters, $\rho = .98$ and $\delta = 0$



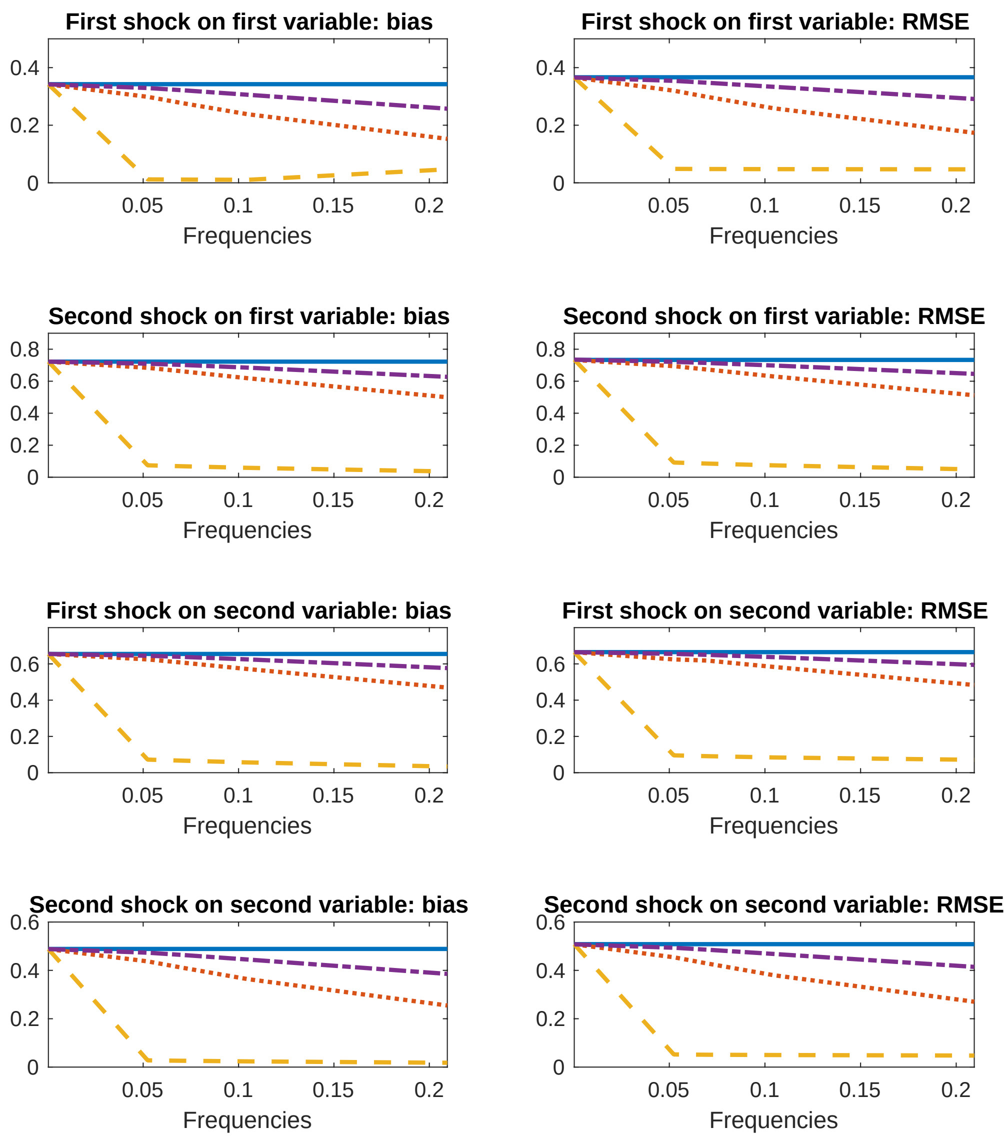
Note: The solid line and dashed line represent the true impulse response function for the first shock on the second variable and the average impulse response function using the BQ, first-step C-ALS, second-step C-ALS or the MS estimator when the long-run restriction is imposed on the frequency interval $\omega_{60} = \left[-\frac{2\pi}{60}, \frac{2\pi}{60}\right]$. Confidence intervals are based on the 95–percentile from 10,000 Monte–Carlo experiments.

Figure 4: Marginal spectral densities of the first variable under the null and the alternative hypothesis



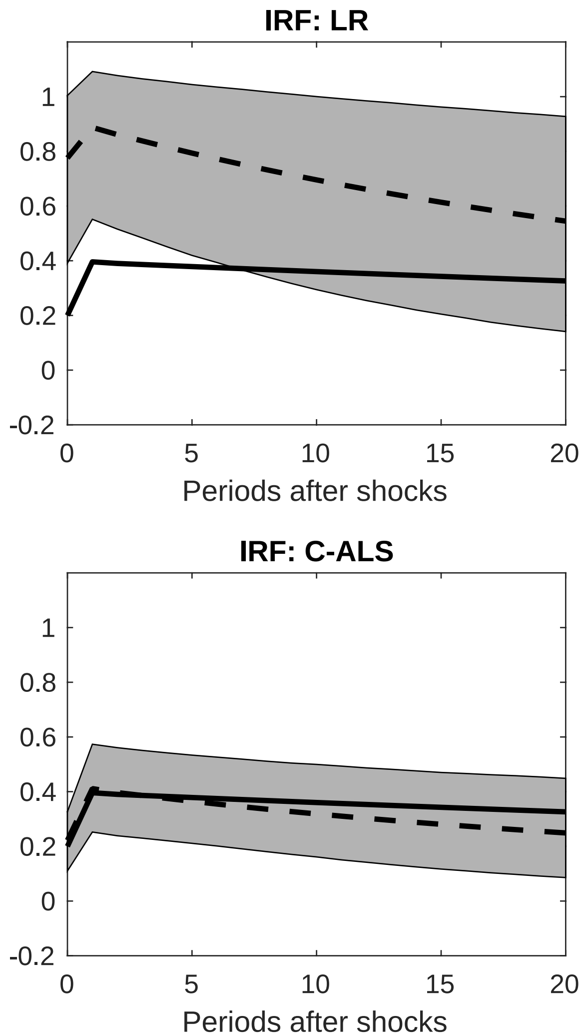
Note: In Figure 4 a), the solid line is the contribution of the first shock on the marginal spectral density of the first variable and the dashed line, the second shock on the first variable.

Figure 5: Contemporaneous bias and RMSE using a VAR(1) model with $\rho = .95$ and $\delta = .1$



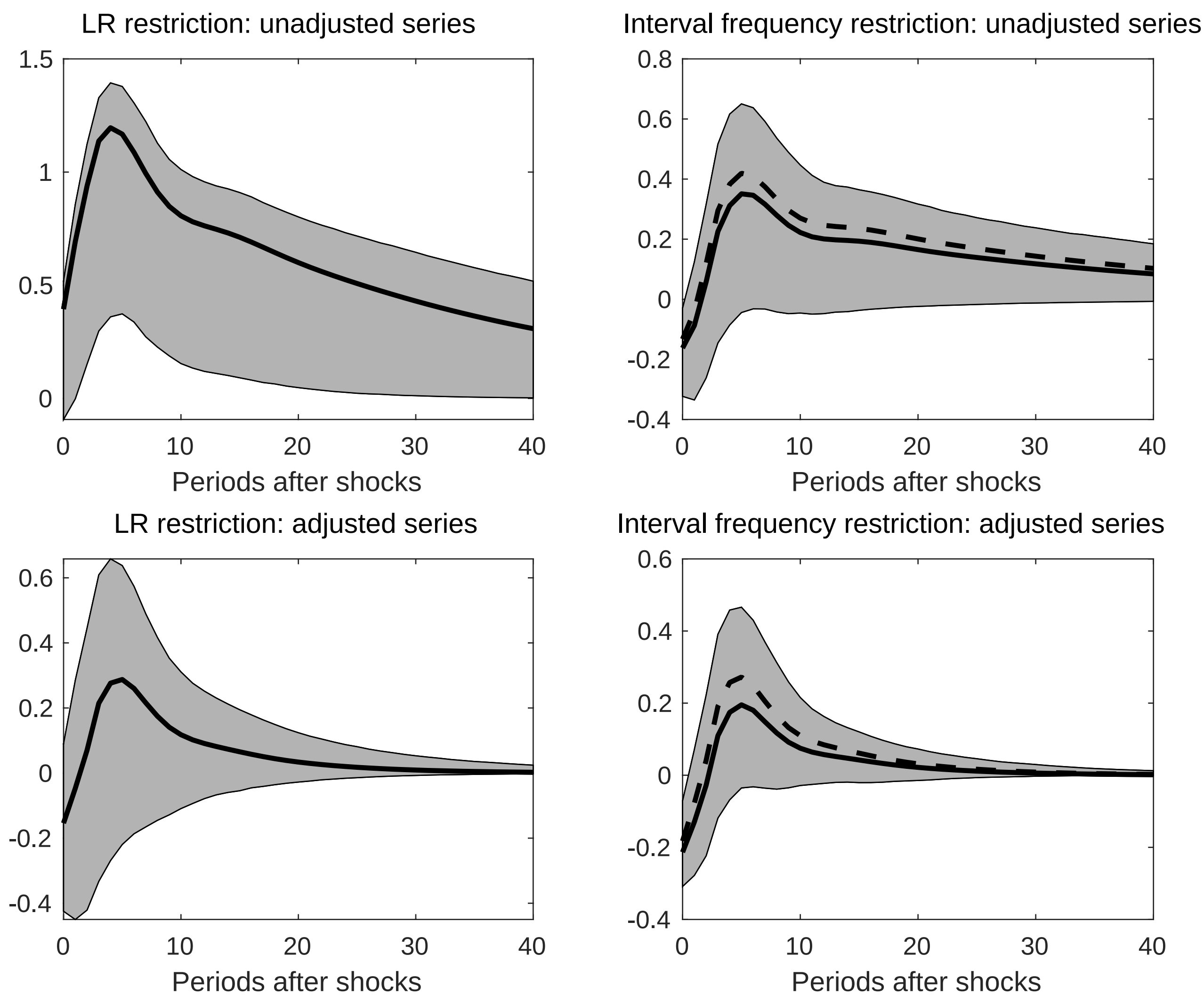
Note: Monte Carlo simulations assume that there is a misspecified exclusion restriction ($\delta = .1$). A zero long-run restriction of the second shock on the first variable is imposed whereas both shocks have a permanent effect on the first variable. The left panel displays the (average) bias of the contemporaneous effect of shock i on variable j ($i, j = 1, 2$), i.e. the average impulse response function minus the true impulse response function for $h = 0$, using subsequently the frequency intervals $\omega_n = \left[-\frac{2\pi}{n}, \frac{2\pi}{n}\right]$ for $n = 30, 60, 90,$ and 120 quarters. The right panel displays the corresponding RMSE. The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

Figure 6: Impulse Responses for the first shock on second variable when $n = 60$, $\rho = .98$ and $\delta = .05$



Note: The solid line and dashed line represent the true impulse response function for the first shock on the second variable and the average impulse response function using the BQ, first-step C-ALS, second-step C-ALS or the MS estimator when there is a misspecified exclusion restriction ($\delta = 0.05$). Confidence intervals are based on the 95–percentile from 10,000 Monte–Carlo experiments.

Figure 7: Impulse responses for the technology shock on hours worked



Note: Confidence intervals are based on the 95–percentile from 2000 bootstraps. In the right panels, the solid line and the dash line represent the C-ALS and the discretized ALS estimators respectively. The identifying interval frequency restriction is imposed over the frequency interval $\omega_{120} = [-\frac{2\pi}{120}, \frac{2\pi}{120}]$.