ESG RECHERCHE

DOCUMENT DE TRAVAIL / WORKING PAPER

No. 2021-06

SVARs in the Frequency Domain using a Continuum of Restrictions

Alain Guay et Florian Pelgrin

Août 2021

ESG UQÀM

Département des sciences économiques
École des sciences de la gestion
Université du Québec à Montréal

SVARs in the Frequency Domain using a Continuum of Restrictions

Alain Guay, Université du Québec à Montréal,
Chaire en macroéconomie et prévisions ESG-UQAM, et CIREQ

Florian Pelgrin, EDHEC Business School

Document de travail No. 2021-06

Août 2021

Département des Sciences Économiques
Université du Québec à Montréal
Case postale 8888,
Succ. Centre-Ville
Montréal, (Québec), H3C 3P8, Canada
Courriel : brisson.lorraine@uqam.ca
Site web : http://economie.esg.uqam.ca

SVARs in the Frequency Domain using a Continuum of Restrictions*

Alain Guay†

Florian Pelgrin‡

This version: June 15, 2021

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 (Gourieroux, Monfort and Trognon, 1985) in which there is a continuum of nonlinear estimating equations. Following Carrasco and Florens (2000), we then develop a continuum asymptotic least squares estimator (C-ALS) that exploits efficiently 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 with a data-driven procedure. Finally, we provide some new results using Monte Carlo simulations and applications regarding the hours-productivity debate and the impact of news shocks.

JEL classification: C12, C32, C51.

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

*The first author gratefully acknowledges financial support of SSHRC and FRQSC. We would also like to thank Jean-Pierre Florens, René Garcia, John Galbraith, Stéphane Gregoir, Christophe Hurlin, Patrick Fève, Nour Meddahi, Benoît Perron, Simon Van Norden, Roberto Pancrazi, Eric Renault and Eva Ortega for comments on an earlier draft of the paper.

†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 effervescent 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).[1]

In this paper, we propose a joint methodology for the identification and inference of structural vector autoregressive models in the frequency domain. Our starting point is that identifying restrictions in the frequency domain condition the impulse response functions and correspond to constraints on (multivariate) spectral density functions. Accordingly, there is a general mapping between identifying restrictions and the power spectra of the variables of interest. Said differently, imposing identifying restrictions on a frequency band is equivalent to constraint the contribution of structural shocks to explain the variance of given variables at those frequencies. To illustrate it, consider the model of Galí (1999) that attributes variation in U.S. labour productivity and hours worked to a technology shock and a non-technology shock. Identification is achieved by imposing that the latter has no long-run effects, i.e. at the frequency $\omega = 0$, on labour productivity. As suggested by Faut and Lepper (1997), Faust (1998) and Pötscher (2002), we can impose restrictions on the effects of the non-technology shock not only at the zero frequency but also its neighborhood, say $\omega \in (-\underline{\omega}, \overline{\omega})$. Intuitively, it means that the contribution of the non-technology shock is expected to be negligible to explain the variance of the (level) real GDP in the medium- to long-term. More generally, our approach allows for identifying restrictions on the frequency bands of interest in finance and macroeconomics, which correspond to short-run, business cycles, medium-term and long-run fluctuations. In this respect, this generalization permits a much richer set of identifying restrictions and possibly to better capture a wide range of structural shocks that are only weakly identified through existing methods.

Our main contribution is to define and solve the identification and inference of structural vector autoregressive models using identifying restrictions in the frequency domain as an appropriate asymptotic least squares problem. Indeed, ignoring for the moment the writing of identifying restrictions in the frequency domain, it is well-known that imposing short-run and/or long-run restrictions leads to a finite set of nonlinear equations.[2] A somewhat different perspective is to see this problem from the angle of the Asymptotic Least Squares (henceforth, ALS) theory developed by Gourieroux, Monfort and Trognon (1985). Indeed, the identification of structural shocks/parameters is obviously concomitant to the estimation of the autoregressive parameters of the reduced-form VAR as well as the variance-covariance matrix of the innovations. In this respect, the set of nonlinear equations that defines the mapping between the structural representation and the reduced-form can be interpreted as nonlinear estimating functions in which the structural parameters of

[1] In a different perspective, Gourieroux, Monfort and Renne (2020) propose a statistical identification of structural shocks in a non-Gaussian framework. See also Guay (2021).

[2] See Kilian and Lütkepohl (2017), Chapters 11 and 12.

interest depend on auxiliary parameters, those of the reduced form, which are estimated in a first step. In doing so, we proceed as in the ALS approach and this framework therefore makes it possible to naturally study the identification and inference of structural VAR models. Returning now in the frequency domain, the counterpart of these (finite) nonlinear estimating functions is a continuum of nonlinear functional equations defined on a frequency band $\omega \in (-\underline{\omega}, \overline{\omega})$.

To do this, after reinterpreting the structural identification in an ALS perspective, we take advantage of the methodology proposed by Carrasco and Florens (2000), and Carrasco et al. (2007), namely the Continuum-GMM estimator (henceforth, C-GMM). Notably, Carrasco and Florens (2000) show that there exists a C-GMM estimator—in the presence of a continuum of (generalized) moment conditions—that closely mimics the efficient two-step GMM of Hansen (1982). The main idea is to minimize a quadratic form, with metrics a regularized inverse of the asymptotic covariance operator, in a Hilbertian functional space than spans the continuum of estimating equations. Then the first step based on the identity operator metrics leads to a consistent estimator and the second step delivers an efficient estimator. In our context, this amounts of considering the Hilbertian functional space that spans the continuum of identifying restrictions or their dual representation with the constrained power spectra, i.e. to project the continuum of moment restrictions onto the space engendered by the auxiliary parameters of the reduced-form VAR. Intuitively, it means that one exploits efficiently the information embedded in the estimation of the reduced-form VAR. More fundamentally, the cornerstone of our approach is that the second step involves a set of overidentified estimating equations, thanks to the projection defined on the Hilbertian functional space. This opens the window to test such restrictions and to select the frequency band.

As a by-product, the fact that the dimension of the vector of auxiliary parameters is finite, say q, has at least three fundamental implications. First, while the Hilbert-Schmidt covariance operator is not invertible on the full reference space as in Carrasco and Florens (2000), it has a finite dimensional closed range (at most) equal to q. Second, the objective function only involves a one-dimensional integral against a well-chosen measure and thus does not require any quadrature method or Monte Carlo integration.[3] Third, the practical implementation of the (regularized) objective function and the optimal weighting matrix does not depend on the inverse of a matrix of dimension T, but rather on a matrix of dimension q. These three properties render our approach quite appealing and attractive from a computational point of view. To summarize, we propose a new efficient estimator (C-ALS estimator) that combines the seminal work of Carrasco and Florens (2000) with the asymptotic least squares method (Gourieroux et al., 1985; Gourieroux and Monfort, 1995) in the presence of a continuum of identifying restrictions. Importantly, our approach is optimal compared to an alternative strategy based on a selection of a sufficiently refined grid through a discretization of the frequency band and thus to a discretized ALS estimator, and does not involve any numerical approximation.[4]

[3] In general, the objective function of the C-GMM involves a d-dimensional integral against a well-chosen measure, where d is the dimension of the multivariate (random) process.

[4] For instance, we discuss thoroughly the case of a bivariate SVAR models and show that the discretized ALS estimator makes use of a different weighting procedure, which results in a loss of efficiency.

Our new methodology for the identification and inference of structural VAR models has also other appealing features. First, 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 mean squared errors irrespective of the IRF horizon. This results from two key elements. On the one hand, using a frequency band allows more substantial information for the identification of structural shocks and thus the structural impulse responses functions. On the other hand, the second-step exploits efficiently this information. Intuitively, and as reported in our Monte Carlo simulations and applications, the IRF confidence bands are sharply reduced with respect to the first-step C-ALS estimator and other alternatives. Second, when imposing restrictions on a neighborhood of $\omega = 0$, we offer a credible alternative to the so-called unreliability problem (Faust, 1996; Faust and Leeper, 1997) and especially the fact that one can neither form asymptotically correct confidence intervals for the impulse response functions nor can offer any consistent test of the nullity of the latter from any inconsistent estimate of the long-run multiplier.

Third, it 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. Following Carrasco and Florens (2000), we propose tests of overidentification with a modified J-stat using any positive definite weighting matrix including the optimal one.[5] The modified J-stat allows for selecting the frequency band (i.e., the reliability of the imposed restrictions), thereby permitting us to conduct a data-driven procedure in order to assess frequency intervals on which the imposed restrictions might be satisfied. We illustrate the usefulness of the overidentification test and the interval selection in two applications regarding the identification of a technology shock and a news shock.[6]

Our paper is closely related to two strands of the literature in the frequency domain, namely the full identification of structural shocks, as well as the (agnostic) partial identification case using the max-share approach. On the one hand, Wen (2001, 2002) proposes a min-effect/max-effect frequency estimator in a bivariate VAR model. We show that it corresponds to the first-step C-ALS estimation in the bivariate case. Similarly, Chahrour and Jurado (2018, 2021) propose an identification condition of structural shocks in the context of expectations-driven fluctuations using spectral density on the whole frequency band $\omega \in [-\pi, \pi]$. Especially they relax the assumption of a fundamental representation with a less restrictive condition, called recoverability—structural shocks can be backed out from all past, present and future observables available to the econometrician. It turns out that an optimal C-ALS two-step estimator can also be performed in this context, thereby permitting to test the imposed identification restrictions to recover the noise and the fundamental structural shocks.

On the other hand, our paper is also related to the strand of the literature on the max-share frequency

[5]Overidentification tests have been proposed when the number of restrictions is greater than required in the identification procedure (see Bernanke and Mihov, 1998).

[6]Section 3 illustrates several possible applications of these two testing procedures.

approach. Starting from the seminal contributions of Faust (1998) and Uhlig (2003, 2004) in the time domain, DiCecio and Owyang (2010) propose a frequency-based max-share identification in the case of bivariate SVAR models, which amounts of finding a structural shock (e.g., the technology shock) that maximizes the share of the forecast-error variance in the two variables of interest (e.g., productivity growth and hours worked) on various frequency bands (see also Dieppe et al., 2019). Using the max-share frequency approach with a larger VAR dimension, Angeletos et al. (2020) and Basu et al. (2021) provide evidence that a single shock mostly explains the variance of a wide range of financial and macroeconomic variables over selected band of frequencies. Such max-share approach in the frequency domain can also be seen as a particular case of our proposed approach in the partial identification situation.

The rest of the paper is organized as follows. Section 2 reviews notation and present our identification strategy at a given frequency and then a frequency band. Section 3 discusses some applications of the C-ALS methodology. Section 4 presents the (optimal) C-ALS estimator using the methodology proposed by Carrasco and Florens (2000). Section 5 first provides the derivations for any multivariate VAR model in the case of full identification and then applies these results to a bivariate structural VAR. We then turn to the partial identification case, especially for a single structural shock. Section 6 proposes a comparative study of competing identification schemes using some Monte-Carlo simulations. Section 7 revisits the empirical evidence of two main applications of structural VAR models. The last section contains concluding comments and future extensions. Proofs are gathered in Appendix.

2 Frequency identification of structural VAR models

In this section, we first introduce preliminary notation and provide an overview of our frequency-based approach.

2.1 Notation

It is assumed that a 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 difference or stationary linear combination of integrated variables.[7] Accordingly, $X_t = (X_{1t}, \cdots, X_{Nt})'$ is a $(N \times 1)$ random vector and $X = (X_1, \cdots, X_T)$ is a $(N \times T)$ random matrix. To simplify the notation, presample values for each variable are assumed to be available. Furthermore all deterministic regressors have been suppressed for notational convenience. X_t is approximated by a stationary, stable, reduced-form $\texttt{VAR}(p)$ process:

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

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, $u_t = (u_{1t}, \cdots, u_{Nt})'$ is a N-dimensional white noise (innovation process), that

[7]The identification strategy proposed here could be extended to VECM.

is, $E(u_t) = 0_{N\times 1}$, $E(u_t u_t') = \Sigma$ is nonsingular and $E(u_t u_s') = 0_{N\times N}$ for $s \neq t$.

The corresponding reduced-form vector moving average representation is defined by:

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

where $C(L) = \sum_{i=0}^{\infty} C_i L^i$, L is the lag operator, and $C_0 = C(0) = I_N$ (with I_N the identity matrix of order N), $C_i = \sum_{j=1}^{i} C_{i-j}\phi_j$.

The structural VAR model can be written as:

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

where $A(L) = \sum_{i=0}^{\infty} A_i L^i$, $A_0 \equiv A(0)$, and ϵ_t is a random $N \times 1$ vector of structural shocks with $E(\epsilon_t) = 0$ and $E(\epsilon_t \epsilon_t') = \Gamma$. A common identification assumption is $\Gamma = I_N$.

Taking equations (2.2) and (2.3), the error terms of the reduced-form model are related to the structural shocks as follows:[8]

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

with $C(L)A(0) = A(L)$, and thus

$$C(0)\Sigma C(0)' \equiv \Sigma = A(0)A(0)'. \tag{2.5}$$

Then the central question is how to recover the elements of $A(0)$ from consistent estimates of the reduced-form parameters. In the sequel, we propose a frequency-based approach of structural identification.

2.2 Imposing restrictions on a frequency band

We first present a unifying framework embedding the short-run and long-run restrictions, and those in the frequency domain. Let $z \in \mathbb{C}$, the mapping of the reduced-form innovations and the structural shocks using equation (2.5) can be written generically as follows:

$$C(z)\Sigma C(z)^* = A(z)A(z)^* \tag{2.6}$$

where $C(z) = \left[I_N - \phi_1 z - \cdots - \phi_p z^p\right]^{-1}$, $C(z)^* = \overline{C(z)}'$ and $A(z)^* = \overline{A(z)}'$ denote the transpose complex conjugate of C and A, respectively. For $z = 0$, this corresponds to equation (2.5) and for $z = 1$, to the equality between the long-run variance-covariance matrices of the reduced-form innovations and the structural shocks. When $z = e^{-i\omega} \in \mathbb{C}$ for any $\omega \in [-\pi, \pi]$, the left-hand side term (respectively, right-hand side

[8] For a more general presentation, see Amisano and Giannini (1997), Lüktepohl (2007), Kilian (2013), and Kilian and Lüktepohl (2017).

term) is the Fourier transform of the reduced-form VAR (respectively, the structural VAR) for any frequency in $[0, \pi]$. Said differently, up to the constant term $1/2\pi$, it is simply the equality between the multivariate spectral density matrix of the reduced-form VAR and the one of the structural VAR for any frequency ω.

Equation (2.6) implies the following mapping between the reduced-form autoregressive matrix and the structural autoregressive matrix:

$$C(z)A(0) = A(z). \tag{2.7}$$

Starting from equation (2.6), the key observation is that Σ and $C(z)$ can be estimated from data, using some estimates of the autoregressive parameters $\beta = (\text{vec}(\Phi_p)', \text{vech}(\Sigma)')'$, denoted $\widehat{\beta}_T$. By equation (2.7), if one puts enough restrictions on $A(z)$, the remaining elements of $A(z)$ and thus $A(0)$ can be pinned down. Especially, identifying restrictions are generally imposed on some (i, j) elements of $A(z)$:

$$a_{ij}(z) = 0 \quad \text{for a given z}$$

or

$$[C(z)]_{i.}\, [A(0)]_{.j} = 0$$

where $[B]_{i.}$ and $[B]_{.j}$ denote the i-th row and j-th column of the B matrix, respectively. Consider the bivariate case with the restriction: $a_{12}(z) = 0$. This gives the following equation:

$$\widehat{c}_{11}(z)a_{12}(0) + \widehat{c}_{12}(z)a_{22}(0) = 0$$

where $\widehat{c}_{ij}(z)$ and $a_{ij}(0)$ are respectively the (i, j) element of the matrices $\widehat{C}(z)$ and $A(0)$. When $z = 0$, one obviously recovers the usual short-run restriction $a_{12}(0) = 0$ corresponding to the Cholesky decomposition since $\widehat{c}_{11}(0) = 1$ and $\widehat{c}_{12}(0) = 0$ using $C(0) = I_2$. Following Galí (1999), using a long-run restriction ($z = 1$), for a bivariate VAR model including labor productivity and hours worked, the technology shock can be identified by assuming that the second structural shock has no long-run impact on the first variable:

$$\widehat{c}_{11}(1)a_{12}(0) + \widehat{c}_{12}(1)a_{22}(0) = a_{12}(1) \equiv 0.$$

In general, when $z = 1$ (or $\omega = 0$), one recovers the standard long-run identification scheme of Blanchard and Quah (1989), $C(1)\Sigma C(1)' = A(1)A(1)'$. Only $N(N-1)/2$ restrictions on $A(1)$ are then required to satisfy the order condition in the case of exact identification and the matrix $A(0)$ is recovered from the relationship: $C(1)A(0) = A(1)$.

Finally, using equations (2.6) and (2.7) and the economic interpretation of the spectral density matrix, imposing structural identifying restrictions on a frequency band is equivalent to minimize the contribution of certain structural shocks for the variables of interest. For instance, in a bivariate structural VAR model, the identifying restriction that the second structural shock has no permanent effect on the first variable at a given frequency, say in the long-run, translates into the fact that the partial spectrum of the first variable with respect to the second structural shock is minimized at this frequency. Without loss of generality, we discuss the estimating equations using equation (2.7) and we differ the discussion of the spectral-based representation in Section 5.

For any given z, the r-identifying restrictions write as follows:

$$H\text{vec}\left(A(z) - C(z)A(0)\right) = \mathbf{0_{r\times 1}} \tag{2.8}$$

or

$$H\left(I_N \otimes C(z)\right)a(0) = b(z)$$

where H is an $r \times N^2$ selection matrix, $a(0) = \text{vec}(A(0))$, and $b(z) = H\text{vec}(A(z))$.

Using equations (2.5) and (2.8), the identifying restrictions can be written as:

$$g(a(0), \beta, z) = \begin{bmatrix} \text{vech}\left(\Sigma - A(0)A(0)'\right) \\ H\left(I_N \otimes C(z)\right)a(0) - b(z) \end{bmatrix} = \mathbf{0} \tag{2.9}$$

where $\mathbf{0}$ is a conformable vector of zeroes.

These moment conditions (2.9) define a system of estimating equations linking the (structural) parameters of interest α (here $a(0)$) and the vector of auxiliary parameters β. This estimation framework corresponds to the asymptotic least squares procedure proposed by Gourieroux et al. (1985) and Gourieroux and Monfort (1995).

Indeed, using some consistent estimates of β, denoted $\widehat{\beta}_T$, an estimator of α can be obtained by minimizing the system of estimating equations in a given metric. More specifically, a first step proceeds with the estimation of the reduced-form parameters β whereas a second step makes use of the asymptotic least squares procedure. To the best of our knowledge, the asymptotic least squares procedure has not yet been used to study SVARs in a unified framework. By appropriately defining the fonction $g(\alpha, \widehat{\beta}_T, z)$ that maps the parameters of interest α and the auxiliary parameters β, we show that the ALS methodology allows to jointly investigate a large class of identification, estimation and inference issues based on SVARs. The next section presents several examples in the case of full or partial identification and estimation of structural IRFs, as well as when estimating and validating structural dynamic models based on structural IRFs.

Turning now to any frequency interval $\omega \in [\underline{\omega}, \overline{\omega}] \subseteq [-\pi, \pi]$, we can consider a continuum of frequency conditions indexed by ω as follows:

$$g(a(0), \beta, \omega) = \begin{bmatrix} \text{vech}\left(\Sigma - A(0)A(0)'\right) \\ H\left(I_N \otimes C(e^{-i\omega})\right)a(0) - b(z) \end{bmatrix} = \mathbf{0}. \tag{2.10}$$

Going back to the example of Galí (1999) and thus the standard Blanchard-Quah long-run restriction (equation (2.8)), the continuum of estimating equations $H\left(I_N \otimes C(e^{-i\omega})\right) a(0) - b(z)$ in a (symmetric) neighborhood of $\omega = 0$, $\omega \in [-\overline{\omega}; \overline{\omega}]$, is given by:

$$g(a_2(0), \widehat{\beta}, \omega) \equiv \widehat{c}_{11}(e^{-i\omega})a_{12}(0) + \widehat{c}_{12}(e^{-i\omega})a_{22}(0) = 0$$

where $b(z) = 0$. Combining the continuum of restrictions with the relationship between the variance-covariance matrix of the reduced form error terms and the structural shocks, $\Sigma = A(0)A(0)'$, leads to identify the vector of structural parameters $a(0)$.[9]

Similar to the case of a given frequency ω, a two-step procedure can be applied but now on a continuum of estimating equations $g(\alpha, \widehat{\beta}_T, \omega)$ for $\omega \in [\underline{\omega}, \overline{\omega}] \subseteq [-\pi, \pi]$ with $\alpha = a(0)$. Importantly, the system of equations based on a continuum of frequencies is overidentified even if the original structural VAR model for a given frequency, say $\omega = 0$, is just-identified. An overidentification test can then be performed to assess the reliability of the imposed restrictions on the frequency band. This also raises the question of the selection of the frequency interval on which the identifying restrictions might be satisfied. This issue is further discussed in Section 4 by means of an information criterion procedure (Hall et al., 2012). Accordingly, using a larger window of frequencies might provide a more efficient identification and estimation of the parameters of interest α.

3 Examples

This section provides some applications of the asymptotic least squares theory using a frequency band or equivalently a continuum of estimating equations. The first two applications, namely the identification of technology and news shocks, are further discussed as applications in Section 7. The third application regarding neutral versus investment technology shocks is presented formally in Section 5. The fourth application is related to the recoverability condition of Chahrour and Jurado (2021) and the structural identification of expectations-driven fluctuations. The last two applications, the identification and inference of common features, and the estimation and validation of dynamic stochastic structural models with structural VAR in the frequency domain, are left for future work.[10]

3.1 The hours-productivity debate using bivariate SVAR models

The predominant role of technology shock as the main source behind movements in macro data has been sharply challenged since the important contribution of Galí (1999). Indeed, bivariate structural vector au-

[9] See also Section 5.2.

[10] It is worth emphasizing that this section is not exhaustive in the sense that our methodology applies for general zero restrictions, i.e. with a mixture of short-run and long-run restriction (e.g., in the monetary policy model of Rubio-Ramirez et al. (2010)), as well as for structural VARs based on present value models (e.g., Campbell and Shiller (1987,1988)), for structual VECM (e.g., King et al. (1991)) and for the identification and contribution of seasonal cycles versus business cyles (Wen, 2002), among others.

toregressive models including labor productivity and hours worked yield conflicting results on the effect of technology shocks on hours worked, generally due to the assumed data generating process for the measure of hours worked (in level or in difference). On the other hand, Francis and Ramey (2009) show that demographic trends and sectoral allocation are important sources of low-frequency movements in hours worked and labor productivity. Consequently, 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. A SVAR model with some identifying restrictions on a frequency band allows to focus in a neighborhood more or less close to $\omega = 0$ rather than just the zero frequency, and thus to assess the effects of other low-frequency movements, such as those advocated by Francis and Ramey (2005), on the identification of technology shocks. In Section 5.2, we show how the technology shock can be identified by restrictions (2.11) and by the equality between the variance-covariance matrix of the reduced form error terms and the structural shocks, $\Sigma = A(0)A(0)'$.

Our strategy has three advantages with respect to the usual long-run restrictions. First, the set of identifying restrictions in the frequency domain has a Lebesgue measure strictly greater than zero. Indeed the fact that standard long-run restrictions (when $\omega = 0$) have a zero Lebesgue measure (Faust, 1996) leads to the so-called unreliability problem (Faust and Leeper, 1997). Notably, unreliable long-run effects of shocks are transferred on estimates of other model parameters through the long-run identification scheme and thus any test of the null hypothesis that the kth coefficient of an autoregressive polynomial in the SVAR equals zero is not consistent, i.e., the test has significance level greater than or equal to maximum power. In addition, one cannot compute asymptotically correct confidence intervals for impulse responses since the unreliability of the long run effect estimator is transferred to the estimator of the dynamic multipliers of the structural shocks. In contrast, and as suggested by Faust and Lepper (1997), Faust (1998) and Pötscher (2002), this issue can be circumvented by imposing restrictions on the (long-run) effect of these shocks at non-zero frequencies and not only at the zero-frequency so that the problem is no-longer ill-posed in the terminology of Pötscher (2002). Second, the overidentification testing procedure proposed in the next section allows to assess the hypothesis that only the first shock drives the long-run movements in labor productivity (Francis and Ramey, 2009). Third, using a wider frequency band relative to the zero frequency should help to better identify and estimate the structural shocks driving long-run movements.

3.2 News shock

Using structural VAR's and partial identification schemes, recent empirical literature delivered controversial results concerning the role of anticipated neutral technology—news—shocks in business cycle fluctuations. By imposing long-run restrictions, Beaudry and Portier (2006) and Beaudry and Lucke (2010) conclude that news shocks about future productivity are one of the main drivers of business cycles and there is a positive (contemporaneous) impact of the news shock on hours worked. These results have been challenged in several dimensions. For instance, adopting partial identification schemes based on different max-share approaches, Barsky and Sims (2011) and Kurmann and Sims (2019) find results incompatible with the news-driven explanation of business cycles and thereby more in line with the implications of the standard neoclassical framework. These alternative identification strategies are based on the forecast error variance decomposition over a horizon of up to 40 quarters (Barsky and Sims, 2011) or 80 quarters (Kurmann and Sims, 2019): both strategies encompassing short-run and business cycle fluctuations. In contrast, using an appropriate frequency band (see Section 7.2.) has the advantage to focus on the medium and long-run frequencies of TFP and thus allows to isolate the identification of the news shocks from the effects of short-run and business cycles fluctuations. In addition we can test whether TFP is driven by one structural shock (the news shock) or two structural shocks (the surprise TFP shock and the news shock) in the medium to long-run. In the latter, any linear combination of the two structural shocks would be a main driver and thus there is a lack of proper identification.

3.3 Neutral *versus* investment-related technology shocks

Fisher (2006) examines the relative importance of neutral technology shocks and the investment-related technology shocks in the explanation of business cycles by incorporating long-run restrictions that separately identify these two sources of technology shocks.[11] Pursuing this decomposition, Chen and Wemy (2015) argue that long-run movements in the capital-producing sector can spread and spillover to the rest of the economy and enhance TFP in long-run. Accordingly, the long-run fluctuations of TFP may be characterized by two stochastic trends driven unequivocally either by the long-run movements of specific TFP or the spillover effect of long-run movements due to investment-specific technological (IST) changes. This implies that long-run movements of the TFP series would be caused by two shocks while long-run movements of investment-specific technology are only driven by its own (structural) shock. In this respect, consider a SVAR in which the first variable is the IST series and the second variable is the TFP series of Fernald (2014). The two restrictions that only one shock has long-run effects on IST, and the same shock and the neutral technology shocks have long-run effects on TFP lead to the following restrictions for an interval of frequencies around zero:

$$C(e^{-i\omega})A(0) = A(e^{-i\omega}) = \begin{bmatrix} a_{11}(e^{-i\omega}) & 0 & 0_{1\times(N-2)} \\ a_{21}(e^{-i\omega}) & a_{22}(e^{-i\omega}) & 0_{1\times(N-2)} \\ \tilde{A}_{31}(e^{-i\omega}) & \tilde{A}_{32}(e^{-i\omega}) & \tilde{A}_{33}(e^{-i\omega}) \end{bmatrix}$$

where $\tilde{A}_{31}(e^{-i\omega})$ is the first column of the matrix $A(e^{-i\omega})$ after dropping the two first elements $a_{11}(e^{-i\omega})$ and $a_{21}(e^{-i\omega})$ and the column vector $\tilde{A}_{32}(e^{-i\omega})$ is the corresponding second column and the submatrix $A_{33}(e^{-i\omega})$ contains the other columns of $A(e^{-i\omega})$ (except the two first rows of those columns). Proceeding with a frequency band in a neighboorhood of $\omega = 0$ allows to perform a statistical test on the assumption that one stochastic trend against the alternative hypothesis of two stochastic trends drives the long-run movements of TFP. We further discuss formally this example in Section 5.1.

[11]See Ramey (2016) for a survey on the debate on the relative importance of the neutral TFP shocks and the investment specific technology shocks. See also Ben Zeev and Khan (2015) for the effect of IST news shocks.

3.4 Recoverability and expectations-driven fluctuations

Chahrour and Jurado (2021) propose an identification condition of structural shocks, which is less restrictive than the usual condition of fundamentalness—the so-called recoverability. This condition only imposes that the structural shocks can be recovered from the past, present and future observables available to the econometrician. Indeed, the econometrician has access to the entire sample to identify the structural shocks and not only to the information available in the observables up to time t as required by fundamentalness. Said differently, the econometrician can also use available observables at time $t+1, \ldots, t+h$ to infer structural shocks at time t. In this respect, the necessary and sufficient condition of recoverability depends on the invertibility of the Fourier transform of the two-side moving average representation of the observable variables as function of the structural shocks.

Consider the example in Section 3 of Chahrour and Jurado (2021) regarding the identification of a noise and a fundamental structural shock about technology as potential drivers of business cycle fluctuations. Their VAR procedure can be summarized as follows. On the one hand, the VAR estimation is conducted with a set of observables/variables including a technology measure (a_t). On the other hand, one can get the joint spectral density of technology a_t as well as the optimal forecast of technology b_t implied by the VAR. Finally, the identification of the structural shocks is then achieved by matching the joint spectral density resulting from the VAR and the joint spectral density implied by the structural two-side moving average of a_t and b_t as function of the two structural shocks for the interval $\omega \in [-\pi, \pi]$. In our framework, this can be written as:

$$g(\alpha, \widehat{\beta}, \omega) = vech\left(\widehat{C}_{ab}(e^{-i\omega})\widehat{\Sigma}\widehat{C}_{ab}(e^{-i\omega})^* - A_{ab}(e^{-i\omega})A_{ab}(e^{-i\omega})^*\right)$$

for $\omega \in [-\pi, \pi]$ where $\widehat{C}_{ab}(e^{-i\omega})\widehat{\Sigma}\widehat{C}_{ab}(e^{-i\omega})^*$ is the estimated joint spectral density of the technology series a_t and its optimal forecast b_t implied by the VAR, and $A_{ab}(e^{-i\omega})A_{ab}(e^{-i\omega})^*$ is the structural joint spectral density compatible with the identification restrictions. Then Chahrour and Jurado (2021) solves a minimization problem by using the factorization proposed by Rozanov (1967), and thus can derive the mapping from observables to structural disturbances, as well as the impulse response functions and the (historical) variance decomposition. An optimal two-step C-ALS estimator can then be performed allowing to formally test the imposed identification restrictions that lead to recover the noise and the fundamental technology shocks. In the event of rejection, the measure of technology a_t could be contaminated by measurement errors or short-run fluctuations such that the identification restrictions hold only for medium-run and long-run frequencies. The C-ALS framework is well-suited to investigate such a conjecture.

3.5 Common features in the frequency domain

Since the seminal contribution of Engle and Kozicki (1993), a common feature can be defined as follows: a feature is common if a group of variables of interest possesses this feature and a combination of these variables does not have the feature. Canonical examples include cointegration in which some (all) variables have stochastic trends but some linear combinations of these variables do not have stochastic trends; common serial correlation in which some linear combinations of serially correlated variables correspond to a weak white noise process (Engle and Kozicki, 1993); common cycles in which linear combinations of the cycle components of a group of variables have no cyclical component (Vahid and Engle, 1993, and Hecq, Palm and Urbain, 2006).[12] It turns out that this concept of a common feature can be extended to the existence of common business cycles, i.e. a set of series is characterized by some common business cycle fluctuations whereas some linear combinations does not have this feature.[13]

To go one step further, suppose that the number of structural shocks is less than the number of variables and thus that there exist some common business cycles, say on a frequency band $\omega \in [\underline{\omega}, \overline{\omega}]$. In this case, it turns out that the matrix $C(e^{-i\omega})$ have less than full rank for all $\omega \in [\underline{\omega}, \overline{\omega}]$, meaning that the left null space of the matrix $C(e^{-i\omega})$ is non-empty:

$$C(e^{-i\omega})A(0) = A(e^{-i\omega}) = \begin{bmatrix} A_1(e^{-i\omega}) & A_2(e^{-i\omega})_{m\times s} \end{bmatrix} = \begin{bmatrix} A_1(e^{-i\omega}) & 0_{N\times s} \end{bmatrix}.$$

Accordingly, there exists a set of s linear independent combinations such that the rank of the matrix $A(e^{-i\omega})$ is equal to $N - s$ and thus this rank restriction allows the identification of a subset of structural shocks. Moreover, since these linear combinations define a set of estimating equations, an overidentification test, such as the one proposed in Section 4.4, can be conducted and interpreted as a reduced-rank test of common business cycles.

3.6 Estimation and validation of dynamic structural models with SVAR in the frequency domain

Rotemberg and Woodford (1999) and Christiano, Eichenbaum and Evans (2005) propose a limited information econometric strategy to estimate and evaluate dynamic structural models by matching impulse response functions. This is based on the minimization of the distance between a selected subset of impulse response functions from a SVAR with the corresponding impulse response functions from a DSGE model. For example, Christiano, Eichenbaum and Evans (2005) select the impulse response functions of eight macroeconomic variables to an identified monetary shock using a SVAR with short-run restrictions. Since DGSE models aim at explaining business cycles, it then appears natural to estimate and evaluate the structural model by matching the impulse responses but at the business cycle frequencies.[14]

Let $\widehat{A}_j(e^{-i\omega})$ denote the vector containing the impulse response functions estimates of the variables of interest to the structural shock j, which depends on the reduced-form parameter estimates of β, and $\widetilde{A}_j(e^{-i\omega}, \alpha)$

[12] Other common features have been proposed in the literature which include common seasonally Engle and Hylleberg, 1996), codependence (Gouriéroux and Peaucelle, 1992), common structural breaks (Hendry, 1999) among others.

[13] Angeletos, Collard and Dellas (2020) find some support for a main business-cycle driver which implies that the business cycles fluctuations can be explained by a small number of structural shocks.

[14] See Christiano and Vigfusson (2003), Diebold, Ohanian and Berkowitz (1998) for the estimation and the evaluation of DSGE models in the frequency domain.

the mapping from the vector of structural parameters α to the corresponding model impulse responses.[15] An estimator of the structural parameter vector α of the DSGE model is given by the following minimization problem :

$$\widehat{\alpha} = \text{argmin}\|g(\alpha, \widehat{\beta}, \omega)\|_W^2$$

with $g(\alpha, \widehat{\beta}, \omega) = \widehat{A}_j(e^{-i\omega}) - \widetilde{A}_j(e^{-i\omega}, \alpha)$ for all $\omega \in [\underline{\omega}, \overline{\omega}]$ and W is a weighting matrix which can depend on the frequency interval. The estimating equations are then defined on a continuum, i.e. the frequency interval, and the methodology developed hereafter can be applied to estimate and evaluate the structural DSGE model.[16]

4 Asymptotic Least Squares in the frequency domain

In this section, we propose a general asymptotic least squares estimator in the presence of a continuum of restrictions (hereafter, C-ALS estimator). After briefly reviewing the asymptotic least squares theory proposed by Gourieroux et al. (1985) and then fixing some notations, we proceed in two steps. 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.

4.1 Notation

Let the q-vector $\widehat{\beta}_T$ denote a first-step M-estimator defined by:

$$\widehat{\beta}_T = \arg\min_{\beta \in \mathcal{B}} Q_T(Z_T, \beta) \tag{4.11}$$

where $\beta_0 = P_0 \lim_{T\to\infty} \widehat{\beta}_T$ denotes the true unknown value of the instrumental parameters, P_0 the true unknown probability distribution of the data generating process, and Q_T the (sample) objective function. Under standard regularity conditions $\sqrt{T}\left(\widehat{\beta}_T - \beta_0\right) \xrightarrow{d} \mathcal{N}(0, \Omega)$ and $\Omega = \lim_{T\to\infty} \text{Var}\left(\sqrt{T}\widehat{\beta}_T\right)$ under P_0.

The asymptotic least squares method consists in estimating the parameters of interest through J constraints:

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

where $\widehat{\beta}_T$ is the vector of auxiliary parameters from the first-step estimation procedure and α is vector of the parameters of interest such that $\alpha_0 = P_0 \lim_{T\to\infty} \widehat{\alpha}_T$ is the true unknown value. The ALS estimator is

[15] The impulse response functions from the model can be computed directly from the state-space representation of the linearized solution of the model or by simulations (indirect inference).

[16] Diebold, Ohanian and Berkowitz (1998) propose to estimate the spectral density using the Blackman-Tukey lag-window approach instead of using the VAR-based spectral density and an approximation of the corresponding integral by a sum at the selected (discretized) frequencies.

thus defined by:

$$\begin{aligned}\widehat{\alpha}_T(S_T) &= \arg\min_{\alpha\in\mathcal{A}} g(\alpha,\widehat{\beta}_T)' S_T g(\alpha,\widehat{\beta}_T) \\ &= \arg\min_{\alpha\in\mathcal{A}} \left\| S_T^{1/2} g(\alpha,\widehat{\beta}_T)\right\|^2\end{aligned}$$

where S_T is a symmetric positive definite matrix which possibly depends on the observations and $S = P_0 \lim_{T\to\infty} S_T$.

Gourieroux et al. (1985) and Gourieroux and Monfort (1995) show that under suitable regularity conditions the estimator $\widehat{\alpha}_T(S_T)$ is asymptotically normally distributed with:

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

where

$$\begin{aligned}Q(S) &= \left(\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0) S \frac{\partial g'}{\partial \alpha}(\alpha_0,\beta_0)\right)^{-1}\left(\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0) S \frac{\partial g}{\partial \beta'}(\alpha_0,\beta_0)\Omega\frac{\partial g'}{\partial \beta}(\alpha_0,\beta_0) S \frac{\partial g'}{\partial \alpha}(\alpha_0,\beta_0)\right) \\ &\times \left(\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0) S \frac{\partial g'}{\partial \alpha}(\alpha_0,\beta_0)\right)^{-1}.\end{aligned}$$

An optimal ALS estimator is then obtained for a sequence of matrices S_T converging to

$$S_0 = \left(\frac{\partial g}{\partial \beta'}(\alpha_0,\beta_0)\Omega\frac{\partial g'}{\partial \beta}(\alpha_0,\beta_0)\right)^{-1}$$

and the corresponding optimal asymptotic variance is equal to:

$$Q(S_0) = \left(\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0)\left(\frac{\partial g}{\partial \beta'}(\alpha_0,\beta_0)\Omega\frac{\partial g'}{\partial \beta}(\alpha_0,\beta_0)\right)^{-1}\frac{\partial g'}{\partial \alpha}(\alpha_0,\beta_0)\right)^{-1}.$$

It is worth emphasizing that local identification requires a full rank condition on the Jacobian matrix of the estimating equations (2.9) with respect to the vector $a(0)$ at the true value:

$$\texttt{rk}\left(\frac{\partial g(\alpha,\beta)}{\partial \alpha'}|_{\alpha=\alpha_0}\right) = N^2$$

where $\alpha = a(0)$. Notably the necessary part implies that $\texttt{dim}(\alpha) \leq \texttt{dim}(g(\alpha,\beta)$.[17] Accordingly, the number of estimating equations is constrained by the dimension of the vector β in the case of an optimal weighting matrix, i.e. $\texttt{dim}(g(\alpha,\beta)) \leq \texttt{dim}(\beta)$. This condition also holds to preserve the validity of standard asymptotic distributions for statistical tests based on impulse response functions.[18] Finally, this rank condition also needs to be satisfied for either a given frequency or a frequency band.[19]

[17]It corresponds to Proposition 9.2 of Lütkepohl (2017) in the context of structural VAR's.

[18]See Inoue and Killian (2016) and Guerron-Quintana, Inoue and Killian (2017) for inference when the dimension of a set of impulse responses investigated exceeds the dimension of the vector β.

[19]Intuitively, suppose that the identifying restrictions are observed on a discrete finite support, i.e. one proceeds with a discretization of the frequency interval, say $\underline{\omega} = \omega_1 < \omega_2 < \cdots < \omega_m = \overline{\omega}$, and thus there are $r \times m$ identifying restrictions. In this respect, the rank condition would be satisfied if the discretization scheme is not too thin and thus avoids redundant moment conditions. To circumvent this issue of a discretized asymptotic least squares estimator on a frequency band, we rather consider a continuum perspective and make use of a suitable kernel operator (and its regularized version). Note that such a discretized asymptotic least squares estimator is derived for a bivariate VAR model in the supplementary material.

In contrast, we consider a system of J constraints defined on a continuum of frequencies (e.g., equation (2.10)) with the true vector of interest α_0 under P_0 . These constraint functions are complex-valued and indexed by a parameter ω taking its values on the interval $\mathcal{I} = [-\pi, \pi]$:

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

where $g(\cdot,\cdot,\omega)$ takes its values in $H = (L^2(\mathcal{I},\varphi))^J$, a Hilbert space with the inner product $< .,. >$ and the norm $\|\cdot\|$. $L^2(\mathcal{I},\varphi) \equiv L^2(\varphi)$ is the space of complex valued functions that are square integrable with respect to the probability density function φ of a distribution for ω.[20] Let S denote a bounded linear operator defined on $(L^2(\mathcal{I},\varphi))^J$ or a subspace of $(L^2(\mathcal{I},\varphi))^J$ and $\overline{g(\cdot,\cdot,\omega)}$ denote the complex conjugate of $g(\cdot,\cdot,\omega)$.[21]

4.2 The class of C-ALS estimators

The C-ALS estimator is defined by replacing the common sequence of symmetric positive definite matrices in the GMM literature by a converging sequence of random bounded linear operators (Carrasco and Florens, 2000; Carrasco et al., 2007). Hence, for a given sequence, say S_T, converging to S an operator from $(L^2(\mathcal{I},\varphi))^J$ in $(L^2(\mathcal{I},\varphi))^J$, the C-ALS estimator is defined by:

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

Therefore the C-ALS estimator renders the constraints, $g(\alpha, \widehat{\beta}_T, \omega) = 0$ for $\omega \in [\underline{\omega}, \overline{\omega}]$, as close as possible to zero by using the metric associated with the inner product defined by S_T and φ is the uniform probability measure on the interval $[-\pi, \pi]$. For instance, if ω belongs to the interval $[\underline{\omega}, \overline{\omega}]$, the C-ALS estimator is given by:

$$\widehat{\alpha}_T(S_T) = \arg\min_{\alpha\in\mathcal{A}} \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\omega}}^{\overline{\omega}} g(\alpha, \widehat{\beta}_T, \omega_1)' S_T(\omega_1, \omega_2) \overline{g(\alpha, \widehat{\beta}_T, \omega_2)} d\omega_1 d\omega_2.$$

The following proposition provides its asymptotic properties.

Proposition 4.1. *Suppose that Assumptions A.1 to A.11 are satisfied and that S_T denote a sequence of random bounded linear operators converging to S, the C-ALS estimator associated with S_T is a solution $\widehat{\alpha}_T(S_T)$ to the problem:*

$$\widehat{\alpha}_T(S_T) = \arg\min_{\alpha\in\mathcal{A}} \|S_T^{1/2} g(\alpha, \widehat{\beta}_T, \omega)\|^2. \tag{4.12}$$

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

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

[20] All assumptions are given in Appendix 1.

[21] 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).

with

$$Q(S) = \left\langle S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0,\omega), S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0,\omega)\right\rangle^{-1}\left\langle S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0,\omega), \left(S^{1/2}K\overline{S^{1/2}}\right)S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0,\omega)\right\rangle \times\left\langle S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0,\omega), S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0,\omega)\right\rangle^{-1}$$

where $K : (L^2(\varphi))^J \to (L^2(\varphi))^J$ *is an integral (Hilbert-Schmidt) operator such that for all* f_l *in* $L^2(\varphi)$*:*[22]

$$Kf(\omega_1) = \left(\sum_{l=1}^{J}\int E^{P_0}\left[k_{jl}(\omega_1,\omega_2)\right]f_l(\omega_2)\varphi(\omega_2)d\omega_2\right)_{j=1,\ldots,J}$$

with

$$k_{jl}(\omega_1,\omega_2) = \frac{\partial g_j}{\partial \beta'}(\alpha_0,\beta_0,\omega_1)\Omega\overline{\frac{\partial g_l}{\partial \beta}(\alpha_0,\beta_0,\omega_2)'}.$$

Proof: See Appendix 2.

Several points are worth commenting. First, the previous proposition is an implication of the following functional convergence (as $T \to \infty$):

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

where $N(0,K)$ is the Gaussian random vector of $(L^2(\varphi))^J$.[23] Second, the Hilbert-Schmidt operator K is not invertible on the full reference space but has a finite dimensional closed range (at most) equal to q—the dimension of the (asymptotic) variance-covariance matrix of the parameter vector β. This differs from the framework of Carrasco and Florens (2000) and Carrasco et al. (2007) in which the inverse of K is not bounded because the range of K is not closed in general.[24,25] In contrast, since the operator K depends on the first step estimation through $\widehat{\beta}_T$ in our approach, the range of K, denoted $R(K)$, is then known and $R(K) = \mathtt{dim}(\beta)$. However, as explained below, the derivation of the (Moore-Penrose) generalized inverse of K can be cumbersome in finite samples due to the presence of tiny eigenvalues and thus a regularization method cannot be precluded in our framework.

Third, since the range of K equals (at most) q, the number of its eigenvalues different from zero is finite and, according to the Mercer's Theorem, K admits the following spectral decomposition:

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

[22]See Wahba (1992) for reproducing kernel Hilbert space (RKHS) of vector-valued functions.

[23]See Chen and White (1998) and Carrasco and Florens (2000, remark 2 p. 803) for a univariate version of the functional central limit theorem. A multivariate version can be established by applying the functional Cramér-Wold device (White, 2000).

[24]For a general discussion of linear inverse problems in econometrics, see Carrasco, Florens and Renault (2007).

[25]Indeed the Moore-Penrose inverse operator K^{-1} is not bounded and the solution $K^{-1}f$ to a Fredholm equation of the first kind $K\phi = f$ is not continuous in f, i.e. K does not admit a generalized inverse over the entire Hilbert space of reference. Consequently, to guarantee the stability of the solution, Carrasco and Florens (2000) replace the operator K by some nearby operator (e.g., using a Tikhonov regularization)—see also Carrasco (2012).

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 $i = 1, \ldots, q$. It follows that

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

This implies that the Moore-Penrose generalized inverse of K, denoted K^{-1}, satisfies:

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

and that a consistent estimator of K^{-1} can be obtained as follows.

Proposition 4.2. *Let $\widehat{\alpha}_T^1$ denote a first-step consistent estimator of α_0. A consistent estimator of the Moore-Penrose generalized inverse is defined by:*

$$K_T^{-1}f(\omega_1) = \sum_{i=1}^{q} \frac{1}{\lambda_{i,T}} \gamma_{i,T}(\omega_1) \langle f, \gamma_{i,T} \rangle$$

where $\gamma_{i,T}(\omega_1)$ is given by:

$$\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega_1)\widehat{\Omega}_T^{1/2} D_i,$$

the eigenvalues $\lambda_{i,T}$ are those of the $q \times q$ matrix:

$$\int \widehat{\Omega}_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega_2)} \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega_2)\widehat{\Omega}_T^{1/2} \varphi(\omega_2) d\omega_2,$$

and the matrix $D = [D_1 \cdots D_q]$ and the diagonal matrix Λ of eigenvalues λ_i satisfy

$$\int \widehat{\Omega}_T^{1/2} \overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega_2)} \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega_2)\widehat{\Omega}_T^{1/2} \varphi(\omega_2) d\omega_2 D = D\Lambda.$$

where $\widehat{\Omega}_T$ is a consistent estimate of Ω.

Proof: See Appendix 2.

The eigenfunctions $\gamma_{i,T}(\omega)$ for $i = 1, \ldots, q$ are linear combinations that result from the (orthogonal) projection of the (nonlinear) estimating equations $g(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega)$ onto the subspace spanned by $\widehat{\beta}_T$, namely $\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1, \widehat{\beta}_T, \omega)\widehat{\Omega}_T^{1/2}$. It makes sense since the estimation of the parameters of interest α is completely determined by the estimator of the auxiliary parameters $\widehat{\beta}_T$.

4.3 The optimal C-ALS estimator

Using Proposition 4.1, it can be shown that choosing $S_T = K_T^{-1/2}$ leads to the estimator of minimum variance and thus the optimal C-ALS estimator.[26]

Proposition 4.3. *Let K_T denote a consistent estimator of K and K_T^{-1} a Moore-Penrose generalized inverse estimator of K^{-1}. The optimal C-ALS estimator of α_0 is given by:*

$$\widehat{\alpha}_T = \arg\min_{a \in \mathcal{A}} \|K_T^{-1/2} g(\alpha, \widehat{\beta}_T, \omega)\|^2 \tag{4.13}$$

and $\widehat{\alpha}_T$ is consistent and asymptotically normally distributed:

$$\sqrt{T}\left(\widehat{\alpha}_T(K_T^{-1/2}) - \alpha_0\right) \xrightarrow{d} N\left(0, \left\langle \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0), K^{-1}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0)\right\rangle^{-1}\right).$$

Proof: See Appendix 2.

Proposition 4.13 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 reaches the asymptotic efficiency by exploiting all the information contained in the interval of frequencies.

Taking the eigenfunctions and eigenvalues decomposition of K_T, the optimal C-ALS objective function to minimize can be rewritten as:

$$\|K_T^{-1/2} g(\alpha, \widehat{\beta}_T, \omega)\|^2 = \sum_{i=1}^{q} \frac{1}{\lambda_{i,T}} \left\langle g(\alpha, \widehat{\beta}_T, \omega), \gamma_{i,T}(\omega)\right\rangle^2.$$

However, this computation can be burdensome, especially for large q. As in Carrasco et al. (2007), we propose a simple expression of the objective function. This requires a first-step consistent C-ALS estimator, denoted $\widehat{\alpha}_T^1$, defined by (using the identity operator as a kernel operator):

$$\widehat{\alpha}_T^1 = \arg\min_{\alpha \in \mathcal{A}} \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\omega}}^{\overline{\omega}} g(\alpha, \widehat{\beta}_T, \omega_1)' \overline{g(\alpha, \widehat{\beta}_T, \omega_2)} d\omega_1 d\omega_2.$$

Proposition 4.4. *A simplified expression for the objective function of the 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

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

is a $q \times q$-matrix and

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

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

[26] Intuitively, S is chosen such that $S^{1/2} K \overline{S^{1/2}}$ is equal to the identity operator.

Proof: See Appendix 2.

Intuitively, the system of functions $\underline{s}(\alpha, \widehat{\beta}_T)$ corresponds to the orthogonality conditions between the estimating equations that link α and $\widehat{\beta}_T$, and the projection of this nonlinear system of equations $g(\alpha, \widehat{\beta}_T, \omega)$ 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, one key issue is that the matrix W might not be of full rank q, especially 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.[27]

Proposition 4.5. *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., 2007; Carrasco, 2012).

Note that the finite sample behavior of this regularized estimator does depend obviously on the choice of the regularization parameter. Using a higher-order expansion of the mean squared error, Carrasco (2012) proposes a data-driven procedure for selecting such a parameter for estimation purposes. However, such a procedure might not be well-suited for testing procedures.[28] One key issue is that the variance-covariance matrix affects obviously the finite sample properties of the test statistics that rest on the parameters of interest. This is also the case when considering some overidentification tests or interval selection tests (see Sections 4.4 and 4.5). Yet the literature has only provided asymptotic theory (for estimation) when the regularization parameter goes to zero at some rate, but there is no guideline regarding the selection of the smoothing parameter in finite samples for statistical tests.

4.4 Test of overidentification

Using Carrasco and Florens (2000), a test of overidentification can also be performed using the following J-statistic.

Proposition 4.6. *The overidentification test is based on the statistic:*

$$J_T = \|\sqrt{T}K_T^{-1/2} g(\widehat{\alpha}, \widehat{\beta}_T, \omega)\|^2$$

[27] Different regularization schemes can be used: Tikhonov, Landweber-Fridman, spectral cut-off or principal components regularization (Carrasco, 2012). The first three aforementioned methods have a long tradition in statistics (Kress, 1999) whereas the principal components approach is widely used in factor models (Stock and Watson, 2002). All of these methods do involve a regularization term η_T that is defined on a continuous support (Tikhonov) or a discrete support (the other three approaches).

[28] See also Carrasco and Kochoni (2017, 2019).

The statistic J_T is χ^2-distributed with $q-r$ degrees of freedom.

In the presence of a rank-order deficiency, Proposition 4.5 can be used to define a regularized version of the test statistic. More specifically, for a given value of η_T, the J_T statistic converges to the following distribution:

$$T\underline{s}(\widehat{\alpha}_T,\widehat{\beta}_T)'[\eta_T I_q+\widehat{W}^2]^{-1}\overline{\underline{s}(\widehat{\alpha}_T,\widehat{\beta}_T)} \xrightarrow{d} \sum_{j=1}^{q}\frac{\lambda_j^2}{\lambda_j^2+\eta_T}Z_i^2 \tag{4.14}$$

where $\{Z_i\}_{i=1}^q$ are independent standard normal variates and the λ_j terms are the eigenvalues of the matrix W. As $\{Z_i\}^2\sim\chi^2(1)$, the limiting distribution is a weighted sum of independent Chi-squared variables (see Arellano, Hansen and Sentana, 2012; Vuong, 1989).

Interestingly, as shown in Proposition 4.7, a specification test can still be performed for any arbitrary positive definite matrix S_T (e.g., the identity operator) that depends on ω.

Proposition 4.7. *The J_T statistic based on the objective function of Proposition 4.1 for any S_T has the following asymptotic distribution:*

$$\|S_T^{1/2}\sqrt{T}g(\widehat{\alpha}_T,\widehat{\beta}_T,\omega)\|^2 \xrightarrow{d} \sum_{i=1}^{q}\lambda_i Z_i^2 \tag{4.15}$$

where λ_i are the eigenvalues of the asymptotic variance matrix of $S_T^{1/2}\sqrt{T}g(\widehat{\alpha}_T,\widehat{\beta}_T,\omega)$ given by:

$$\lim_{T\to\infty} Var\left(S^{1/2}\sqrt{T}g(\widehat{\alpha}_T,\widehat{\beta}_T,\omega)\right) = \left\langle S^{1/2}[I-M(\omega)],(S^{1/2}K\overline{S^{1/2}}')S^{1/2}[I-M(\omega)]\right\rangle$$

where $M(\omega)=S^{1/2}\frac{\partial g}{\partial\alpha'}(\alpha_0,\beta_0,\omega)\left\langle S^{1/2}\frac{\partial g}{\partial\alpha'}(\alpha_0,\beta_0,\omega),S^{-1/2}\frac{\partial g}{\partial\alpha'}(\alpha_0,\beta_0,\omega)\right\rangle^{-1}\overline{\frac{\partial g'}{\partial\alpha}(\alpha_0,\beta_0,\omega)S^{1/2}}$.

Critical values of the limiting distribution (4.14) and (4.15) can be obtained either by implementing the numerical inversion of the characteristic function proposed by Imhof (1961) or by simulating independent Chi-squared distributions (see Robin and Smith, 2000)—both methods requiring a *plug-in* procedure and thus to replace the true eigenvalues by some consistent estimates. These critical values can also be approximated by a gamma distribution (Shorack, 2000).[29]

4.5 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 sake of simplicity, we consider the class of symmetric intervals of ω around zero, i.e. $I_\omega=(-\omega,\omega)$ and we use the information criteria-based methodology of

[29]The theoretical justification of a CLT approximation based on a Gamma distribution for a summand of random variables when the underlying distribution is positively skewed is given in Shorack (2000), Theorem 4.1. In this case, the distribution is function of two parameters: the shape parameter ζ and the scale parameter δ. These parameters are obtained by using an estimator of the mean and standard deviation of $\sum_{i=1}^m\lambda_i Z_i^2$ given by $\zeta\delta$ and $\zeta\delta^2$ respectively. Taking that $\mathbb{E}\left[\sum_{i=1}^m\lambda_i Z_i^2\right]=\sum_{i=1}^m\lambda_i$ and $\mathbb{V}\left[\sum_{i=1}^m\lambda_i Z_i^2\right]=2\sum_{i=1}^m\lambda_i$, an estimator of the mean and the variance is obtained by replacing λ_i by $\hat{\lambda}_i$.

Hall et al. (2012) in order to propose a statistical criterion which selects the largest interval I_ω that might guarantee consistent estimation of $\widehat{\alpha}_T$. In so doing, we select the frequency interval by minimizing the Valid Interval Selection Criterion (VISC) defined by:

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

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

$$\texttt{VISC}_T(\omega) = J_T(\omega) - h(|w|)\kappa_T \tag{4.16}$$

where $h(|w|)\kappa_T$ is a deterministic penalty, which is an increasing function of the length of the interval. Proposition 4.8 shows that $\hat{\omega}_T$ converges in probability to the unique ω_0 that chooses the maximal bound for a valid consistent estimation of $\widehat{\alpha}_T$.

Proposition 4.8. *Suppose that (1) There exists a lower bound ω_{lb} such that the restrictions are respected for the interval $(\omega_{lb}, \omega_{lb})$, (2) $\omega_{max} = (-\omega_0, \omega_0)$, 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 (4.16) 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.

5 Asymptotic least squares in the frequency domain for structural VAR models

In this section, we make use of the general results of Section 4 and show how they can be applied to identify structural VAR models with frequency-based restrictions. We first consider the full identification case for any N-variate VAR model. Then we apply these results in the case of a (structural) bivariate VAR model. Finally, we discuss the case of partial identification, especially for a single structural shock.

5.1 N-variate VAR

Going back to the example of identifying neutral versus investment-related technology shocks (Section 3), one can impose a lower triangular $A(z)$ matrix in the case of a trivariate SVAR model and thus a Cholesky decomposition. In this respect, the frequency-based identifying restrictions are, $\forall \omega \in [\underline{\omega}, \overline{\omega}]$:

$$\left[\widehat{C}(e^{-i\omega})A(0)\right]_{\substack{\iota j \\ j>\iota}} = 0 \Leftrightarrow \sum_{\ell=1}^{3} c_{\iota\ell}(e^{-i\omega})a_{\ell j}(0) = 0$$

where $(\iota, j) = \{(1,2),(1,3),(2,3)\}$ and $c_{\iota\ell}$ is the (ι,ℓ) element of $C(e^{-i\omega})$, and the estimating equations resulting from the variance-covariance matrices of the reduced-form innovations and the structural shocks, $vech\,(\Sigma - A(0)A(0)')$ are given by:

$$\widehat{\sigma}_{\iota j,T} - \sum_{\ell=1}^{3} a_{\iota\ell}(0)a_{j\ell}(0) = 0$$

where $\iota = 1,\cdots,3$ and $j \leq \iota$, and $\widehat{\sigma}_{\iota j,T} = \widehat{\sigma}_{j\iota,T}$ is a consistent estimate of the (ι, j) element of Σ. Then, the C-ALS estimator of $A(0)$ is obtained by solving jointly these estimating equations. In a first-step, using the identity operator, the system is just-identified.[30] In contrast, the optimal second-step estimator is obtained from an over-identified system of estimating equations, $\underline{s}\left(a(0),\widehat{\beta}_T\right)$ (e.g., in Proposition 4.4), which correspond to the $q = \texttt{dim}(\beta)$ orthogonality conditions between the estimating equations and the projection of the latter onto the subspace spanned by $\widehat{\beta}_T$ in a functional Hilbert space.

Moving on to the general case of a N-variate SVAR model and by defining[31]

$$g(a(0),\beta,z) = \begin{bmatrix} g_1(a(0),\beta,z) \\ g_2(a(0),\beta,z) \end{bmatrix} = \begin{bmatrix} H\left(I_N \otimes C(z)\right)a(0) \\ \text{vech}\,(\Sigma - A(0)A(0)') \end{bmatrix} = \mathbf{0}, \tag{5.17}$$

the first-step C-ALS estimator with the identity operator is given by the following minimization problem:

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

Proposition 5.1 establishes the main results irrespective of the selection of the estimating equations, H, as long as the SVAR is just-identified or over-identified in the first-step.

Proposition 5.1. *Consider the vector of just- or over-identified estimating equations defined by equation (5.17). Suppose that the moments of order three of u_t are zero. Let $\beta = (vec(\Phi_p)', vec(\Sigma)')' \equiv (\Phi',\sigma')'$ 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. Then,*

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

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

  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:*

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

[30]Derivations are provided in Appendix 4.

[31]Without loss of generality, we assume that $b(z) = 0$.

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

$$\widehat{a}_T = \arg\min_a \left[\underline{s_1}(a(0),\widehat{\beta}_T)'\tilde{W}_{1T}^2\underline{s_1}(a(0),\widehat{\beta}_T) + g_2(a(0),\widehat{\beta}_T)'W_{2T}g_2(a(0),\widehat{\beta}_T)\right]$$

where $\underline{s_1}\left(a(0),\widehat{\beta}_T\right)$ *is the first set of estimating equations,* $\tilde{W}_{1T}$ *is the generalized inverse of* W_{1T}*:*

$$W_{1T} = \Omega_\Phi^{1/2}\int_{\underline{\omega}}^{\overline{\omega}} \overline{\frac{\partial g_1'}{\partial \Phi}\left(\widehat{a}_T^1,\widehat{\beta}_T,\omega\right)}\frac{\partial g_1}{\partial \Phi'}\left(\widehat{a}_T^1,\widehat{\beta}_T,\omega\right)d\omega\Omega_\Phi^{1/2}$$

and W_{2T} *is the inverse of* $2D_N^+(\widehat{\Sigma}\otimes\widehat{\Sigma})D_N^{+\prime}$.

Proof: See Appendix 2.

On the one hand, the first-step C-ALS estimator can be easily obtained from the first-order conditions and only requires to solve N^2 quadratic functions of the parameters of interest $a(0) = \mathrm{vec}(A(0))$. Interestingly, for a given ω, the two terms $(I_N \otimes C(z)')H'$ and $D_N^+(A(0)\otimes I_N)$ are those that provide the rank condition and thus the local identification of structural VARs.[32] On the other hand, taking the block diagonal structure of the optimal variance-covariance matrix when the moments of order three of u_t are zero, the second-step C-ALS estimator can also be derived in a straightforward way by solving numerically 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_1 onto the subspace spanned by the reduced-form estimates of the autoregressive parameters and the standard $N(N+1)/2$ estimating equations, g_2, that defines the mapping between the variance-covariance of the innovations and the structural shocks. At the same time, note that the objective function takes into account the optimal weighting matrix for g_2 through W_{2T}.

5.2 Bivariate VAR

As an application of Proposition 5.1, we consider a bivariate VAR model. Suppose that the second shock has no impact on the first variable over a (symmetric) interval around zero $[\underline{\omega},\overline{\omega}]$ (possibly, with $\underline{\omega} = -\overline{\omega}$). The continuum of identifying restrictions is given by

$$\widehat{c}_{11}(e^{-i\omega})a_{12}(0) + \widehat{c}_{12}(e^{-i\omega})a_{22}(0) = 0 \tag{5.18}$$

for all $\omega \in [\underline{\omega},\overline{\omega}]$, and the system of estimating equations by:

$$g(a(0),\beta,\omega) = \begin{bmatrix}\widehat{c}_{11}(e^{-i\omega})a_{12}(0) + \widehat{c}_{12}(e^{-i\omega})a_{22}(0)\\ \mathrm{vech}\left(\widehat{\Sigma}_T - A(0)A(0)'\right)\end{bmatrix} = \mathbf{0}.$$

Then one can use the results of Proposition 5.1 and derive the C-ALS estimator of $a(0) = \mathrm{vec}(A(0))$. However, for the bivariate (structural) VAR, there exists a simpler procedure, which exploits the linearity

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

of the just-identifying restriction and the existence of a nontrivial solution for any $z = \exp(-i\omega)$.[33] Indeed, the identifying restriction can be written as $a_{22}(0)\left(\widehat{c}_{11}(e^{-i\omega})\tilde{a}_{12}(0) + \widehat{c}_{12}(e^{-i\omega})\right) = 0$ for all $\omega \in [\underline{\omega}, \overline{\omega}]$ with $\tilde{a}_{12} = a_{12}(0)/a_{22}(0)$, or equivalently as $\widehat{c}_{11}(e^{-i\omega})\tilde{a}_{12} + \widehat{c}_{12}(e^{-i\omega}) = 0$ since $a_{22}(0) \neq 0$. Therefore one can proceed with a two-step estimation procedure. In a first step, a consistent estimate of $\tilde{a}_{12}$ is obtained by minimizing the restrictions (5.18) using the objective function in Proposition 4.4 or 4.5. In a second step, the estimator of $a = (a_{11}(0), a_{12}(0), a_{22}(0))'$ is obtained as the solution of the locally just-identified nonlinear system of equations:

$$\widehat{a}_T = \arg\min_{a \in \mathcal{A}} vech\left(\widehat{\Sigma}_T - \tilde{A}(0)\tilde{A}(0)'\right). \tag{5.19}$$

where $\tilde{A}(0)$ is equal to $A(0)$ but replacing $\widehat{a}_{12}(0)$ by $\widehat{\tilde{a}}_{12,T}\widehat{a}_{22}(0)$.

In this respect, the local identification of the parameters of $A(0)$ is stated in Proposition 5.2.

Proposition 5.2. *Consider the identifying restrictions that the effect of the second structural shock on the first variable in a bivariate structural VAR is zero over a frequency interval:*

$$\widehat{c}_{11}(e^{-i\omega})a_{12}(0) + \widehat{c}_{12}(e^{-i\omega})a_{22}(0) = 0 \quad \forall \omega \in [\underline{\omega}, \overline{\omega}].$$

Then the matrix $A(0)$ is locally identified (up to a sign restriction) when

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

where $\tilde{a}_{12} = a_{12}(0)/a_{22}(0)$, and $\Sigma = A(0)A(0)'$.

Using the identity operator, the first-step C-ALS estimator of $\tilde{a}_{12}(0)$ solves:

$$\begin{aligned} \widehat{\tilde{a}}^1_{12,T} &= \arg\min \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\omega}}^{\overline{\omega}} g(\tilde{a}_{12}, \widehat{\beta}_T, \omega_1)' I_{\{\omega_1 = \omega_2\}} \overline{g(\tilde{a}_{12}, \widehat{\beta}_T, \omega_2)} d\omega_1 d\omega_2 \\ &= \arg\min \int_{\underline{\omega}}^{\overline{\omega}} |g(\tilde{a}_{12}, \widehat{\beta}_T, \omega)|^2 d\omega. \end{aligned}$$

with $I_{\{.\}} = 1$ for $\omega_1 = \omega_2$ and zero otherwise. The objective function can be easily interpreted from equation (2.6). Indeed, the (1,1) element of the reduced-form spectral density matrix (up to a constant term), $|C(z)A(0)|^2$, defines the power spectrum of the first variable:

$$G^r_{11}(z) = |c_{11}(z)a_{11}(0) + c_{12}(z)a_{21}(0)|^2 + |c_{11}(z)a_{12}(0) + c_{12}(z)a_{22}(0)|^2.$$

Therefore minimizing the continuum of identifying restrictions (with the identity operator) such that the second structural shock has no effect on the first variable, i.e. $c_{11}(z)a_{12}(0) + c_{12}(z)a_{22}(0) = 0$ for any $z = e^{-i\omega}$ such that $\omega \in [\underline{\omega}; \overline{\omega}]$, is equivalent to maximize the partial spectrum of the first variable w.r.t. the first

[33] This is no longer true when $N > 2$ and there are at least two identifying restrictions in the frequency domain.

structural shock or to minimize the partial spectrum of the first variable w.r.t. the second structural shock. This dual interpretation turns out to be very convenient, especially in the case of partial identification (see Section 5.3).

Deriving the first-order condition, it is straightforward to show that:

$$\hat{\hat{a}}^{1}_{12,T} = -\frac{\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\hat{c}_{11,j}\hat{c}_{12,l}\int_{\underline{\omega}}^{\overline{\omega}}\cos((l-j)\omega)d\omega}{\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\hat{c}_{11,j}\hat{c}_{11,l}\int_{\underline{\omega}}^{\overline{\omega}}\cos((l-j)\omega)d\omega}. \tag{5.20}$$

In the bivariate case, this first-step estimator corresponds to the Min-effect/Max-effect frequency estimator proposed by Wen (2001, 2002). On the other hand, taking that the standard long-run Blanchard and Quah restriction (i.e., $\omega = 0$) yields:

$$\hat{a}_{12,T} = -\frac{\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\hat{c}_{11,j}\hat{c}_{12,l}}{\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\hat{c}_{11,j}\hat{c}_{11,l}} = -\frac{\hat{c}_{12}(1)}{\hat{c}_{11}(1)}, \tag{5.21}$$

the one-step C-ALS estimator can be seen as a generalized least squares estimator in which the weights (i.e., the sine terms) capture not only the information at the zero frequency but also its neighborhood. In the supplementary material, we show that the first-step C-ALS estimator can be written equivalently as a function of the autoregressive parameters of the reduced-form representation, and we also compare this estimator with a discretized ALS estimator.

Using the first-step estimator $\hat{\hat{a}}^{1}_{12,T}$ and Proposition 4.4, one obtains the objective function and the explicit expression of the optimal second-step estimator $\hat{\hat{a}}_{12,T}$ in the form of a generalized least squares estimator.

Proposition 5.3. *Consider the identifying restrictions that the effect of the second structural shock on the first variable in a bivariate structural VAR is zero over a frequency interval:*

$$\hat{c}_{11}(e^{-i\omega})a_{12}(0) + \hat{c}_{12}(e^{-i\omega})a_{22}(0) = 0 \quad \forall\omega \in [\underline{\omega}, \overline{\omega}].$$

Then, using the objective function defined in Proposition 4.4, the optimal C-ALS is:

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

where $\hat{s}_{11,T}$, $\hat{s}_{12,T}$ and $\widehat{W}_T$ are given by:

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

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

and

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

where $\widehat{\tilde{a}}^{1}_{12,T}$ *is a first-step estimator and* $\int_{\underline{\omega}}^{\overline{\omega}} \cos((l-j)\omega)d\omega = \frac{1}{l-j}\left[\sin((l-j)\omega)\right]_{\underline{\omega}}^{\overline{\omega}} = \frac{2}{l-j}\sin((l-j)\overline{\omega})$ *with a symmetric interval* $[-\overline{\omega}, \overline{\omega}]$ *for* $(l-j) \neq 0$ *and* $\int_{\underline{\omega}}^{\overline{\omega}} \cos((l-j)\omega)d\omega = \overline{\omega} - \underline{\omega} = 2\overline{\omega}$ *for* $l = j$.

Proof: See Appendix 2.

Finally, an estimate of $a(0)$ is obtained as the solution of equation (5.19).

5.3 Partial identification of a structural shock

As a last application, we consider the (partial) identification of a single structural shock in a VAR with more than two variables. Without loss of generality, we assume that the structural shock of interest is the first one. For instance, this corresponds to the identification of a technology shock in a multivariate VAR without requiring the identification of other shocks (see Christiano et al., 2006b; Francis and Ramey, 2009) or the identification of a news shocks (Beaudry and Portier, 2006; Barsky and Sims, 2011; Kurmann and Sims, 2019). For sake of completeness, we first provide the common sign and (long-run) exclusion restrictions and then turns to the frequency identifying restrictions. Notably we use the equivalent representation of the identifying restrictions in equation (2.6).

Following Christiano et al. (2006a), the dynamic effects of the first structural shock can be computed by identifying only $[A(0)]_{\cdot 1}$, the first column of $A(0)$, since combining a sign restriction and zero restrictions on the long-run impact does uniquely identify the vector $[A(0)]_{\cdot 1}$. In this respect, one needs to impose $N-1$ zero-restrictions:

$$\widehat{C}(1)A(0) = A(1) = \begin{bmatrix} a_{11}(1) & 0_{1\times(N-1)} \\ \tilde{A}_{21}(1) & \tilde{A}_{22}(1) \end{bmatrix}$$

where $\tilde{A}_{21}(1)$ is the first column of the long-run impact matrix $A(1)$ after dropping the first element $a_{11}(1)$ and the submatrix $\tilde{A}_{22}(1)$ contains the other columns of $A(1)$ (except the first row of those columns). Imposing that only the first structural shock has a long-run impact on the first variable yields the following specification of the long-run variance-covariance matrix:[34]

$$\widehat{C}(1)A(0)A(0)'\widehat{C}(1)' = A(1)A(1)' = \begin{bmatrix} a_{11}(1)^2 & a_{11}(1)\tilde{A}_{21}(1)' \\ \tilde{A}_{21}(1)a_{11}(1) & \tilde{A}_{21}(1)\tilde{A}_{21}(1)' + \tilde{A}_{22}(1)\tilde{A}_{22}(1)' \end{bmatrix} = \widehat{C}(1)\widehat{\Sigma}_T\widehat{C}(1)'.$$

This implies that $a_{11}(1)^2$ is the (1,1)-element of the matrix $\widehat{C}(1)\widehat{\Sigma}_T\widehat{C}(1)'$ and that $\tilde{A}_{21}(1)$ is equal to the corresponding elements of the matrix $\widehat{C}(1)\widehat{\Sigma}_T\widehat{C}(1)'$ divided by $a_{11}(1)$. Since the first column of the matrix $C(1)A(0)$, denoted $[A(1)]_{\cdot 1}$, is known, the column vector $[A(0)]_{\cdot 1}$ is uniquely identified by the relation

[34] Note that many matrices $A(0)$ are comfortable with these restrictions but the first column of each of these matrices $A(0)_1$ is the same (see Christiano et al., 2006b).

$[A(0)]_{\cdot 1} = \widehat{C}(1)^{-1}[A(1)]_{\cdot 1}$. Now consider the same restrictions but for a general operator z, one has:

$$
\begin{aligned}
|a_{11}(z)|^2 &= \left[\widehat{C}(z)\widehat{\Sigma}_T\overline{\widehat{C}(z)} =' \right]_{11} \\
\tilde{A}_{21}(z)a_{11}(z) &= \left[\widehat{C}(z)\widehat{\Sigma}_T\overline{\widehat{C}(z)}'\right]_{n1,\ n=2,\dots,N}
\end{aligned}
$$

where $\left[\widehat{C}(z)\widehat{\Sigma}_T\overline{\widehat{C}(z)}'\right]_{n1}$ is the element $(n,1)$ of the matrix $\widehat{C}(z)\widehat{\Sigma}_T\overline{\widehat{C}(z)}'$.

Proposition 5.4. *Consider the following identification constraints* $\sum_{j=1}^{N}\widehat{c}_{1j}(e^{-i\omega})a_{jn}(0) = 0$ *for* $n = 2,\dots,N$ *and* $\forall\omega \in [\underline{\omega},\overline{\omega}]$. *Let* $\widehat{\beta}_T = \left(vec\left(\widehat{\Phi}_p\right)', vech\left(\widehat{\Sigma}_T\right)'\right)'$ *denote the vector of dimension* $q = N^2 \times p + \frac{N(N+1)}{2}$ *of the reduced-form parameters estimates. The estimating equations,* $g(\alpha_0,\widehat{\beta}_T,\omega) = \mathbf{0}$ *for* $\omega \in [\underline{\omega},\overline{\omega}]$, *defined by*

$$
g(\alpha_0,\widehat{\beta}_T,\omega) = \left(g_1(\alpha_0,\widehat{\beta}_T,\omega), g_2(\alpha_0,\widehat{\beta}_T,\omega),\cdots,g_N(\alpha_0,\widehat{\beta}_T,\omega)\right)'
$$

with

$$
\begin{aligned}
g_1(\alpha_0,\widehat{\beta}_T,\omega) &= \left|[\widehat{C}(e^{-i\omega})A(0)]_{11}\right|^2 - \left[\widehat{C}(e^{-i\omega})\widehat{\Sigma}_T\overline{\widehat{C}(e^{-i\omega})}'\right]_{11} \\
g_n(\alpha_0,\widehat{\beta}_T,\omega) &= \left[\widehat{C}(e^{-i\omega})A(0)\right]_{n1}[A(0)'\overline{\widehat{C}(e^{-i\omega})}']_{11} - \left[\widehat{C}(e^{-i\omega})\widehat{\Sigma}_T\overline{\widehat{C}(e^{-i\omega})}'\right]_{n1}
\end{aligned}
$$

for $n = 2,\dots,N$, *uniquely identify the first column of the matrix* $\alpha_0 = [A(0)]_{\cdot 1}$ *up to a sign restriction.*

Note that the moment conditions can be written, for $n = 1,\cdots,N$:

$$
g_n(\alpha_0,\widehat{\beta}_T,\omega) = \sum_{r=1}^{N}\sum_{s=1}^{N}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty} c_{1s,j}c_{nr,l}\left(a_{r1}(0)a_{s1}(0) - \widehat{\sigma}_{sr,T}\right)cos((j-l)\omega)
$$

where $\widehat{\sigma}_{sr,T}$ is a consistent estimate of the (s,r) element of Σ. As in Section 5.1, a first-step consistent C-ALS estimator of $\alpha_0 = [A(0)]_{\cdot 1}$ solves the following minimization problem (using the identity operator as a weighting matrix):

$$
\widehat{\alpha}_T = \arg\min_{a}\int_{\underline{\omega}}^{\overline{\omega}} g(\alpha,\widehat{\beta}_T,\omega)'\overline{g(\alpha,\widehat{\beta}_T,\omega)}d\omega.
$$

Then the second-step C-ALS results from the minimization of the simplified (regularized) objective function in the light of Proposition 4.3. or 4.4. Appendix 2 provides the relevant analytical first-order partial derivatives. Finally, as shown in Section 5.1, note that an equivalent solution can be obtained using a Cholesky decomposition.

6 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 is a bivariate VAR model (with different parameter configurations) 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:

$$\Delta X_{1,t} = \rho_{11,1}\Delta X_{1,t-1} + (\rho_{12,1}+\delta)X_{2,t-1} - \rho_{12,1}X_{2,t-2} + \epsilon_{1,t} \quad (6.22)$$

$$X_{2,t} = \rho_{21,1}\Delta X_{1,t-1} + \rho_{22,1}X_{2,t-1} + \rho_{22,2}X_{2,t-2} + b_{21}\epsilon_{1,t} + \epsilon_{2,t} \quad (6.23)$$

where the vector $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ represents some structural shocks, with $\epsilon_t \sim N(0, I_2)$. 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, e.g., the identification of a technology shock with some measures of (labor or total) productivity and hours worked (see Section 7). It is worth emphasizing that the VAR(1) specification is the data generating process of Gospodinov et al. (2013) and Chevillon et al. (2020), whereas the VAR(2) corresponds to the one of Gospodinov (2010) and Gospodinov et al. (2011).

Using equations (6.22) and (6.22), 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.[35] Our method denoted `C-ALS`, which is based on the two-step C-ALS procedure (Section 4.2), is compared with three approaches. The first one, denoted `LR`, is a standard long-run identification scheme à la Blanchard-Quah, i.e. we only impose the identification constraint at $\omega = 0$.[36] The second alternative is the first-step C-ALS estimator defined in equation (5.20) when the kernel operator is the identity operator. The last alternative is the max-share procedure of DiCeccio and Owyang (2010) and Francis et al. (2014), denoted `MS`, which identifies the structural shock of interest that maximizes the share of the forecast-error variance of a given variable in the frequency domain.[37]

With the exception of the `LR` method, we consider four fixed symmetric frequency intervals $\omega_n = (-\frac{2\pi}{n}, \frac{2\pi}{n})$ for $n = 30, 60, 90$ and 120 quarters. Results are then assessed along three dimensions. First, we compute the initial impact of each structural shock on each variable and determine the corresponding mean absolute bias and root mean squared errors (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.[38] More specifically, the

[35] Several robustness exercises, which are available upon request, have been experimented to control for the lag order misspecification. All in all, our results remain unchanged and our estimator performs better than the competing estimators.

[36] We also implement the methodology of Christiano et al. (2006b), i.e. 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.

[37] The max-share approach is developped in Faust (1998) and Uhlig (2003, 2004), and subsequently used by Barsky and Sims (2011), Francis et al. (2014), Kurmann and Sims (2019), Angeletos et al. (2020).

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

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 (6.22) and (6.23), and $\texttt{irf}_h(\texttt{svar}) = (1/N)\sum_{j=1}^{N} irf_h(\texttt{svar})^j$ the average impulse response function over the N simulation experiments. Say differently, the cumulative mean absolute bias is a measure of the area between the impulse response function up to a certain 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 estimated impulse response functions provided by the competing methods.

In all simulation experiments, as n increases and the length of the symmetric frequency interval reduces, the matrix W might not be of full rank so that we make use of a generalized inverse through a regularization method and define the objective function as in equation (4.5). Indeed one main difficulty of solving the Moore-Penrose pseudo-inverse of W without regularization is that the matrix W has tiny positive singular values and this might lead to severe numerical instability due to round-off errors and unstable behavior of the solution; these problems being more and more accurate as the frequency interval shrinks toward zero. To circumvent this issue, we determine the rank of the matrix, say k, and take as regularization parameter the kth (ordered) eigenvalue.[39]

In the sequel, we consider two assumptions for each parameter configuration of our Monte-Carlo simulations. First, we assume that only the first structural shock has a permanent effect on the first variable, i.e. the identifying restriction is correctly specified (null hypothesis). Second, we proceed with a misspecified exclusion restriction (alternative hypothesis) in the sense that both shocks have a permanent effect whereas we do only impose that the first structural shock matters permanently for the first variable.

In the first set of experiments, we 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.[40] 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. Therefore 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 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 at different frequencies ω_n. Three points are worth commenting. First, an eye inspection of 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 comparing with the `MS` approach. Second, the second-step `C-ALS` approach outperforms other methods for both statistical criteria. Notably it turns out that the mean absolute bias differences between the second-step `C-ALS` estimator and other alternatives are substantial irrespective of the frequency. Moreover, the discrepancy between the first-step C-ALS and the second-step C-ALS illustrates that there is an effective gain to

[39]We also consider a truncated singular value decomposition and a pseudo-inverse method: our resuts are qualitatively the same.

[40]These calibrated values are those of Gospodinov et al. (2013).

exploit the relevant information of the reduced-form VAR estimation. All in all, 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, as stated in Section 5.2, the objective function of the first-step estimator is equivalent to minimize (respectively, to maximize) the contribution of the second (respectively, the first) structural shock to the partial spectrum of the first variable. Under the null hypothesis of a well-specified identifying restriction, it amounts of 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 twelve periods for $n = 30, 60, 90, 120$, Figure 2 provides also very supportive evidence for the `C-ALS` approach. Indeed 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` methodology 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 $\omega_{60} = (-\frac{2\pi}{60}, \frac{2\pi}{60})$.[41] 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 included in level in the 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. On the other hand, 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 an horizon $H = 20$. Therefore, there is a huge gain of efficiency by computing the optimal weighting matrix relative to the first-step C-ALS, and the max-share approach. Finally, we implement the overidentification test (Proposition 4.6) and especially its regularized version in equation (4.14).[42] 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 the alternative hypothesis, i.e. the exclusion restriction is misspecified. As reported in Table 1, when $\delta = 0.05$ or 0.1 and n augments and thus the length of the frequency band reduces, the proportion of the variance explained by the second structural shock for the first variable increases. In this respect, Figures 4 and 5 provide support that our methodology clearly outperforms other methods in terms of (cumulative) mean absolute bias and (cumula-

[41] Results are qualitatively the same for ω_{30}, ω_{90} and ω_{120}.

[42] As explained before, 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). However, our worst case leads to reject 4 simulations over 10,000 for the frequency interval ω_{120}.

tive) RMSE. Again here, 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 estimation 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 relative good performance of our proposed estimator when the estimating equations are not satisfied can be explained by the fact that the minimization solution yields a pseudo-true value, which is the closest to the true value, and that the contribution of the second structural shock is minimized. Finally, our results seem to be consistent with the restricted IRF estimator proposed by Gospodinov (2010) when the second variable is a strongly dependent process that can be parameterized as local to unity and thus leads to a weak identification problem in the instrumental variable-based framework of structural VAR identification. On the other hand, panels a,b, and c of Table 1 shows that the overidentification test performs well. As δ augments, 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 $\delta = 0.1$.

We now turn to the second set of Monte Carlo simulations in which we consider a VAR(2) specification with $(\rho_{11,1}, \rho_{12,1}, \rho_{21,1}, \rho_{22,1}, \rho_{22,2}, b_{21}, \delta) = (0, -0.08, 0.2, \rho + 0.55, -0.55\rho, 0.2, 0)$ where $\rho = 0.9$, 0.95, or 0.98. Looking at Figures 7, 8, and 9, the results are qualitatively similar to those of the VAR(1) under the null hypothesis. Interestingly, when $\delta = 0$, the contribution of the second structural shock to the first variable is not zero but rather close to zero and decreases when the length of the frequency band increases, and, as such, this case can be interpreted as a local alternative to the null hypothesis of a well-specified identifying restriction.[43] Taking that this local alternative hypothesis is often considered as a very plausible data generating process in SVARs applications, the finite sample properties of the second-step C-ALS estimator are again remarkable and appealing. At the same time, as to be expected, the J-stat has less power but can still provide useful information in the case of a local alternative.

When $\delta = 0.04$, three points can be stressed. First, the `C-ALS` estimator still dominates the first-step, `MS` and `LR` approaches and displays a small (cumulative) mean absolute bias for interval greater than ω_{30} meanwhile the max-share estimator is only slightly improving relative to the standard `LR` approach. Notably, the mean absolute bias of the two-step C-ALS estimator is close to zero for the widest interval ω_{30} while it increases with decreasing intervals. This comes from the fact that the variance contribution of the second shock to the first variable augments when the frequency interval becomes smaller and smaller and that the (regularized) minimization problem of the `C-ALS` estimator seeks to find the optimal linear combination of the reduced-form shocks such that the contribution of the second structural shock to the first variable is minimized. Second, the bias reduction of the second-step `C-ALS` estimator is achieved with a lower (cumulative)

[43] In contrast to other cases, the partial spectral density of the first variable with respect to the second structural shock is U-shaped in a neighborhood of $\omega = 0$.

RMSE relative to other estimators. Third, all of these results are robust with respect to the chosen horizon H.

Finally, Panel D of Table 1 show that the J-test has good power properties for this case. Indeed, in the former, the probability of rejecting the null hypothesis in the presence of misspecification is close to 95% (resp., 97%)

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 also 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.

7 Applications

This section provides two applications of our methodology. First, we discuss the effect of a technology shock on hours worked within the framework of bivariate structural VAR model (Section 5.2). Second, we proceed with a (partial) identification of a news shocks (Section 5.3) using a quadrivariate structural VAR model.

7.1 The hours-productivity debate using bivariate SVAR models

Continuing from Sections 2 and 3.1, one key issue of the technology-hours debate is the assumed data generating process for the hours worked (per capita) measures. On the one hand, using a first difference specification of the hours measure, structural VAR models predict a decline of hours in response to a positive technological shock (e.g., Galí, 1999, or Francis and Ramey, 2005), opposite of that implied by Real Business Cycles models.[44] On the other hand, entered in level, hours rise 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 of these conflicting results is that the standard measure of hours per capita and productivity have significant low-frequency movements and these movements can conduct 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.[45] Consequently, 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 problem, Francis and Ramey (2009) propose using new measures of hours worked per capita

[44]While standard unit root tests can not reject the presence of an 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).

[45]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).

and labor productivity that are more comfortable with the imposed long-run restriction(s). Taking the adjusted series, it turns out that hours worked now respond negatively in the short run, and then become slightly positive after a year for a structural VAR model in which the adjusted hours worked per capita series is included in level. In this respect, a more complete test of their results asks: (1) Is there any 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 defined in Section 5.2 (Proposition 5.3), 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 enters 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, standard error bands are 95 percent confidence bands based on bootstrap standard errors with 1,000 replications.

As reported in Figure 13, using the standard `LR` restriction, Panel A shows 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 of this apparent discrepancy is that the identifying assumption, namely the technological shock does explain alone 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.

On the other hand, as reported in Figure 13, 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 has now a p-value of 0.6512. Say differently, 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 long-run effect on labor productivity. Finally, we 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 $\hat{\omega}_T = 120$.[46]

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.

7.2 Total factor productivity and news shocks

Continuing from Section 3.3, recent empirical literature delivered controversial results concerning the role of anticipated technology—news—shocks in business cycle fluctuations using structural VAR approaches. Starting with Beaudry and Portier (2004, 2006), news shocks are related to innovations driving the long-run variations of the total factor productivity.[47] Using short-run and long-run restrictions, Beaudry and Portier (2006) conclude that news shocks about future productivity are one of the main drivers of business cycles and there is a positive (contemporaneous) impact of the news shock on hours worked.

These results have been challenged in several directions. Especially, using a cumulative max-share approach, Barsky and Sims (2011) provide evidence that there are a negative (contemporaneous) impact of the news shocks on hours worked and a negative conditional correlation between consumption growth and hours, which leans against news-driven business cycles and thus is more in line with the implications of the standard neoclassical framework. Kurman and Sims (2019) further relax the assumption that news shocks are orthogonal to TFP at the impact to avoid cyclical mismeasurement of the long-run productivity.[48] Using a max-share approach, their results provide weak support for the news shock as the main driver of business fluctuations.

Taking the debate regarding the importance of news shocks as a main driver of business fluctuations and the mixed evidence regarding the response of hours worked following a news shock, we reexamine the empirical evidence in light of our results in Sections 4 and 5. Our starting point is that the max-share approach advocated by Barsky and Sims (2001) or Kurmann and Sims (2019) still relies on both short-run and business cycle fluctuations since the forecast error variance decomposition is based on the summation of the first forty quarters (Barksy and Sims, 2011) or the first eighty quarters (Kurmann and Sims, 2019) and thus

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

[47] For an overview of the main insights and challenges of news shocks, see Beaudry and Portier (2014).

[48] Kurman and Sims (2019) also discuss the importance of the revisions of the utilization-adjusted series on total factor productivity (Fernald, 2014). Using the 2016-vintage instead of the 2007-vintage leads to an opposite conclusion: hours worked can respond positively to a news shocks at the impact (with confidence bands encompassing negative values) using the approach of Barsky and Sims (2011).

might partially disentangle the long-run productivity shock from lower persistent shocks. In contrast, our frequency-based identification strategy has the advantage to focus on medium and long-run frequencies of TFP and thus allows to exclude short-run and business cycles fluctuations. In so doing, we suppose that the news shock is the only shock driving the productivity growth for frequency intervals corresponding to medium and long-run fluctuations of TFP.

To compare our results, we apply four identification strategies to the same four-variable VAR investigated by Kurmann and Sims (2019). Especially, we focus on the standard long-run identification strategy of Blanchard and Quah (1989) denoted `BQ`, the cumulative forecast error variance decomposition identification of Barsky and Sims (2001) denoted `BS`, the max-share identification of Kurmann and Sims (2019) denoted `KS`, and our C-ALS identification. The structural VAR comprises the utilization-adjusted TFP series, real (personal) consumption expenditures per capita, total hours worked per capita in the non-farm business sector and inflation as measured by the growth rate of the GDP price deflator. The (quarterly) sample period is fixed at 1960:q1 to 2018:q4. An important caution concerns the transformation of the initial series. On the one hand, following the methodology of Fernald (2014) and the prerequisite of the (weakly) stationary assumption for the spectral density matrix, the TFP series is included in first-difference. In contrast, Kurmann and Sims (2019) and Barsky and Sims (2011) consider the TFP series in level by taking the cumulative sum of the series constructed by Fernald (2014). By construction, the cumulative sum leads to the presence of a unit root in the (level) series.[49] Furthermore, the consumption series is also first-differenced as one can not reject the presence of a unit root at conventional significance levels whereas the hours worked and inflation series are left unchanged in levels.[50] On the other hand, all series are in level in the case of the BS and KS identification strategies.

Figure 14 displays the impulse response functions for the aforementioned macro series using the `BQ` long-run restrictions (dash-dotted line), `BS` approach (dotted line), max-share of `KS` (dashed line) and interval frequency restrictions (solid line). In the case of the `C-ALS` approach, using both the overidentification test (Section 4.4) and the data-driven selection procedure (Section 4.5) provides empirical support for $\omega \in [-\frac{\pi}{120}, \frac{\pi}{120}]$.[51] Interestingly, the `KS` identification strategy leads to a negative contemporaneous effect (-.0546) of the news shock on hours worked, which is consistent with their findings, whereas the `BQ`, `BS` and `C-ALS` identification procedures leads to a positive response of .0320, 0.0752 and 0.0660 respectively.[52] In addition, all impulse response functions display the same hump shaped dynamics.

Regarding now other macro series, the impulse response functions of consumption and inflation are rather close. In the case of inflation, both the `KS`- and `BS`-based impulse response functions display a more long-

[49] See Phillips (1998) for the properties of impulse response functions in the case of VAR including unit or near unit roots.

[50] We also investigate a specification with the consumption series in level. The results are close to the ones with the specification with consumption in difference and are available upon request.

[51] The corresponding p-value of the J-stat equals .245.

[52] All in all, using more recent data we find that the contemporaneous effect is less pronounced than in the original paper of Kurman and Sims (2019), which covers the period 1961q1-2007q3.

lasting (persistent) negative effect than the one identified with the `BQ` or `C-ALS` procedure. The result goes the other way for the consumption series. Furthermore, whether the consumption series is in first-difference or in level does not qualitatively alter the shape of the impulse response function, at least for the first forty quarters. Looking at the TFP series, the `BS` approach, as to be expected, leads to a zero contemporaneous effect and then can be characterized by a slow diffusion process. In contrast, by dropping out the short-run identifying restriction, the most noticeable feature of the `KS` approach relative to the `BS` strategy is the presence of a positive inverted hump shaped effect over the first five quarters—the two corresponding impulse response functions being almost the same for horizons greater than 10 quarters. Finally, the impulse response based on the standard long-run and the frequencies restrictions are almost the same with a statistically significant positive higher impact relative to other methods. As a first conclusion, taking the impulse response function of the hours worked and the plausible responses of other macro variables, our empirical evidence suggests that news shocks can not be ruled out as a main driver of business fluctuations. At least, the identification scheme of news shocks and its implications as well as the specification of the variables (in level *versus* first-difference) deserve more attention.

As a robustness experiment, Figure 15 reports the impulse response functions of the hours worked to a news shocks for different frequency intervals. The impact response is negative when one considers frequency intervals on which the restrictions that only one shock drives the fluctuations of TFP is rejected or close to be rejected at a level of 10%. Indeed, for the interval including business cycle frequencies $\omega \in [-\frac{\pi}{3}, \frac{\pi}{3}]$, the p-value of the J-stat is .003 whereas, for $\omega \in [-\frac{\pi}{20}, \frac{\pi}{20}]$ the p-value is .112. This suggests that one possible explanation regarding the negative contemporaneous effect of the news shock on hours using the `KS` identification strategy is related to the selection of the frequency band or the horizons of the max share approach, and thus the fact that the news shock can confound short-run/business fluctuations with long-run fluctuations of the TFP. As a second conclusion, our empirical evidence suggests that the determination of the frequency band or the selection of the horizons of the forecast variance error decomposition must be further investigated.

8 Conclusion

In this paper, we propose a new identification scheme and the corresponding estimation method using frequency interval restrictions for structural VAR models. Using the methodology of Carrasco and Florens (2000) and Carrasco and al. (2004), we derive a continuum asymptotic least squares estimator, the C-ALS estimator, that allows to obtain reliable estimates of the dynamic responses of macroeconomic variables to structural shocks and also to assess formally 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 first application regarding the hours-productivity debate provides some new insights and, especially, the relevance of the argument of Francis and Ramey (2009) whereas the second application suggest that the relevance of news shocks as a main business-cycle driver can not be ruled out when selecting adequately the frequency band.

From an empirical point of view, our methodology and the associated testing procedure (overidentification, interval selection) can be used to reassess several debates, e.g. the identification and reliability of news' shocks (Beaudry and Portier, 2006; Barsky and Sims, 2011; Beaudry et al., 2013), 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). On the other hand, the derivation of optimal rules for the choice of the regularization parameter for testing procedures, the extension to SVAR models with integrated and cointegrated variables (Lütkepohl, 2007; Lütkepohl and Velinov, 2014), the existence of nonfundamental representations (Gourieroux and Monfort, 2020) and the recoverability condition (Chahrour and Jurado, 2021) deserve, among others, some future work.

References

[1] Amisano, G., and C. Giannini (1997), Topics in Structural VAR Econometrics, Springer Verlag, second edition.

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

[3] Arellano, M, Hansen, L.P., and E. Sentana (2012), "Underidentification?", *Journal of Econometrics*, vol. 170(2), 256-280.

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

[5] Basu, S., Candian, G., Chahrour, R. and R. Valchev (2021), "Risky Business Cycles", Working paper.

[6] Beaudry, P., and B. Lucke (2010), "Letting Different Views about Business Cycles Compete", in *NBER Macroeconomics Annual 2009,* Volume 24, NBER Chapters, pp. 413-455. National Bureau of Economic Research, Inc.

[7] Beaudry, P., and F. Portier (2004), "An Exploration into Pigou's Theory of Cycles", *Journal of Monetary Economics*, vol. 51(6), 11831216.

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

[9] Beaudry, P., and F. Portier (2014), "News-driven business cycles: Insights and challenges", *Journal of Economic Literature*, vol. 52(4), 993-1074.

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

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

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

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

[14] Campbell, J.Y., and R.J. Shiller (1987), "Cointegration and Tests of Present Value Models", *Journal of Political Economy*, vol. 95, 1062-1088.

[15] Campbell, J.Y., and R.J. Shiller (1988), "Stock Prices, Earnings and Expected Dividends", *Journal of Finance*, vol. 43, 661-676.

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

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

[18] Carrasco, M., and J.P. Florens (2000), "Generalization of GMM to a continuum of moment conditions", *Econometric Theory*, vol. 16, 797-834.

[19] Carrasco, M., and R. Kochoni (2017), "Efficient Estimation Using the Characteristic Function", *Econometric Theory*, vol. 33(2), 479-526.

[20] Carrasco, M., and R. Kochoni (2019), "The Continuum-GMM Estimation: Theory and Application" in *International Financial Markets*, vol. 1.

[21] Carrasco, M., Florens, J.P. and E. Renault (2007), 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.

[22] Chari, V., Kehoe, P., and E. McGrattan (2005), "Critique of Structural VARs Using Real Business Cycle Theory", Federal Reserve Bank of Minneapolis, Working Paper, No. 631.

[23] Chahrour, R. and K. Jurado (2018), "News or Noise? The Missing Link", American Economic Review, vol. 108, 1702-1736.

[24] Chahrour, R. and K. Jurado (2021), "Recoverability and Expectations-Driven Fluctuations", *Review of Economic Studies*, forthcoming.

[25] Chen, K. and E. Wemy (2015), "Investment-specific Technological Changes: The source of Long-run TFP Fluctuations", *European Economic Review*, vol. 80, 230-252.

[26] Chen, X. and H. White (1998), "Central Limit and Functional Central Limit Theorems for Hilbert Space-valued Dependent Processes", *Econometric Theory*, vol. 14, 260-284.

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

[28] Christiano, L.J., and R.J. Vigfusson (2003), "Maximum Likelihood in the Frequency Domain: The Importance of Time-to-Plan", *Journal of Monetary Economics*, vol. 50(4), 789-815.

[29] Christiano, L.J., Eichenbaum, M. and C.L. Evans (2005),"Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy", *Journal of Political Economy*, vol. 113 (1), 1-45

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

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

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

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

[34] Diebold, F.X., Ohanian, L.E. and J. Berkowitz (1998),, "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data", *Review of Economic Studies*, vol. 65(3), 433-451.

[35] Dieppe, A., Neville, F. and G. Kindberg-Hanlon (2019), "New Approaches to the Identification of Low-Frequency Drivers: An Application to Technology Shocks", Policy Research Working Paper 9047, World Bank group.

[36] Engle, R.F. and S. Hylleberg (1996), "Common Seasonal Features: Global Unemployment", *Oxford Bulletin of Economics and Statistics*, vol. 58, 615-630.

[37] Engle, R.F. and S. Kozicki (1993), "Testing for Common Features", *Journal of Business & Economic Statistics*, vol. 11(4), 369-380.

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

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

[40] Faust, J. (1999), "Conventional Confidence Intervals for Points on Spectrum Have Confidence Level Zero", *Econometrica*, vol. 67, 629-637.

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

[42] Fernald, J. (2014), "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity", Federal Reserve Bank of San Francisco Working Paper Series 2012-19.

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

[44] Francis, N., Owyang, M.T., Roush, J.E., and R. DiCecio (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.

[45] Francis, N., and V. Ramey (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.

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

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

[48] Gospodinov, N. (2010), "Inference in Nearly Nonstationary SVAR Models with Long-Run Identifying Restrictions", *Journal of Business & Economic Statistics*, vol. 28(1), 1-12.

[49] Gospodinov, N., Maynard, A. and E. Pesavento (2011), "Sensitivity of Impulse Responses to Small Low-Frequency Comovements: Reconciling the Evidence on the Effects of Technology Shocks", *Journal of Business & Economic Statistics*, vol. 29(4), 455-467.

[50] Gospodinov, N., Herrera, A.M. and E. Pesavento (2012), "Unit Roots, Cointegration and Pretesting in VAR Models", *Advances in Econometrics*, vol. 32, 81115

[51] Gourieroux, C., and A. Monfort (1995), Statistics and Econometric Models: Volume 1, Cambridge University Press.

[52] Gourieroux, C., Monfort, A. and J.P. Renne (2020), "Identification and Estimation in Non-Fundamental Structural VARMA Models", Review of Economic Studies, vol. 87, 19151953.

[53] Gourieroux, C., Monfort, A., and A. Trognon (1985), "Moindres Carrés Asymptotiques", *Annales de l'INSEE*, vol. 58, 91-121.

[54] Gourieroux, C. and I. Peaucelle (1992), "Séries codépendantes", *Actualité Economique*, vol. 68(1-2), 283-304.

[55] Guay, A. (2021), "Identification of Structural Vector Autoregressions Through Higher Unconditional Moments?", *Journal of Econometrics*, *forthcoming.*

[56] Guerron-Quintana, P., Inoue, A. and L. Killian (2017), "Impulse Response Matching Estimators for DSGE Models", *Journal of Econometrics*, vol.196(1), 144-155.

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

[58] Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, vol. 50, 1029-1054.

[59] Hendry, D.F. (1999), "A Theory of Co-Breaking", in *Forecasting Non-Stationary Economic Time Series*, eds. Hendry D.F. and P. Clements, Camridge, MIT Press.

[60] Hecq, A., Palm, F.C. and J-P. Urbain (2006), "Common Cyclical Features Analysis in VAR Models with Cointegration", *Journal of Econometrics*, vol. 132, 117-141.

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

[62] Inoue, A. and L. Killian (2016), "Joint Confidence Sets for Structural Impulse Responses", *Journal of Econometrics*, vol. 192(2), 421-432.

[63] 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.

[64] 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.

[65] Kilian,L. and H. Lütkepohl (2017), Structural Vector Autoregressive Analysis, Cambridge University Press.

[66] King, R., Plosser, C., Stock, J. and M. Watson (1991), "Stochastic Trends and Economic Fluctuations", *American Economic Review*, vol. 81(4), 819-40.

[67] Kress, R. (1999), Linear Integral Equations, Springer-Verlag, New-York.

[68] Kurmann, A. and E. Sims (2019), "Revisions in Utilization-Adjusted TFP and Robust Identification of News Shocks", *The Review of Economics and Statistics*, forthcoming.

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

[70] Lütkepohl, H. and A. Velino (2014), "Structural Vector Autoregressions: Checking Identifying Long-run Restrictions via Heteroskedasticity, Working Paper No. 1356, DIW Berlin.

[71] Mertens, E. (2012), Are Spectral Estimators Useful for Long-run Restrictions in SVARs?", *Journal of Economic Dynamics and Control*, vol. 36(12), 1831-1844.

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

[73] Newey, W. K. and D. McFadden, (1994). "Large sample estimation and hypothesis testing", Handbook of Econometrics, in: R. Engle & D. McFadden (ed.) volume 4, chapter 36, Elsevier.

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

[75] 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,* **70,** 1035-1065.

[76] Ramey, V.A. (2016), Macroeconomic Shocks and Their Propagation, edited in *Handbook of Macroeconomics.*

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

[78] Rotemberg, J.J., and M. Woodford (1999), "The Cyclical Behavior of Prices and Costs", NBER Working Papers 6909, National Bureau of Economic Research.

[79] Rubio-Ramirez, J. F., Waggoner, D. F. and T. Zha (2010), "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference", *Review of Economic Studies*, vol. 77, 665696.

[80] Shorack, G.R. (2000), Probability for Statisticians, New-York: Springer Verlag.

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

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

[83] Sims, C. (1986), "Are Forecasting Models Usable for Policy Analysis?", *Federal Reserve Bank of Minneapolis Quartely Review*, vol. 10(1), 2-16.

[84] Uhlig, H. (2004), "Do Technology Shocks Lead to a Fall in Total Hours Worked?", *Journal of the European Economic Association*, vol. 2(2-3), 361-371.

[85] Uhlig, H. (2003), "What Moves Real GNP?", Mimeo, Humbold University Berlin.

[86] Urga, G. (2007), "Common Features in Economics and Finance: An Overview of Recent Developments", *Journal of Business & Economic Statistics*, vol. 25(1), 2-11.

[87] Vahid, F. and R.F. Engle (1993), "Common Trends and Common Cycles", *Journal of Applied Econometrics*, vol. 8, 341-360.

[88] Vuong, Q.H. (1989), "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses", *Econometrica*, vol. 57(2), 307-333.

[89] Wen, Y. (2001), "A Generalized Method of Impulse Identification",*Economic Letters*, vol. 73, 367-374.

[90] Wen, Y. (2002), "The Business Cycle Effects of Christmas", *Journal of Monetary Economics*, vol. 49, 1289-1314.

[91] Wahba, G., (1992), Spline Models for Observational Data, SIAM.

[92] White, H. (2000), Asymptotic Theory for Econometricians, New York Academic Press.

Appendix 1: Assumptions

Assumption A.1 The stochastic process Z_t is a $N \times 1$-vector of random variables. Z_t is second-order stationary such that the Wold's Decomposition Theorem holds. The true unknown probability distribution of the $\{Z_t\}_{t=1}^T$ is denoted P_0 which belongs to a family of probability distribution.

Assumption A.2 It exists a sequence of estimators $\widehat{\beta}_T$ such that

$$\widehat{\beta}_T = \arg\min_{\beta \in \mathcal{B}} Q_T(Z_T, \beta),$$

with $\beta_0 = P_0 \lim_{T\to\infty} \widehat{\beta}_T$ is the true unknown value of the instrumental parameters and β_0 is an interior point of $\mathcal{B}$ a compact subset of $\mathbb{R}^q$. Under standard regularity conditions, $\sqrt{T}\left(\widehat{\beta}_T - \beta_0\right) \xrightarrow{d} \mathcal{N}(0, \Omega)$ where $\Omega = P_0 \lim_{T\to\infty} Var\left(\sqrt{T}\widehat{\beta}_T\right)$.

Assumption A.3 φ is the p.d.f. of a distribution that is absolutely continuous with respect to Lebesgue measure on the interval $\mathcal{I} = [-\pi, \pi]$ and admits all its moments. $\varphi(\omega) > 0$ for all $\omega \in \mathcal{I}$. $L^2(\mathcal{I}, \varphi) \equiv L^2(\varphi)$ is the Hilbert space of complex-valued functions that are square integrable with respect to φ:

$$L^2(\varphi) = \{f : \mathcal{I} \to C | \int_{\mathcal{I}} |f(\omega)|^2 \varphi(\omega) d\omega < \infty\}.$$

with $\langle ., . \rangle$ and $\|.\|$ the inner product and the norm on $L^2(\varphi)$. The inner product is

$$\langle f, h\rangle = \int f(\omega)\overline{h(\omega)}\varphi(\omega)d\omega$$

where $\overline{h(\omega)}$ denotes the complex conjugate of $h(\omega)$.

Assumption A.4 Let $g(\alpha, b, \omega)$ be a vector of measurable functions from $\mathbb{R}^d \times \mathcal{A} \times \mathcal{B}$ into $H = (L^2(\mathcal{I}, \varphi))^J$, a Hilbert space with the inner product $< .,. >$ and the norm $\| \cdot \|$. An element of H is a vector $f(\omega) = (f_1(\omega), f_2(\omega), \ldots, f_J(\omega))'$ of square integrable functions $f_i(\omega)$ for $i = 1, \ldots, J$. The inner product is $\langle f, h\rangle = \int f(\omega)'\overline{h(\omega)}\varphi(\omega)d\omega$ with the norm $\|f\|_M = \left[\int f(\omega)'M\overline{f(\omega)}\varphi(\omega)d\omega\right]^{1/2} = \left[\sum_{l,k=1,\ldots,J} m_{lk} \int f_l(\omega)'\overline{f_k(\omega)}\varphi(\omega)d\omega\right]^{1/2}$ where M is a real-value positive definite symmetric $J \times J$ matrix with element m_{lk}. Finally, for matrices of vector of functions $F = (F_1, \cdots, F_p)'$ and $G = (G_1, \cdots, G_q)'$ with elements of H^p and H^q respectively, we denote $\langle F, G\rangle$ is the $p \times q$ matrix with (i, j) element defined by $\int F_i(\omega)'\overline{G_j(\omega)}\varphi(\omega)d\omega$.[53]

Assumption A.5 The J-vector of functions

$$g(\alpha_0, \beta_0, \omega) = 0$$

$\forall \omega \in \mathcal{I}$, φ— almost everywhere, has a unique solution α_0 which is an interior point of $\mathcal{A}$ a compact set and α_0 and β_0 denotes the unknown value under P_0.

Assumption A.6 (i) $g(\alpha, \beta, \omega)$ is continuously differentiable with respect to α and β and $g(\alpha, \beta, \omega) \in (L_\infty(\varphi \otimes P_0))^J$ where $L_\infty(\varphi \otimes P_0)$ is the set of measurable bounded functions of (ω, Z_t).
(ii) $\sup_{\alpha\in\mathcal{A}} \|g(\alpha, \beta, \omega) - g(\alpha_0, \beta, \omega)\| = O_p\left(\frac{1}{\sqrt{T}}\right)$ for all $\beta \in \mathcal{B}$ and $\omega \in \mathcal{I}$.
(iii) $\sup_{\alpha\in\mathcal{A}_0} \|\partial g(\alpha, \beta, \omega)/\partial\alpha' - \partial g(\alpha_0, \beta, \omega)/\partial\alpha'\| = O_p\left(\frac{1}{\sqrt{T}}\right)$ for all $\beta \in \mathcal{B}$ and $\omega \in \mathcal{I}$ where $\mathcal{A}_0$ is some neighborhood about α_0.
(iv) $\sup_{\beta\in\mathcal{B}_0} \|\partial g(\alpha, \beta, \omega)/\partial\beta' - \partial g(\alpha, \beta_0, \omega)/\partial\beta'\| = O_p\left(\frac{1}{\sqrt{T}}\right)$ for all $\alpha \in \mathcal{A}$ and $\omega \in \mathcal{I}$ where $\mathcal{B}_0$ is some neighborhood about β_0.

Assumption A.7 Let S be a nonrandom bounded linear operator defined on $\mathcal{D}(S) \subset H$ valued in H. The operator S does not depend on α but may depend on α_0 and $g(\alpha, \beta, \omega) \in \mathcal{D}(S), \forall a$ and $\forall \beta$ under P_0.

[53] This notation differs from the definition of Frobenius inner product which is $\langle F, G\rangle_{\mathbf{F}} = trG^*F$.

Assumption A.8 Let $N(S)$ denote the null space of S, $N(S) = \{f \in H) | Sf = 0\}$. Suppose that $g(\alpha_0, b_0, \omega) \in N(S)$ implies $g(\alpha_0, b_0, \omega) = 0$.

Assumption A.9 Let S_T be a sequence of bounded linear operators converging in probability to S defined on $\mathcal{D}(S_T) \subset H \to H$. Suppose that $g(\alpha, \widehat{\beta}_T, \omega) \in \mathcal{D}(S_T)$, $\forall \alpha \in \mathcal{A}$ and that $\|S_T g(\alpha, \widehat{\beta}_T, \omega)\|$ is a continuous function of α. Moreover, $\partial g(\alpha, \beta, \omega)/\partial \alpha \in \mathcal{D}(S)$ for all $\alpha \in \mathcal{A}$, and $g(\alpha, \beta, \omega)/\partial b \in \mathcal{D}(S)$ for all $\beta \in \mathcal{B}_0$ under P_0.

Assumption A.10

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

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

$$Kf(\omega_1) = \int E^{P_0} k(\omega_1, \omega_2) f(\omega_2)\varphi(\omega_2) d\omega$$

for all f in H where under P_0

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

with $\int\int k(\omega_1, \omega_2)\varphi(\omega_1)\varphi(\omega_2) d\omega_1 d\omega_2 < \infty$. K is a compact Hilbert-Schmidt operator and K is self-adjoint ($K = K^*$). Here $\xi \in D(S)$ with probability one.

Assumption A.11 The matrices $\left\langle S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle$, $\left\langle \Omega^{1/2}\frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0, \omega), \Omega^{1/2}\frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0, \omega)\right\rangle$ and $\left\langle K^{-1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), K^{-1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle$ are positive definite and symmetric which implies that $dim(a) \leq dim(g) \leq dim(\beta)$.

Appendix 2: Proofs of Section 4 and 5

Proof of Proposition 4.1

The estimator is given by

$$\widehat{\alpha}_T = \arg\min_{\alpha \in \mathcal{A}} \left\| S_T^{1/2} g(a, \widehat{b}_T, \omega) \right\|$$

where S_T is a sequence of random bounded linear operators.

First, under Assumption A.1 to Assumption A.8, $\widehat{\alpha}_T \xrightarrow{p} \alpha_0$ by Theorem 2.1 of Newey and McFadden (1994).
Now, differentiating the objective function with respect to α and β by a mean value expansion leads to:

$$\left\langle S_T^{1/2} \tfrac{\partial g}{\partial \alpha'}(\widehat{\alpha}_T, \widehat{\beta, \omega_T}), S_T^{1/2} g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) \right\rangle = 0$$

$$\iff \left\langle S_T^{1/2} \tfrac{\partial g}{\partial \alpha'}(\widehat{\alpha}_T, \widehat{\beta}_T, \omega), S_T^{1/2} \left\{ g(\alpha_0, \beta_0, \omega) + \tfrac{\partial g}{\partial \alpha'}(\bar{a}_T, \widehat{\beta}_T, \omega)(\widehat{\alpha}_T - \alpha_0) + \tfrac{\partial g}{\partial \beta'}(\widehat{\alpha}_T, \bar{\beta}_T, \omega)(\widehat{\beta}_T - \beta_0) \right\} \right\rangle = 0$$

where $\bar{a}_T$ is on the line segment joining $\widehat{\alpha}_T$ and α_0, $\bar{\beta}_T$ is on the line segment joining $\widehat{\beta}_T$ and β_0 and under Assumption A.4, $g(\alpha_0, \beta_0, \omega) = 0$.

Using the linearity of the operator and $g(\alpha_0, \beta_0, , \omega) = 0$, we obtain:

$$\widehat{\alpha}_T - \alpha_0 = -\left\langle S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\widehat{\alpha}_T, \widehat{\beta}_T, \omega), S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\bar{a}_T, \widehat{\beta}_T, \omega) \right\rangle^{-1} \left\langle S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\widehat{\alpha}_T, \widehat{\beta}_T, \omega), S_T^{1/2} \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T, \bar{\beta}_T, \omega)(\widehat{\beta}_T - \beta_0) \right\rangle.$$

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

$$\begin{aligned} \sqrt{T}(\widehat{\alpha}_T - \alpha_0) \quad &= \quad -\left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle^{-1} \left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0, \omega)\sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle \\ & \quad +o_p(1) \end{aligned}$$

Using Assumption A.8 and Assumption A.10, one has

$$S_T^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, \widehat{\beta}_T, \omega)\sqrt{T}(\widehat{\beta}_T - \beta_0) \Rightarrow Y$$

and $Y = \mathcal{N}(0, S^{1/2} K \overline{S^{1/2}}')$.

The asymptotic variance-covariance matrix of $\sqrt{T}(\widehat{\alpha}_T - \alpha_0)$ depends on the following expression:

$$E \left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0), S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0)\sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle \overline{\left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0), S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0)\sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle}' =$$

$$E \left[\left(\int_{\underline{\omega}}^{\overline{\omega}} \frac{\partial g'}{\partial \alpha}(\alpha_0, \beta_0) S^{1/2} \overline{S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0)} \sqrt{T}(\widehat{\beta}_T - \beta_0) d\omega \right) \left(\int_{\underline{\omega}}^{\overline{\omega}} \sqrt{T}(\widehat{\beta}_T - \beta_0)' \frac{\partial g'}{\partial \beta}(\alpha_0, \beta_0) S^{1/2} \overline{S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0)} d\omega \right) \right] =$$

$$E \left[\int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\omega}}^{\overline{\omega}} \frac{\partial g'}{\partial \alpha}(\alpha_0, \beta_0) S^{1/2} \overline{S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0)} \sqrt{T}(\widehat{\beta}_T - \beta_0)\sqrt{T}(\widehat{\beta}_T - \beta_0)' \frac{\partial g'}{\partial \beta}(\alpha_0, \beta_0) S^{1/2} \overline{S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0)} d\omega \right].$$

Using

$$K^* = E \left[\overline{\frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0)} \sqrt{T}(\widehat{\beta}_T - \beta_0)\sqrt{T}(\widehat{\beta}_T - \beta_0)' \frac{\partial g'}{\partial \beta}(\alpha_0, \beta_0) \right] = \left[\overline{\frac{\partial g}{\partial \beta'}(\alpha_0, \beta_0)} \Omega \frac{\partial g'}{\partial \beta}(\alpha_0, \beta_0) \right]$$

and $K = K^*$ (the operator K is self-adjoint) imply

$$= \int_{\underline{\omega}}^{\overline{\omega}} \frac{\partial g'}{\partial \alpha}(\alpha_0, \beta_0) S^{1/2} \left(\int_{\underline{\omega}}^{\overline{\omega}} S^{1/2} K \overline{S^{1/2} (\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0) d\omega} \right) d\omega = \left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0), (S^{1/2} K \overline{S^{1/2}}') S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0) \right\rangle.$$

under the assumption that S is Hermitian. Then, for

$$\sqrt{T}(\widehat{\alpha}_T - \alpha_0) = -\left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0), S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0) \right\rangle^{-1} \left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0), Y \right\rangle + o_p(1).$$

using the previous result,

$$\left\langle S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0), Y\right\rangle \sim \mathcal{N}\left(0, \left\langle S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0), (S^{1/2}K\overline{S^{1/2}}')S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0)\right\rangle\right).$$

The result for the asymptotic distribution for a given sequence of random linear operators S_T follows.

Proof of Proposition 4.2

We first show that $\|K_n - K\| \to 0$ in probability. Consider for a given element (j,l) of $K_T(\omega_1,\omega_2)$ and the corresponding element of $K(\omega_1,\omega_2)$, the following expression:

$$\int\int \widehat{k}_{jl,T}(\omega_1,\omega_2) - k_{jl}(\omega_1,\omega_2)\varphi(\omega_1)\varphi(\omega_2)d\omega_1 d\omega_2.$$

where $\widehat{k}_{jl,T}(\omega_1,\omega_2) = \frac{\partial g_j}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_1)\widehat{\Omega}\overline{\frac{\partial g_l}{\partial \beta}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)'}$ for $\widehat{\alpha}_T^1$ ia consistent first step estimator of α_0 and $k_{jl}(\omega_1,\omega_2) = \frac{\partial g_j}{\partial \beta'}(\alpha_0,\beta_0,\omega_1)\Omega\overline{\frac{\partial g_l}{\partial \beta}(\alpha_0,\beta_0,\omega_2)'}$. Under $\widehat{\alpha}_T^1 \xrightarrow{p} \alpha_0$, $\widehat{\beta}_T \xrightarrow{p} \beta_0$ and $\widehat{\Omega} \xrightarrow{p} \Omega$, the expression above converges to zero. This holds for $\forall j,l = 1,\dots,J$ which implies that $\|K_n - K\| \to 0$ in probability. Let $\widehat{K}_T f(\omega_1)$ denote

$$\widehat{K}_T f(\omega_1) = \left(\sum_{l=1}^J \int \widehat{k}_{jl,T}(\omega_1,\omega_2)f_l(\omega_2)\varphi(\omega_2)d\omega_2\right)_{j=1,\dots,J}.$$

Then $\widehat{K}_T f(\omega_1)$ can be written as:

$$\widehat{K}_T f(\omega_1) = \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_1)\Omega_T^{1/2}\int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)}f(\omega_2)\varphi(\omega_2)d\omega_2.$$

for $f(\omega) = (f_1(\omega), f_2(\omega),\dots,f_J(\omega))'$ where $f_j(\omega)$ for $j = 1,\dots,J$ are scalar functions in $L^2(\varphi)$. In this case, $R(K_T)$ is the space spanned by $\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)\Omega_T^{1/2}$ with rank at most equals to q. The eigenfunctions γ_i is necessarily of the form $\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)\Omega_T^{1/2}D_i$ where the matrix D_i is of dimension $q\times 1$ and $D = [D_1 \;\; D_2 \;\; \dots \;\; D_q]$ where D is of dimension $q \times q$. By virtue of the Mercer's theorem, the vector $\gamma_i(\omega)$ of eigenfunctions satisfies

$$(K_T\gamma_i)(\omega_1) = \lambda_i\gamma_i(\omega_1)$$

where λ_i is the corresponding eigenvalue of the eigenfunctions vector γ_i.

Using $\gamma_{i,T}(\omega) = \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega)\Omega_T^{1/2}D_i$ yields:

$$K_T\gamma_i(\omega_1) = \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_1)\Omega_T^{1/2}\int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)}\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)\Omega_T^{1/2}D_i\varphi(\omega_2)d\omega_2$$

Let $D = [D_1 \;\; D_2 \;\; \dots \;\; D_q]$ and Λ denote the matrices containing the eigenvectors and the eigenvalues of the following $q\times q$ matrix:

$$\int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)}\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)\Omega_T^{1/2}\varphi(\omega_2)d\omega_2.$$

More specifically, the eigenvectors $D_i, i = 1,\dots,q$ and the corresponding eigenvalues λ_i solve the following system of q equations:

$$\int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)}\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)\Omega_T^{1/2}\varphi(\omega_2)d\omega_2 D_i = \lambda_i D_i.$$

Using the spectral decomposition,

$$\begin{aligned} K_T\gamma_i(\omega_1) &= \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_1)\Omega_T^{1/2}\int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)}\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_2)\Omega_T^{1/2}D\varphi(\omega_2)d\omega_2 \\ &= \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_1)\Omega_T^{1/2}D_i\lambda_i = \lambda_i\gamma_i(\omega_1), \end{aligned}$$

which implies that $\gamma_i(\omega)$ is given by $\frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T^1,\widehat{\beta}_T,\omega_1)\Omega_T^{1/2}D_i$. A consistent estimator of the Moore-Penrose generalized inverse is then given by:

$$K_T^{-1}f(\omega_1) = \sum_{i=1}^q \frac{1}{\lambda_{i,T}}\gamma_{i,T}(\omega_1)\langle f, \gamma_{i,T}\rangle.$$

Proof of Proposition 4.3

After imposing that $S^{1/2} = K^{-1/2}$, one obtains

$$\left\langle K^{-1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0), Y\right\rangle \sim \mathcal{N}\left(0, \left\langle \frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0), K^{-1}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0)\right\rangle\right).$$

Consequently, using $S^{1/2} = K^{-1/2}$,

$$\begin{aligned}\lim_{T\to\infty} Var\left(\sqrt{T}(\widehat{\alpha}_T - \alpha_0)\right) &= \left\langle \frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0), K^{-1}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0)\right\rangle^{-1}\left\langle \frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0), K^{-1}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0)\right\rangle \\ &\quad\times \left\langle \frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0), K^{-1}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0)\right\rangle^{-1} \\ &= \left\langle \frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0), K^{-1}\frac{\partial g}{\partial \alpha'}(\alpha_0,\beta_0)\right\rangle^{-1}.\end{aligned}$$

Proof of Proposition 4.4

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

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

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

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

This yields

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

Using the notation,

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

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

After multiplying $K_T h(\omega)$ by $\Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\omega_1)}$ and integrating, one obtains:

$$\begin{aligned}&\int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\omega_1)}\frac{\partial g}{\partial \beta'}(\omega)\Omega_T^{1/2}\varphi(\omega)d\omega \int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\omega_1)}h(\omega_1)\varphi(\omega_1)d\omega_1 \\ &\qquad = \int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\omega_1)}g(\omega_1)\varphi(\omega_1)d\omega_1 = s.\end{aligned}$$

using $K_T h(\omega) = g(\omega)$. Denoting $W = \int \Omega_T^{1/2}\overline{\frac{\partial g'}{\partial \beta}(\omega)}\frac{\partial g}{\partial \beta'}(\omega)\Omega_T^{1/2}\varphi(\omega)d\omega$, 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 $s'\widetilde{W}^2\bar{s}$. This provides the result. When W is of rank equal to p, then $\widetilde{W} = W^{-1}$.

Proof of Proposition 4.7

To derive the limiting distribution of the statistic, we need the following Lemma from Vuong (1989):

Lemma 8.1. *(Vuong, 1989) Let W be a vector of q random variables distributed as $N(0,\Sigma)$ with rank $\Sigma \leq q$. Let Q be a $q \times q$ real symmetric matrix. Then*

$$W'QW \sim \sum_{i=1}^{q} \lambda_i Z_i^2$$

where $\lambda = (\lambda_1, \cdots, \lambda_q)$ is the vector of eigenvalues of $Q\Sigma$ and $\{Z_i\}_{i=1}^q$ are m independent standard normal variables. Moreover, the q eigenvalues are all real and nonnegative if Q is positive semi-definite.

Using the mean value expansion, one has:

$$S_T^{1/2} g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) = S_T^{1/2} g(\alpha_0, \beta_0, \omega) + S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\bar{a}, \widehat{\beta}_T, \omega)(\widehat{\alpha}_T - \alpha_0) + S_T^{1/2} \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T, \bar{\beta}, \omega)(\widehat{\beta}_T - \beta_0)$$

where $\bar{a}$ 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 $\sqrt{T}(\widehat{\alpha}_T - \alpha_0)$ derived above and $g(\alpha_0, \beta_0, \omega) = 0$, it follows that:

$$\begin{aligned}
\sqrt{T} S_T^{1/2} g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) &= -S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\bar{a}, \widehat{\beta}_T, \omega) \left\langle S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\widehat{\alpha}_T, \widehat{\beta}_T, \omega), S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\bar{a}, \widehat{\beta}_T, \omega) \right\rangle^{-1} \\
&\quad \times \left\langle S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\widehat{\alpha}_T, \widehat{\beta}_T, \omega), S_T^{1/2} \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T, \bar{\beta}, \omega) \sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle \\
&\quad + S_T^{1/2} \frac{\partial g}{\partial \beta'}(\widehat{\alpha}_T, \bar{\beta}, \omega) \sqrt{T}(\widehat{\beta}_T - \beta_0).
\end{aligned}$$

Since $\widehat{\alpha}_T \stackrel{p}{\to} \alpha_0$, $\widehat{\beta}_T \stackrel{p}{\to} \beta_0$ and under the assumption that $\|S_T - S\| \to 0$ in probability:

$$\begin{aligned}
\sqrt{T} S_T^{1/2} g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) &= -S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle^{-1} \\
&\quad \times \left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega) \sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle \\
&+ \quad S_T^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega) \sqrt{T}(\widehat{\beta}_T - \beta_0) + o_p(1).
\end{aligned}$$

Using results obtained from the proof of Proposition 4.1, the variance of the first right-hand side term of the previous expression above is given by:

$$\begin{aligned}
S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) &\left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle^{-1} \left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), (S^{1/2} K \overline{S^{1/2}}') S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle \\
&\times \overline{\left\langle S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega), S^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega) \right\rangle'}^{-1} \overline{\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)'} \overline{S^{1/2}}',
\end{aligned}$$

whereas the variance of the second right-hand side term is $S^{1/2} K \overline{S^{1/2}}'$, and the covariance between the first and the second terms is function of

$$\begin{aligned}
E\left[S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega) \sqrt{T}(\widehat{\beta}_T - \beta_0) \overline{\left\langle S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, b_0, \omega), S_T^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega) \sqrt{T}(\widehat{\beta}_T - \beta_0) \right\rangle}' \right] &= \\
E\left[S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega) \sqrt{T}(\widehat{\beta}_T - \beta_0) \int_{\underline{\omega}}^{\overline{\omega}} \sqrt{T}(\widehat{\beta}_T - \beta_0)' \overline{\frac{\partial g'}{\partial \beta}(\alpha_0, b_0, \omega)} \overline{S_T^{1/2}}' S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, b_0, \omega) d\omega \right] &= \\
S^{1/2} \frac{\partial g}{\partial \beta'}(\alpha_0, b_0, \omega) \Omega \int_{\underline{\omega}}^{\overline{\omega}} \overline{\frac{\partial g'}{\partial \beta}(\alpha_0, b_0, \omega)} \overline{S_T^{1/2}}' S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, b_0, \omega) d\omega &= \\
\left(S_T^{1/2} K \overline{S_T^{1/2}}' \right) S_T^{1/2} \frac{\partial g}{\partial \alpha'}(\alpha_0, b_0, \omega).
\end{aligned}$$

Collecting the previous results and by Assumption A.10, one gets:

$$S_T^{1/2} \sqrt{T} g(\widehat{\alpha}_T, \widehat{\beta}_T, \omega) \stackrel{d}{\to} \mathcal{N}\left(0, \left\langle S^{1/2}[I - M(\omega)], (S^{1/2} K \overline{S^{1/2}}') S^{1/2}[I - M(\omega)] \right\rangle \right)$$

with $M(\omega) = S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\left\langle S^{1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, , \omega), S^{-1/2}\frac{\partial g}{\partial \alpha'}(\alpha_0, \beta_0, \omega)\right\rangle^{-1}\overline{\frac{\partial g'}{\partial \alpha}(\alpha_0, \beta_0, \omega)S^{1/2}}'$.

Finally, using Lemma 8.1,

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

where λ_i are the eigenvalues of $\left\langle S^{1/2}[I - M(\omega)], (S^{1/2}K\overline{S^{1/2}}')S^{1/2}[I - M(\omega)]\right\rangle$.

Proof of Proposition 4.8

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 $|\omega| < |\omega_0|$ and by Assumption 4.3, $-h(|\omega_0|\kappa_T + h(|\omega_0|\kappa_T \to -\infty$ 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 5.1

Consider the vector of just- or over-identified estimating equations defined by Eq. 2.10. 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:

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

where

$$\frac{\partial}{\partial a}\left\{\text{vech}\left(\Sigma - A(0)A(0)'\right)' \text{vech}\left(\Sigma - A(0)A(0)'\right)\right\} = -2(D_N^+\left(A(0) \otimes I_N\right))'\text{vech}\left(\widehat{\Sigma}_T - A(0)A(0)'\right)$$

by the result in Lütkepohl, (2007, p. 363) and

$$\begin{aligned}
\frac{\partial}{\partial a}\left\{\text{vec}\left(C(z)A(0)\right)' H'H\overline{\text{vec}\left(C(z)A(0)\right)}\right\} &= \frac{\partial}{\partial a}\left\{\text{vec}(A(0))'\left(I_N \otimes C(z)\right)' H'H\overline{(I_N \otimes C(z))}\text{vec}\left(A(0)\right)\right\} \\
&= 2\left(I_N \otimes C(z)\right)' H'H\overline{(I_N \otimes C(z))}\text{vec}\left(A(0)\right) \\
&= 2\left|\left(I_N \otimes C'(z)\right) H'\right|^2 \text{vec}\left(A(0)\right).
\end{aligned}$$

This gives the first order conditions. 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 4.4 and $\frac{\partial \text{vec}(C(z))A(0))}{\partial \Phi'} = (A(0)' \otimes I)\frac{\partial \text{vec}(C(z))}{\partial \Phi'}$ where $\frac{\partial \text{vec}(C(z))}{\partial \Phi'}$ 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).

Proof of Proposition 5.3

The objective function of the C-ALS problem of a bivariate VAR model is based on the W and $\underline{s}(\cdot,\cdot)$ matrices, with

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

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

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

and

$$\begin{aligned}
\widehat{W}_T &= \int_{\underline{\omega}}^{\overline{\omega}} \sum_{j=0}^{\infty} \left[\left(\frac{\partial \widehat{c}_{11,j}}{\partial \beta'} e^{i\omega j} \widehat{a}^1_{12,T} + \frac{\partial \widehat{c}_{12,j}}{\partial \beta'} e^{i\omega j} \right) \Omega^{1/2} \right]' \sum_{l=0}^{\infty} \left[\left(\frac{\partial \widehat{c}_{11,l}}{\partial \beta'} e^{-i\omega l} \widehat{a}^1_{12,T} + \frac{\partial \widehat{c}_{12,l}}{\partial \beta'} e^{-i\omega l} \right) \Omega^{1/2} \right] d\omega \\
&= \int_{\underline{\omega}}^{\overline{\omega}} \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \left[\left(\left(\frac{\partial \widehat{c}_{11,j}}{\partial \beta'} \widehat{a}^1_{12,T} + \frac{\partial \widehat{c}_{12,j}}{\partial \beta'} \right) \Omega^{1/2} \right)' \left(\left(\frac{\partial \widehat{c}_{11,l}}{\partial \beta'} \widehat{a}^1_{12,T} + \frac{\partial \widehat{c}_{12,l}}{\partial \beta'} \right) \Omega^{1/2} \right) \right] \exp((j-l)\omega) d\omega \\
&= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \left[\left(\left(\frac{\partial \widehat{c}_{11,j}}{\partial \beta'} \widehat{a}^1_{12,T} + \frac{\partial \widehat{c}_{12,j}}{\partial \beta'} \right) \Omega^{1/2} \right)' \left(\left(\frac{\partial \widehat{c}_{11,l}}{\partial \beta'} \widehat{a}^1_{12,T} + \frac{\partial \widehat{c}_{12,l}}{\partial \beta'} \right) \Omega^{1/2} \right) \right] \int_{\underline{\omega}}^{\overline{\omega}} \cos((l-j)\omega) d\omega
\end{aligned}$$

.

and the last equality holds by the symmetry of the interval around zero. The objective function can be then rewritten as:

$$\underline{s}(\tilde{a}_{12}, \hat{\beta}_T)' \left(\widehat{W}^2 \right)^{-1} \overline{\underline{s}(\tilde{a}_{12}, \hat{\beta}_T)} = (\widehat{s}_{11,T} \tilde{a}_{12} + \widehat{s}_{12,T})' (\widehat{W}^2_T)^{-1} (\widehat{s}_{11,T} \tilde{a}_{12} + \widehat{s}_{12,T}) \tag{8.24}$$

where $\widehat{s}_{11,T}$, $\widehat{s}_{12,T}$ and $\widehat{W}_T$ are given by:

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

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

and

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

The minimizer of the objective function 8.24 is given by $\widehat{a}_{12,T} = -\frac{\widehat{s}'_{11,T} (\widehat{W}^2_T)^{-1} \widehat{s}_{12,T}}{\widehat{s}'_{11,T} (\widehat{W}^2_T)^{-1} \widehat{s}_{11,T}}$.

First-order conditions with restrictions on $|C(z)A(0)|^2 = |A(z)|^2$ in Proposition 5.4. We provide the relevant first-order partial derivatives in the following proposition.

Proposition 8.1. *The first-order partial derivatives of the estimating equations,*

$$g(a, \hat{\beta}_T, \omega) = vech\left(\widehat{C}(z) \widehat{\Sigma}_T \widehat{C}^*(z) - \widehat{C}(z) A(0) A(0)' \widehat{C}^*(z) \right)$$

with respect to a, Φ, and σ are respectively given by:

$$\frac{\partial g}{\partial a'}(a, \hat{\beta}_T, \omega) \quad = \quad -L_N \left(\overline{C}(z) \otimes C(z) \right) \left(I_{N^2} + K_{NN} \right) \left(A(0) \otimes I_N \right)$$

and

$$\begin{aligned}\frac{\partial g}{\partial \Phi'}(a,\hat{\beta}_T,\omega) &= L_N\left[\left(I_N\otimes \widehat{C}(z)(\widehat{\Sigma}_T - A(0)A(0)')\right)\frac{\partial vec(\widehat{C}^*(z))}{\partial \Phi'} + \left(\overline{\widehat{C}}(z)(\widehat{\Sigma}_T - A(0)A(0)')\ \otimes I_N\right)\frac{\partial vec(\widehat{C}(z))}{\partial \Phi'}\right]\\ \frac{\partial g}{\partial \sigma'}(a,\hat{\beta}_T,\omega) &= L_N(\overline{C}(z)\otimes C(z))\end{aligned}$$

with $\widehat{C}^{*\prime}(z) = \overline{\widehat{C}(z)}$, L_N *is an* $\left(\frac{1}{2}N(N+1)\times N^2\right)$ *elimination matrix,* K_{NN} *is the commutator matrix for which* $K_{NN}vec(X) = vec(X')$ *and* X *is an arbitrary* $N\times N$ *matrix.*

Proof: One has

$$\begin{aligned}\frac{\partial g}{\partial a'}(a,\hat{\beta}_T,\omega) &= -L_N\frac{\partial}{\partial a'}\text{vec}\left(\widehat{C}(z)A(0)A(0)'\widehat{C}^*(z)\right)\\ &= -L_N\left(\overline{\widehat{C}}(z)\otimes\widehat{C}(z)\right)\frac{\partial}{\partial a'}\text{vec}\left(A(0)A(0)'\right)\\ &= -L_N\left(\overline{\widehat{C}}(z)\otimes\widehat{C}(z)\right)\left[(I_N\otimes A(0))\frac{\partial\text{vec}(A(0)')}{\partial a'} + (A(0)\otimes I_N)\frac{\partial\text{vec}(A(0))}{\partial a'}\right]\\ &= -L_N\left(\overline{C}(z)\otimes C(z)\right)\left((I_N\otimes A(0))K_{NN} + (A(0)\otimes I_N)\right)\\ &= -L_N\left(\overline{C}(z)\otimes C(z)\right)(I_{N^2}+K_{NN})(A(0)\otimes I_N)\end{aligned}$$

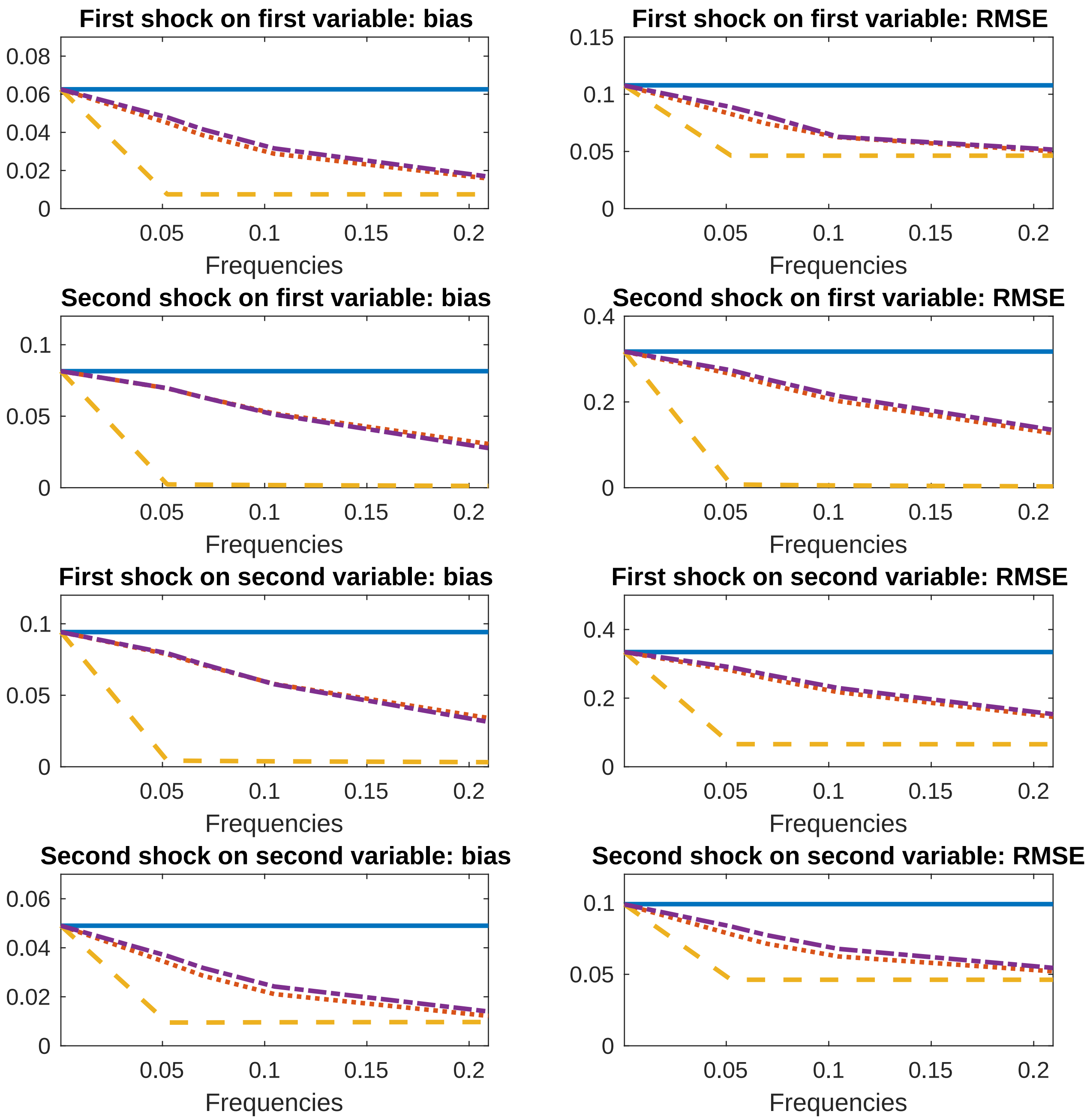
using that $K_{NN}(A(0)\otimes I_N) = (I_N\otimes A(0))K_{NN}$. On the other hand, the partial derivatives with respect to Φ are obtained using the standard product rule for vector differentiation with the vec operator:

$$\frac{\partial\text{vec}\,(A(\theta)CD(\theta))}{\partial\theta'} = (I_N\otimes A(\theta)C)\frac{\partial\text{vec}\,(D(\theta))}{\partial\theta'} + \left(D(\theta)'C'\otimes I_N\right)\frac{\partial\text{vec}\,(A(\theta))}{\partial\theta'}.$$

Moreover, $\frac{\partial\text{vec}(C(z))}{\partial\Phi'}$ can be derived from Lütkepohl (2007, p. 111). Finally, using the property $\text{vec}(PQP^*) = (P^{\star\prime}\otimes P)\,\text{vec}(Q)$ and $P^{\star\prime} = \overline{P}$ where P is a complex-valued matrix, one has

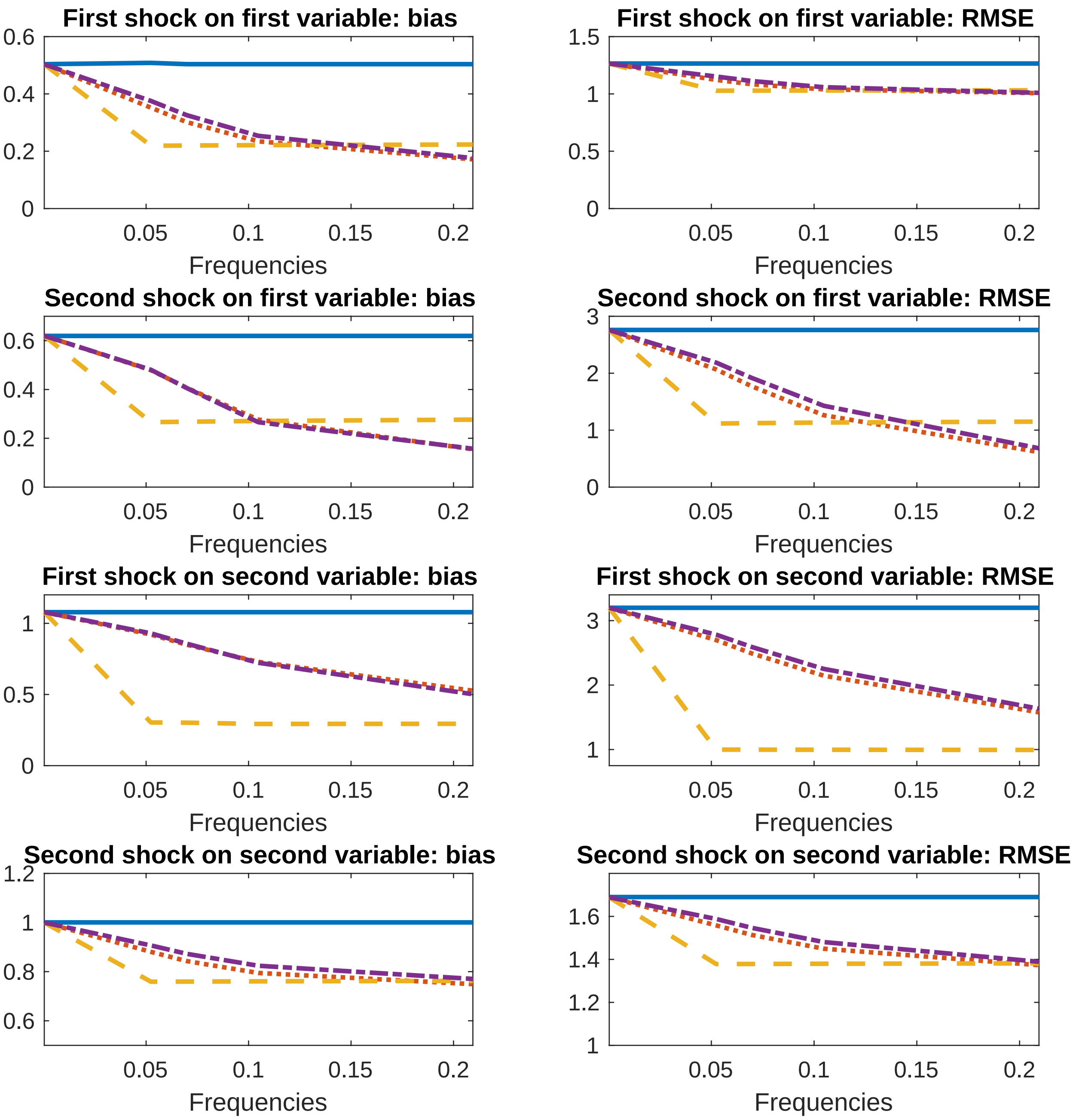
$$\frac{\partial g}{\partial\sigma'}(a,\hat{\beta}_T,\omega) = L_N(\overline{C}(z)\otimes C(z))\frac{\partial\text{vec}(\Sigma)}{\partial\sigma'} = L_N(\overline{C}(z)\otimes C(z)).$$

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



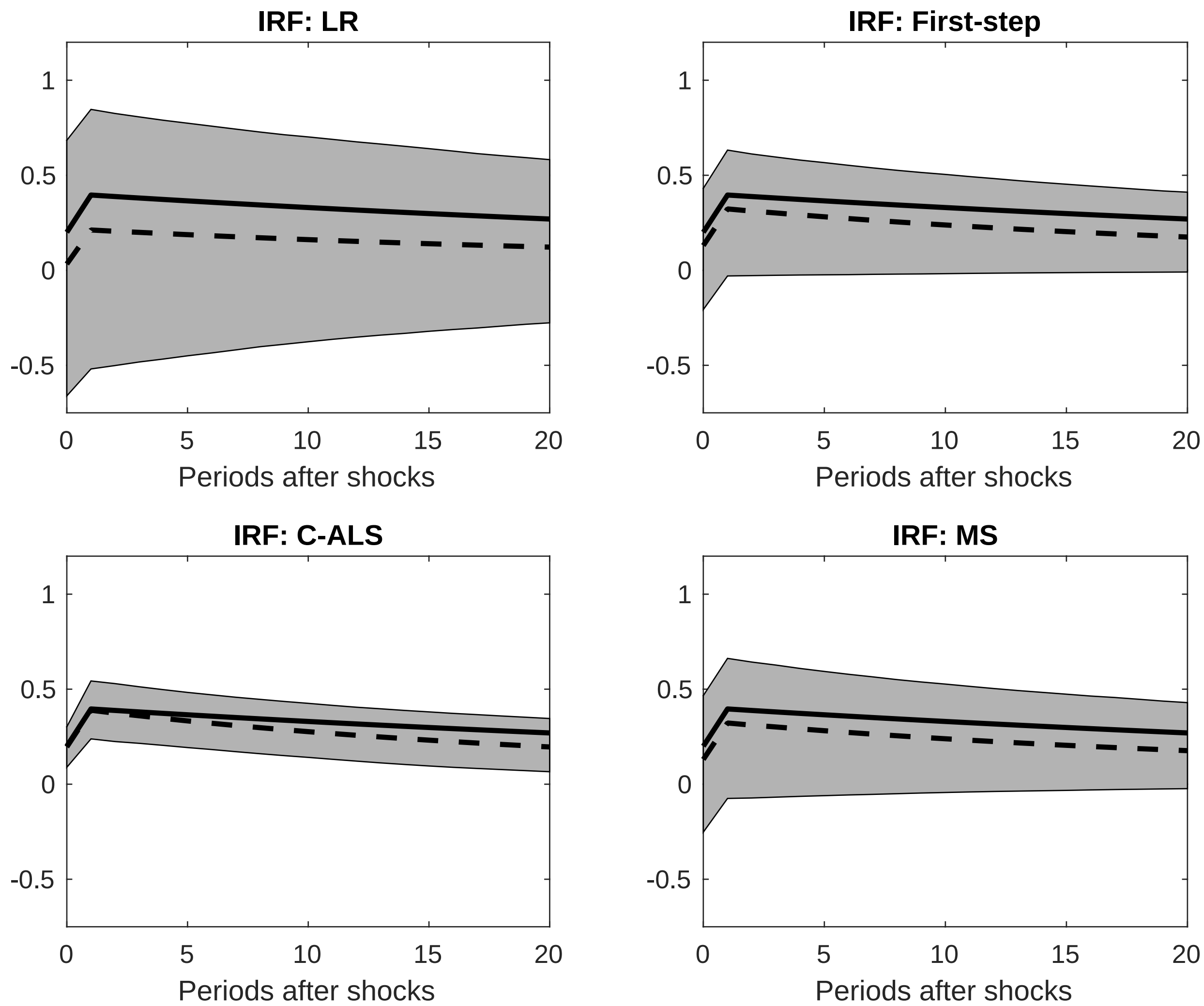
Note: The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

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



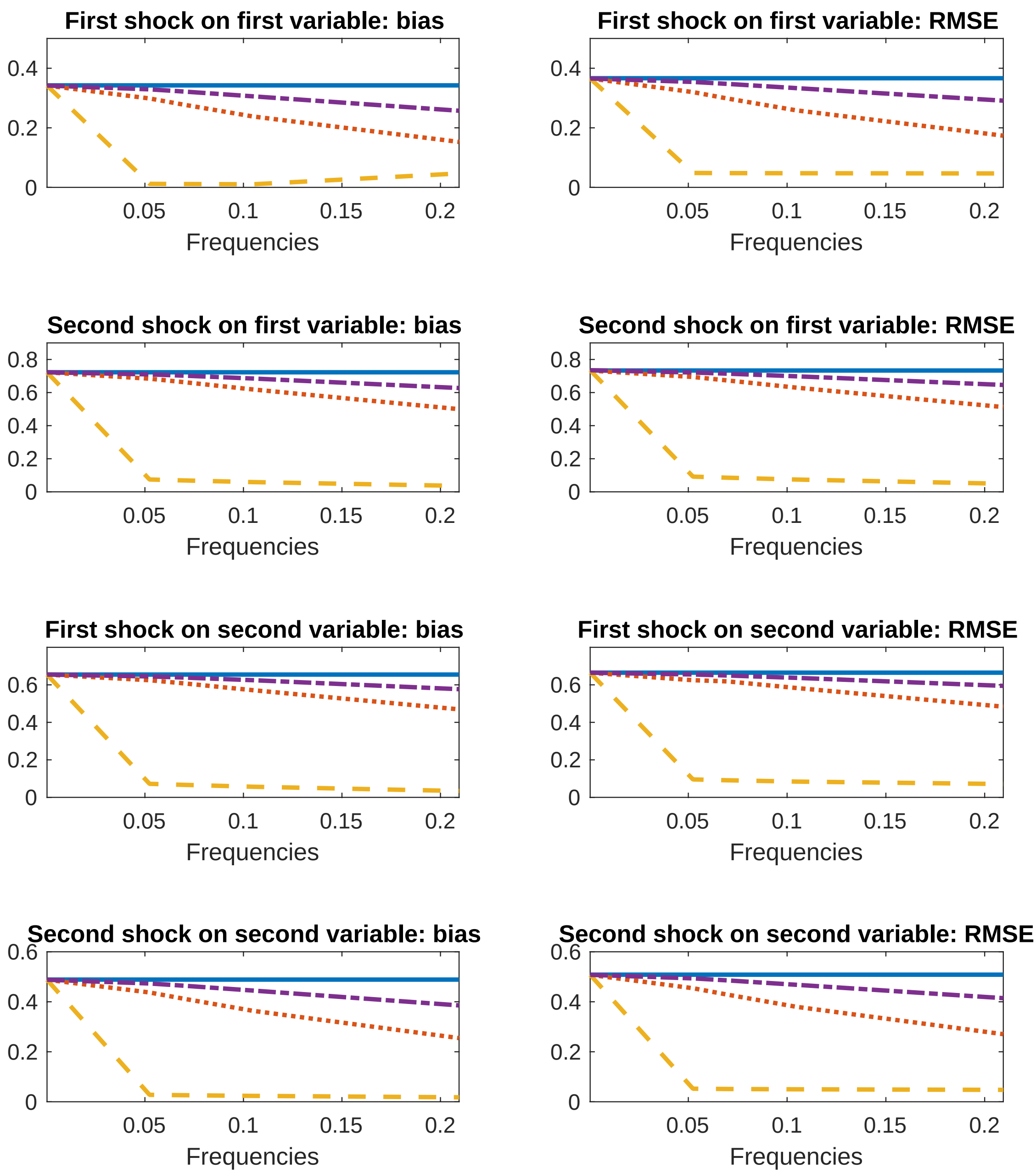
Note: The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

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



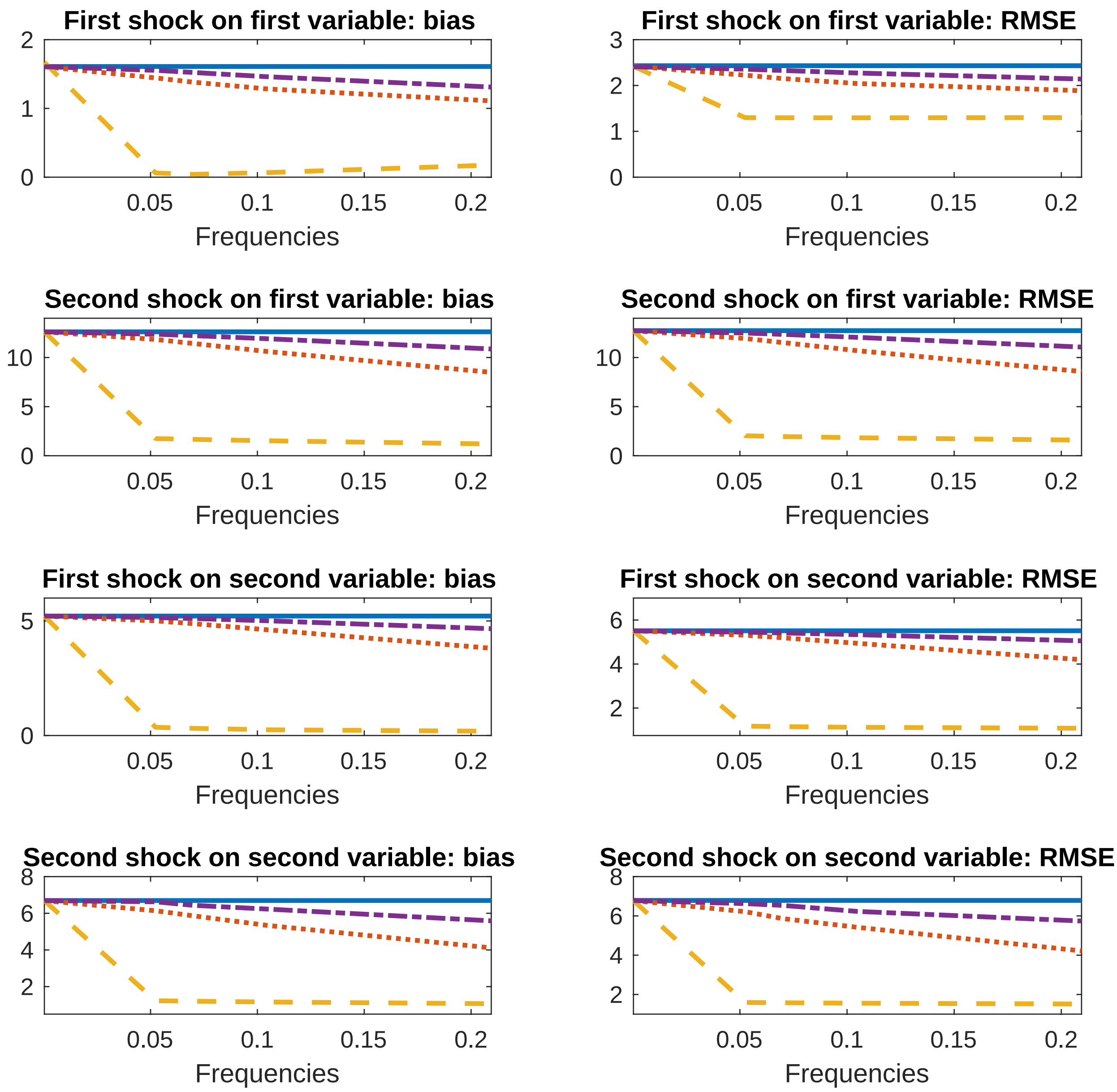
Note: Confidence intervals are based on the 95–percentile from 10,000 Monte–Carlo experiments.

Figure 4: Contemporaneous bias and RMSE using a VAR(1) model with $\rho = .95$ and $\delta = .1$



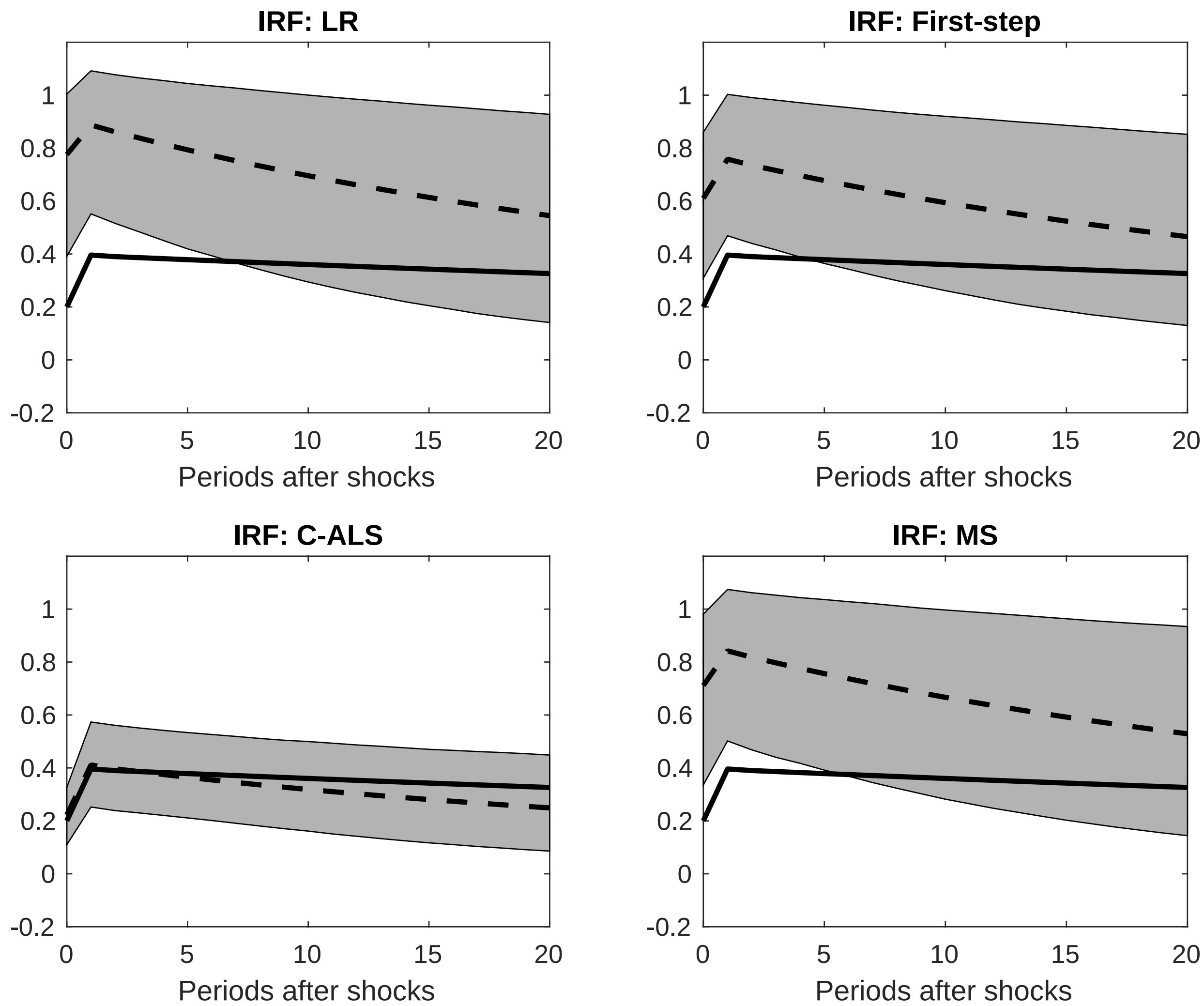
Note: The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

Figure 5: Cumulative Bias and RMSE up to 12 quarters using a VAR(1) model with $\rho = .95$ and $\delta = .1$



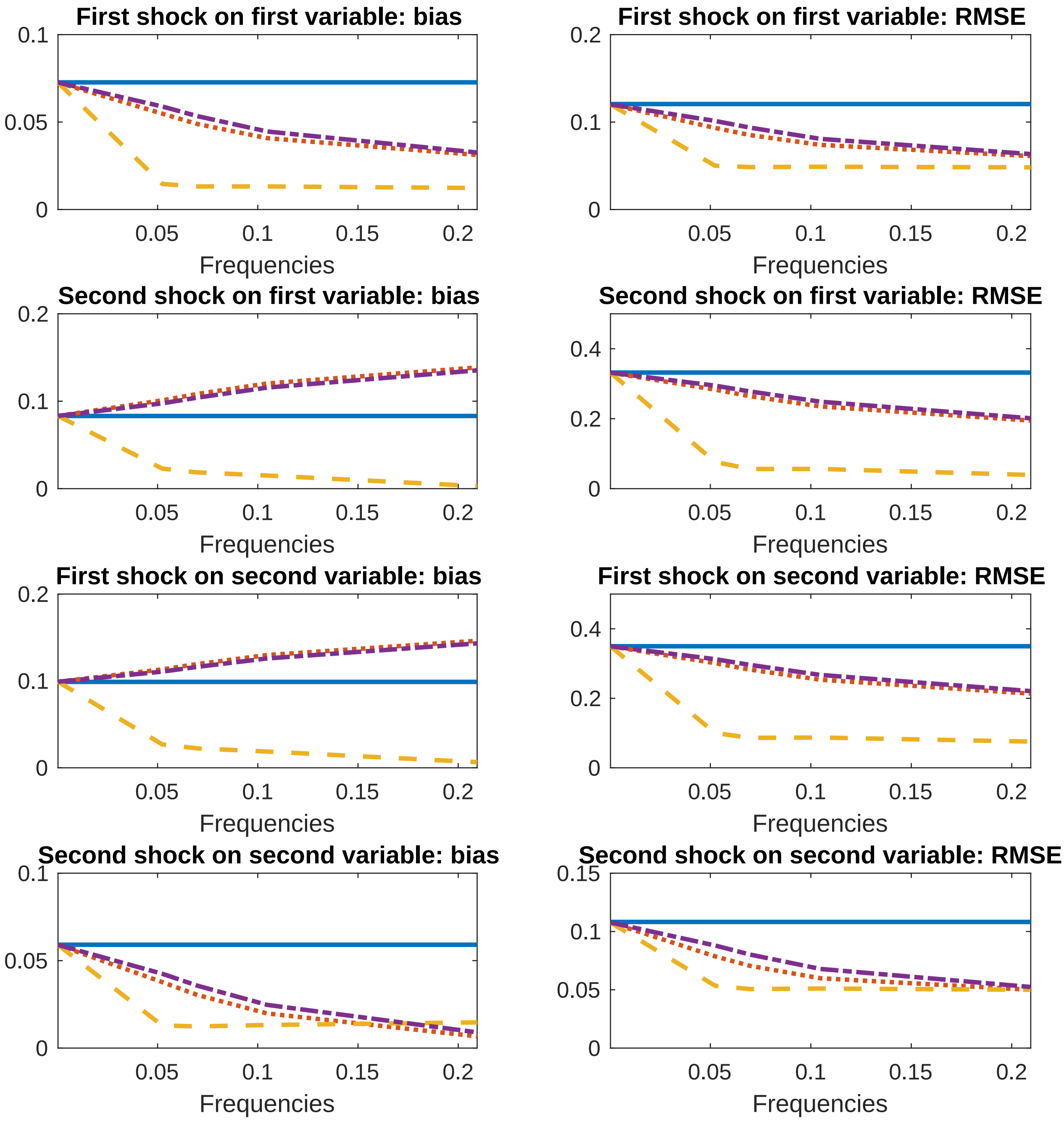
Note: The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

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



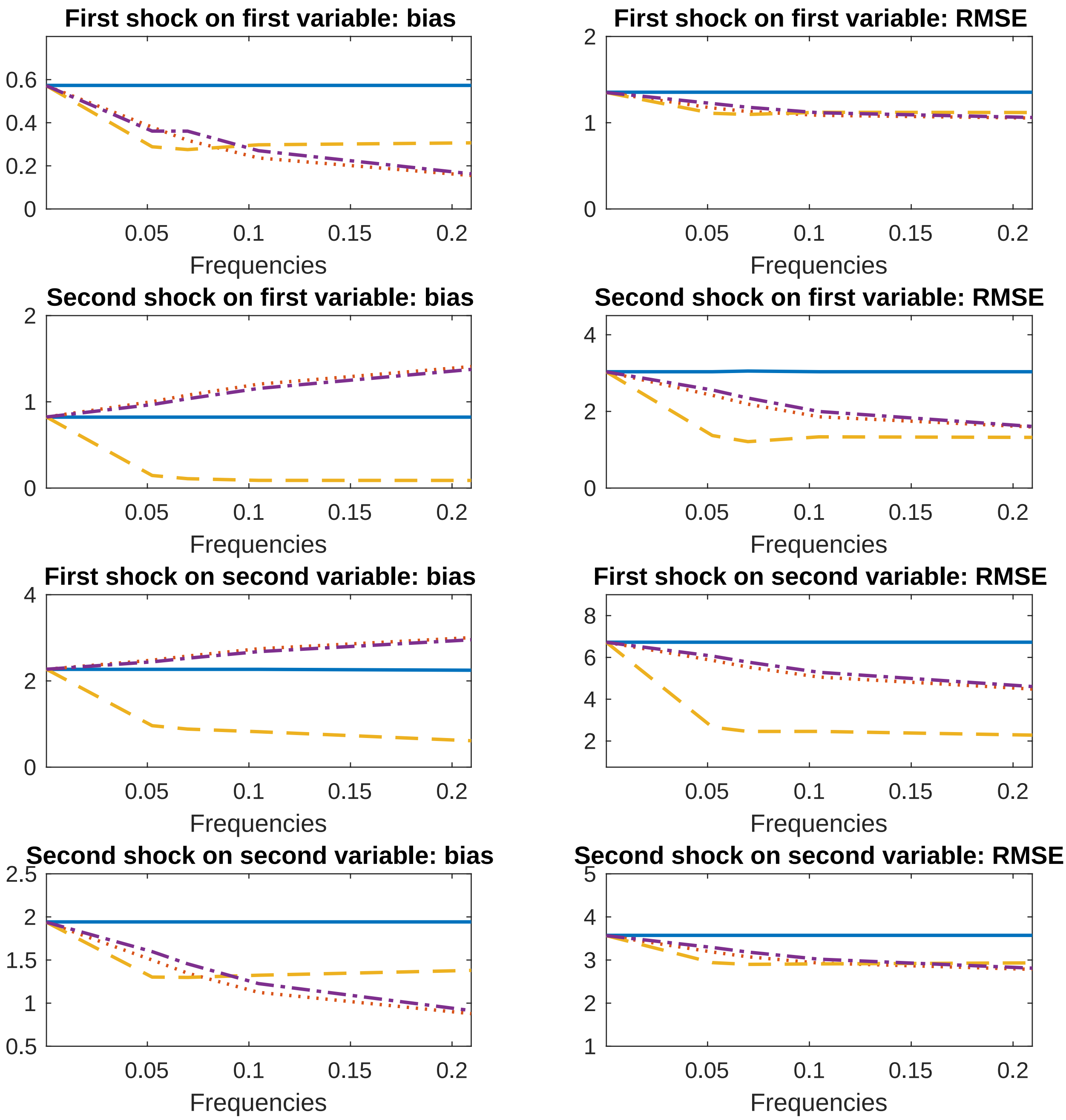
Note: Confidence intervals are based on the 95–percentile from 10,000 Monte–Carlo experiments.

Figure 7: Contemporaneous bias and RMSE using a VAR(2) model with $\rho = .95$ and $\delta = 0$



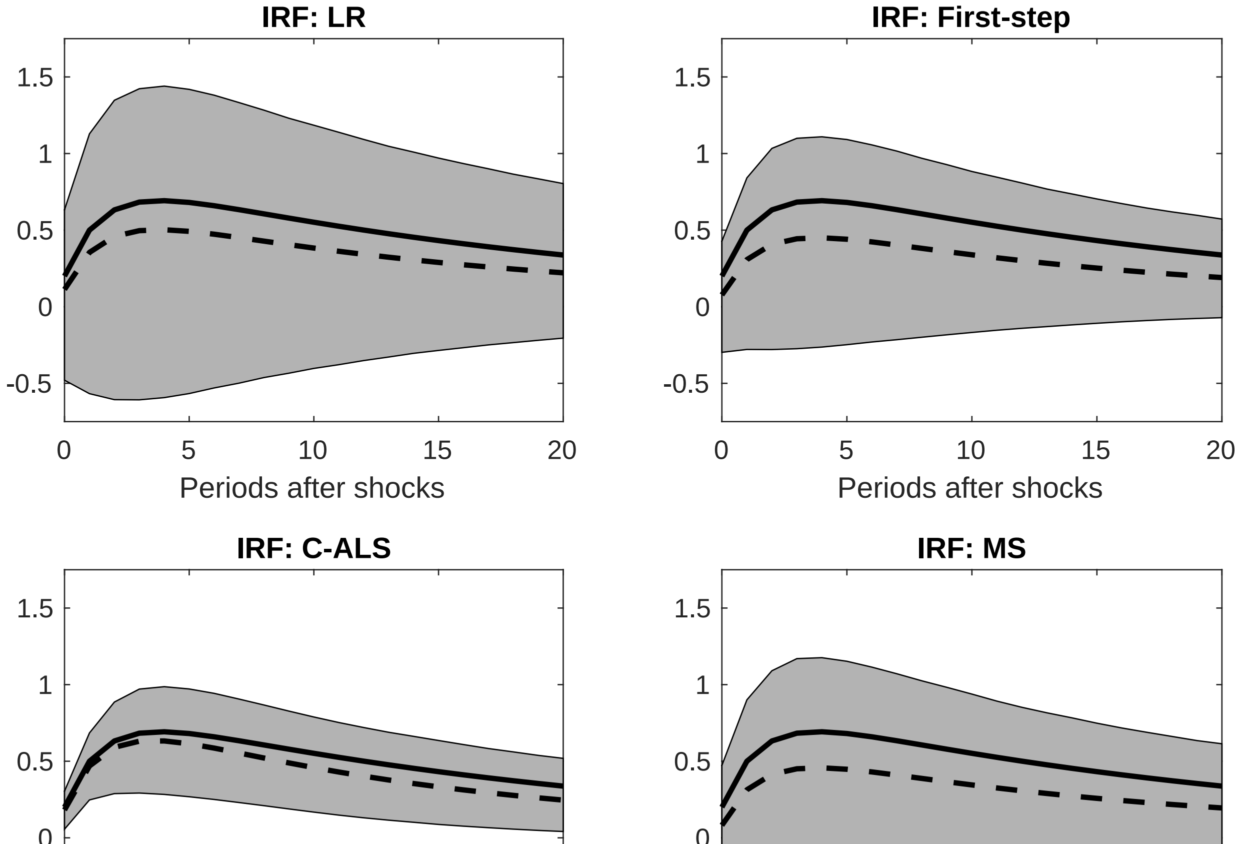
Note: The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

Figure 8: Cumulative Bias and RMSE up to 12 quarters using a VAR(2) model with $\rho = .95$ and $\delta = 0$



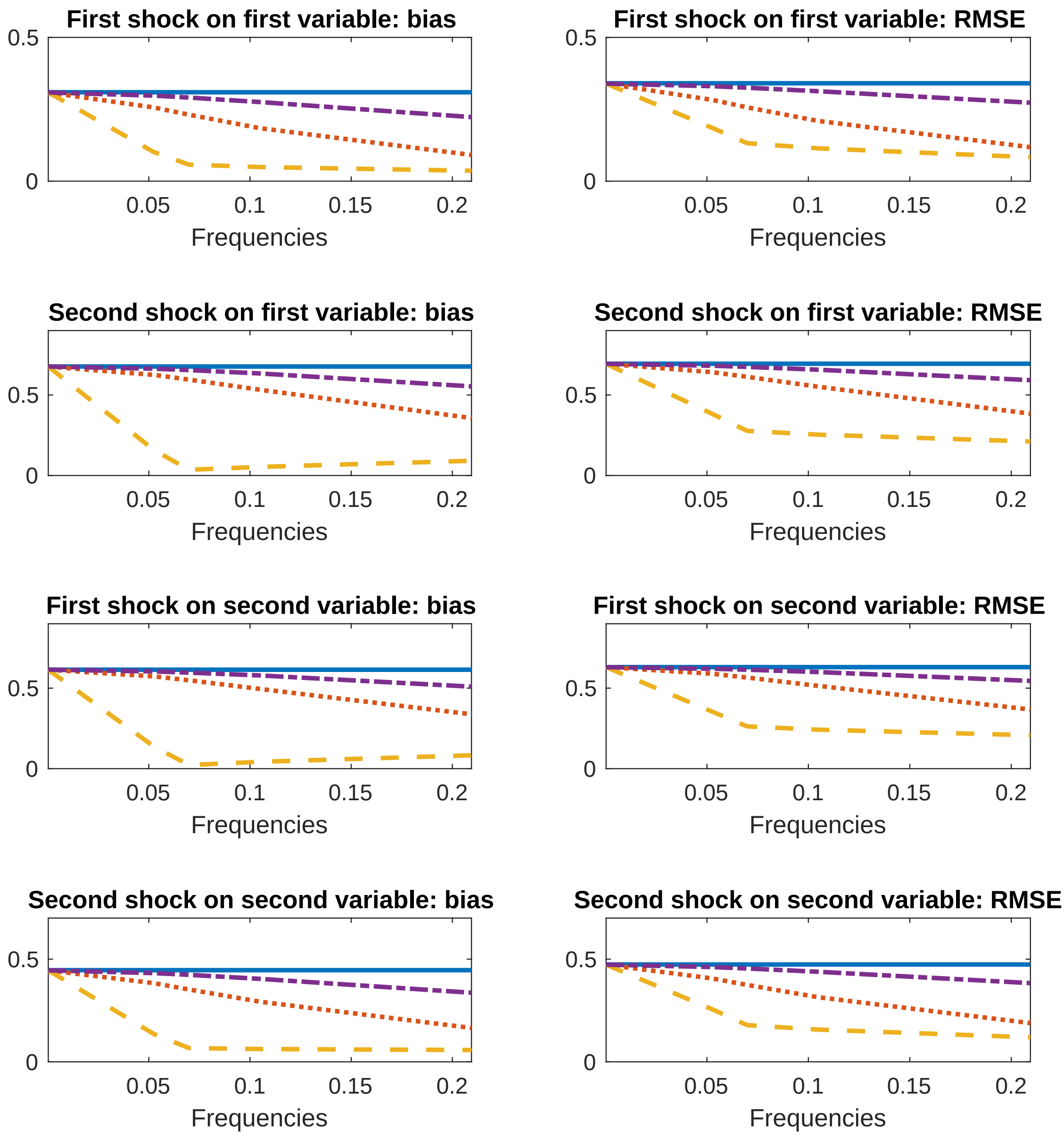
Note: The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

Figure 9: Impulse Responses for the first shock on second variable with $n = 60$, $\rho = .95$ and $\delta = 0$



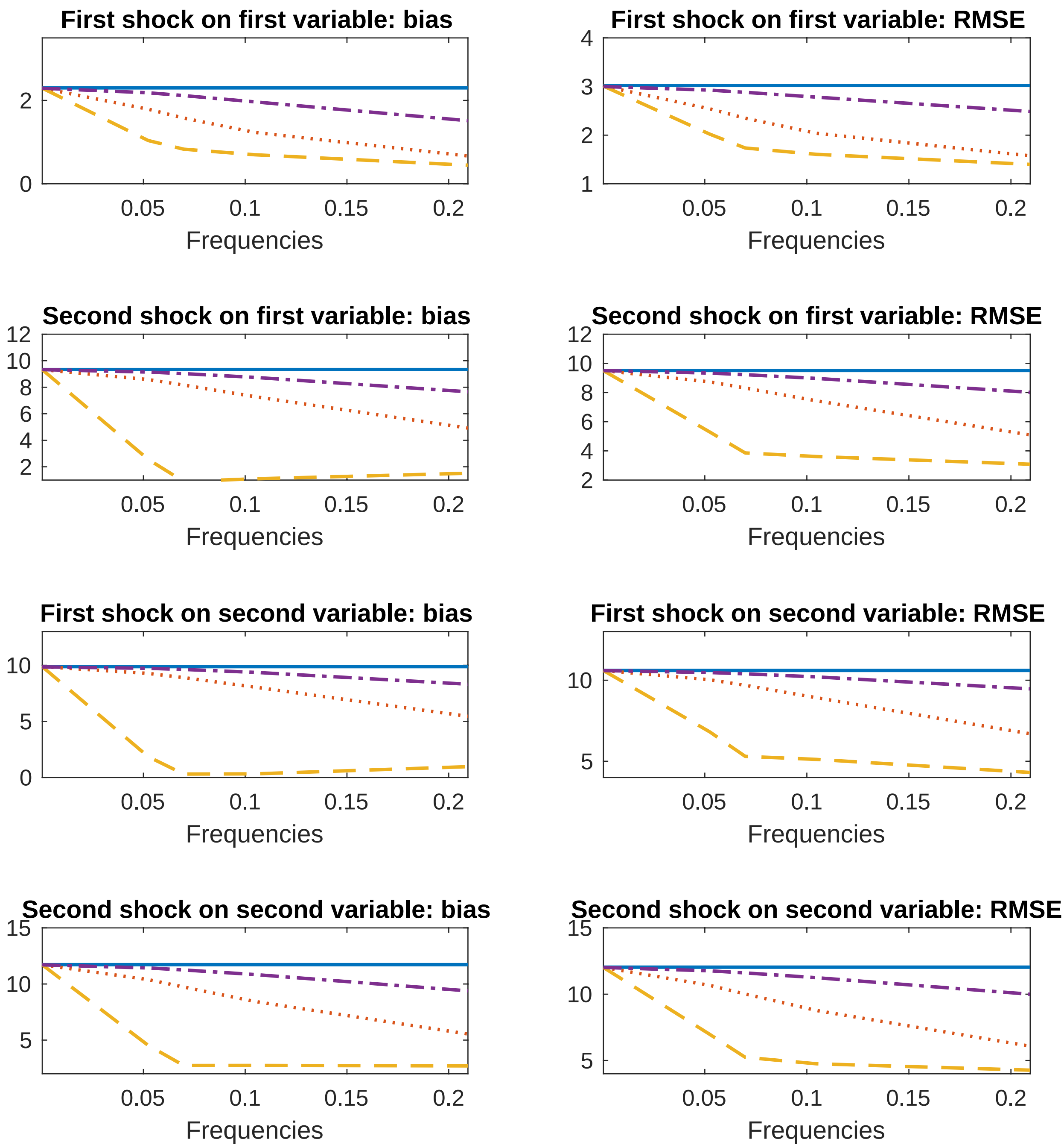
Note: Confidence intervals are based on the 95–percentile from 10,000 Monte–Carlo experiments.

Figure 10: Contemporaneous bias and RMSE at the impact using a VAR(2) model with $\rho = .95$ and $\delta = .04$



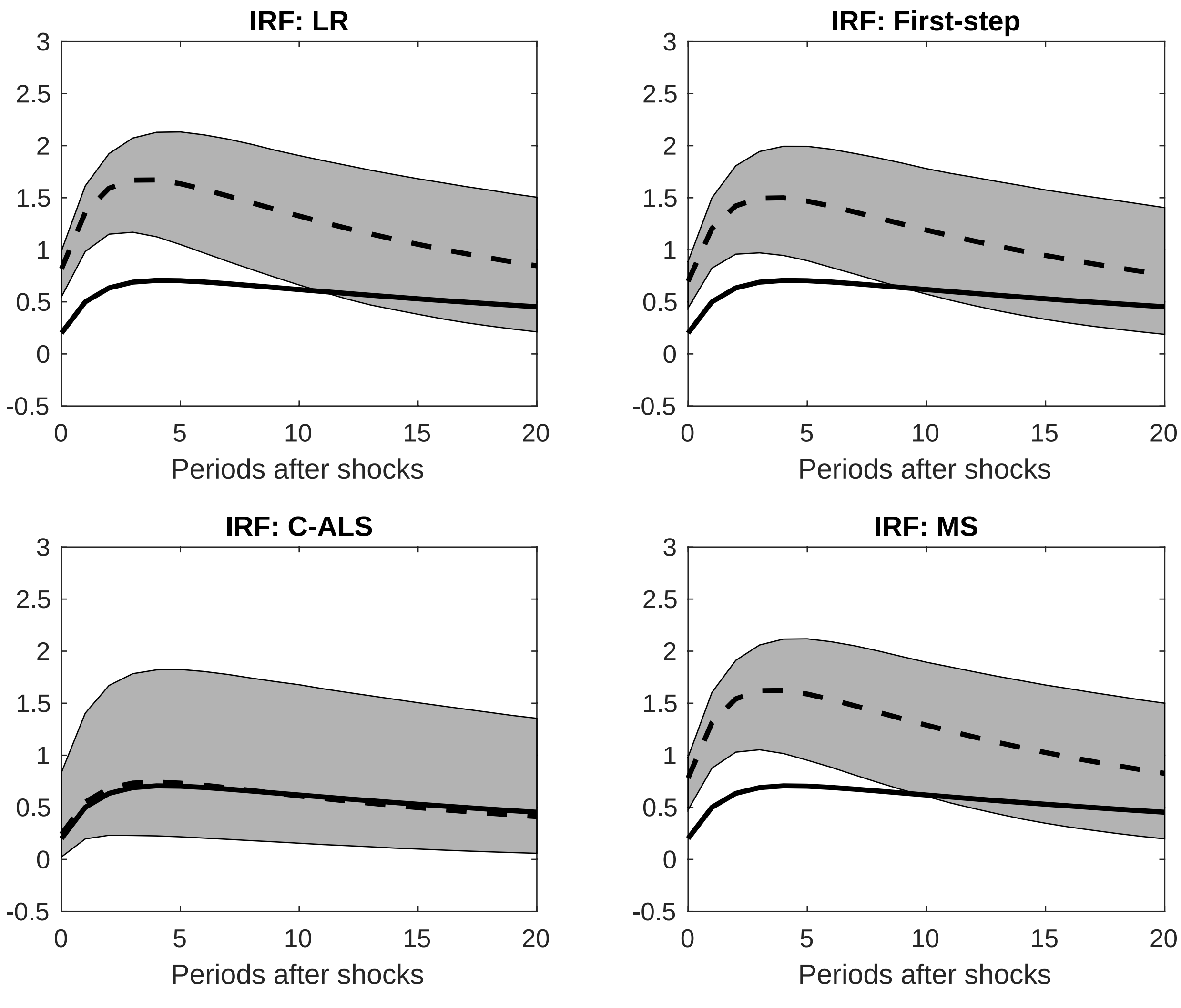
Note: The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

Figure 11: Cumulative Bias and RMSE up to 12 quarters using a VAR(2) model with $\rho = .95$ and $\delta = .04$



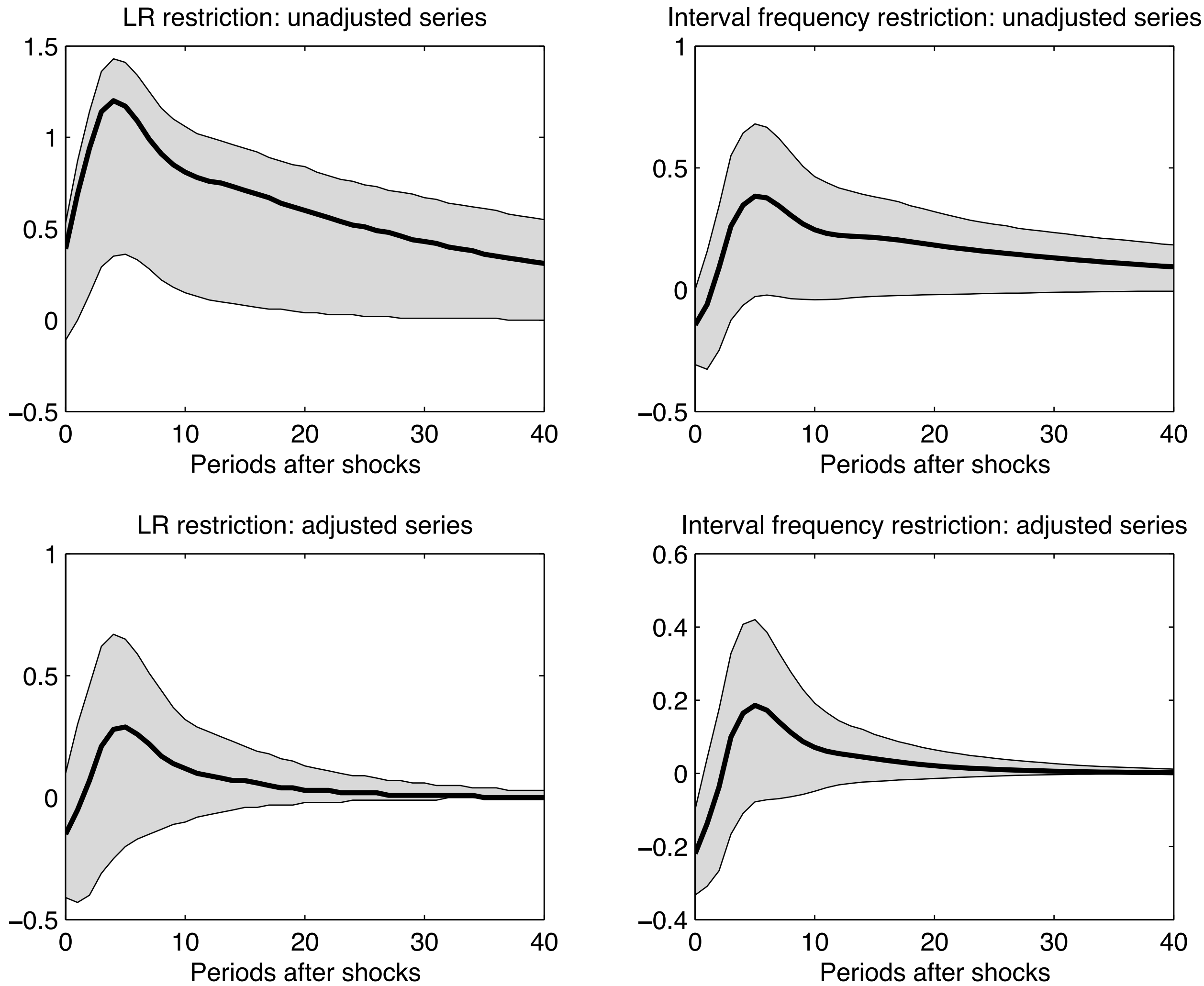
Note: The solid line, long dashed line, dash-dotted line, and dotted lines represent the LR, second-step C-ALS, Max-share and first-step estimators, respectively.

Figure 12: Impulse Responses for the first shock on second variable with $n = 60$, $\rho = .95$ and $\delta = .04$



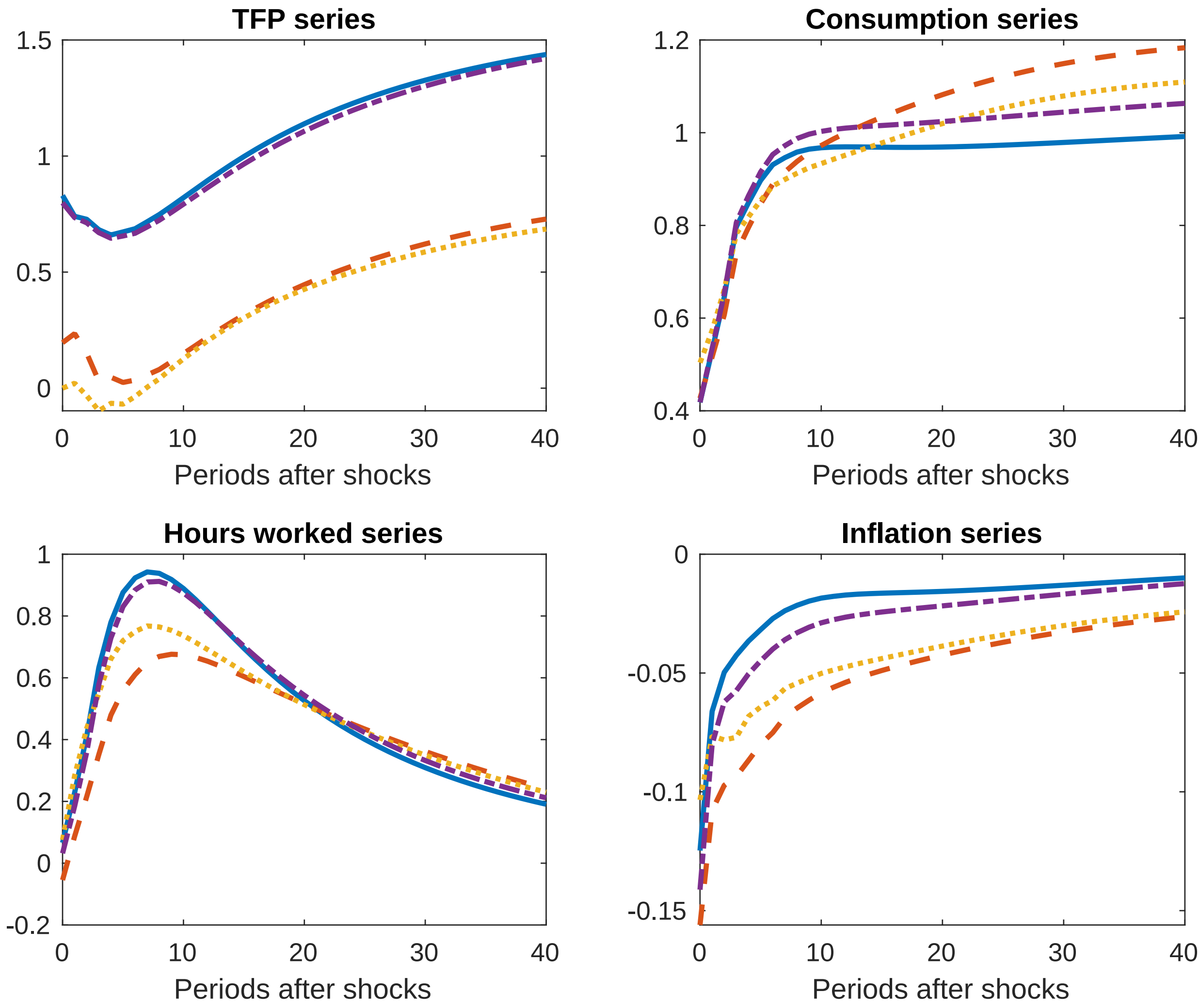
Note: Confidence intervals are based on the 95–percentile from 10,000 Monte–Carlo experiments.

Figure 13: Impulse responses for the technology shock on hours worked



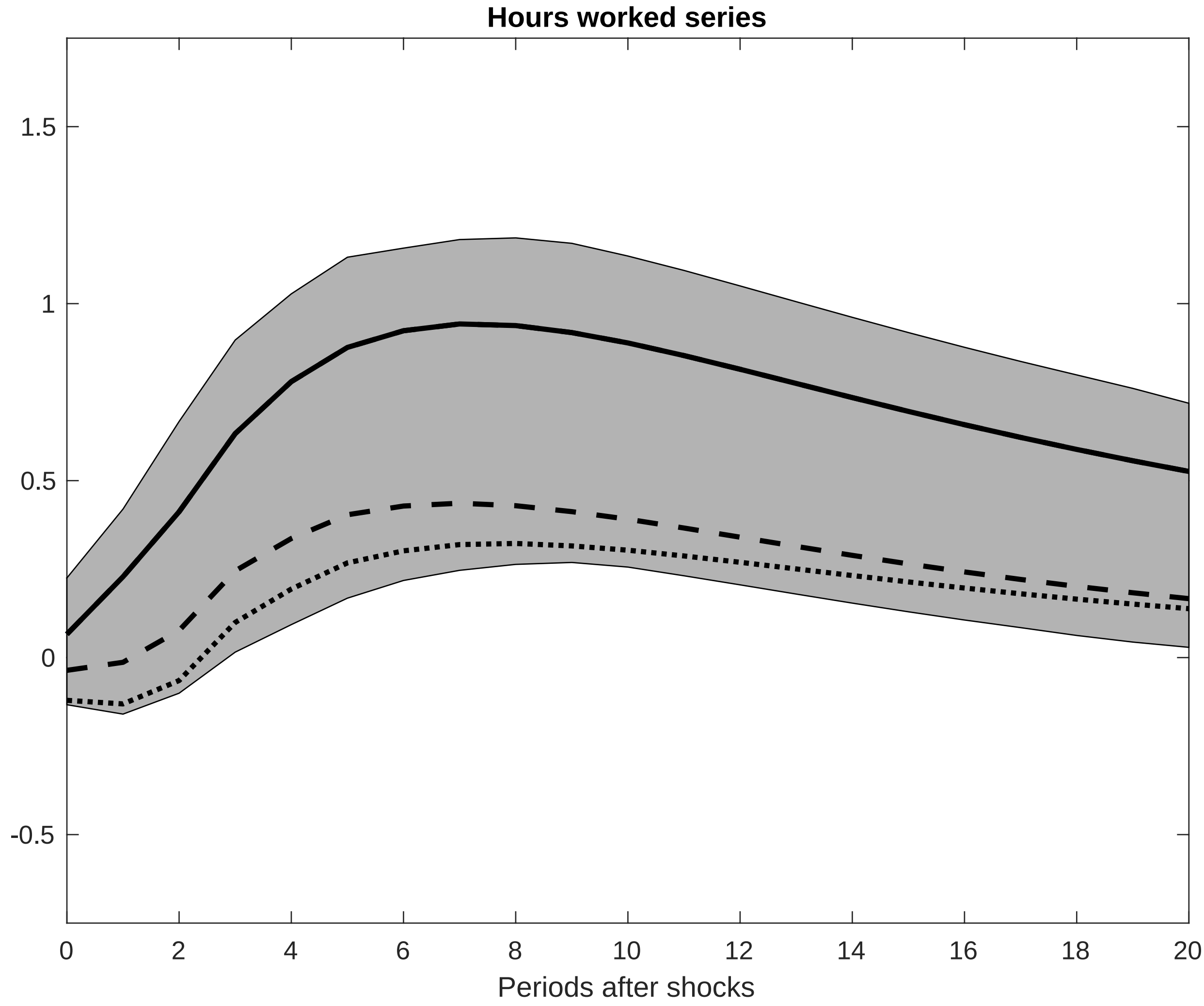
Note: Confidence intervals are based on the 95–percentile from 1000 bootstraps.

Figure 14: Impulse response functions for the news shock



Note: The solid line, dashed line, dash-dotted line, and dotted lines represent the second-step C-ALS, max-share of KS, BQ long-run, max-share BS, respectively.

Figure 15: Impulse response functions for the news shock on hours worked for different frequency intervals



Note: The solid line is for $\omega = \left(\frac{-\pi}{120}, \frac{\pi}{120}\right)$, the dotted line for $\omega = \left(\frac{-\pi}{3}, \frac{\pi}{3}\right)$ and the dashed line for $\omega = \left(\frac{-\pi}{20}, \frac{\pi}{20}\right)$. Confidence intervals are based on the 90–percentile from 1000 bootstraps.

Table 1: J-test

n	Quarters											
	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
	d. VAR(2) : $\rho = .95$											
	$\delta = 0$				$\delta = .02$				$\delta = .04$			
% 2nd struct. shock	.0213	.0150	.0103	.0073	.2201	.2846	.3139	.3297	.5101	.5773	.5966	.6083
.05	.1471	.1474	.1374	.2070	.4874	.5084	.5658	.6222	.9060	.9336	.9420	.9458
.10	.2018	.2079	.2119	.2758	.5553	.5850	.6388	.6830	.9298	.9518	.9574	.9601

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.

Supplementary Material: SVARs in the Frequency Domain using a Continuum of Restrictions

Alain Guay*

Florian Pelgrin†

This version: June 15, 2021

*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.

This supplementary material presents some elements regarding:

- The expression of the one-step and two-step C-ALS estimator using the autoregressive parameters of the reduced-form of a bivariate VAR representation;
- The derivation of a discretized ALS estimator (in the bivariate case);
- The comparison between the C-ALS estimator and the discretized ALS estimator;
- The unreliability issue of a long-run identification scheme.

1 Derivation of the C-ALS estimator using the autoregressive VAR parameters

Taking the reduced-form specification,

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + u_t$$

where $u_t = A_0 \epsilon_t$, one has

$$\Phi(L) X_t = u_t$$

where $\Phi(L) = I_2 - \phi_1 L - \cdots - \phi_p L^p$. Therefore the (estimate of) matrix of spectral densities of the structural VAR model at frequency ω is given by:

$$\widehat{f}_X^{\mathrm{FS}}(\omega) = \frac{1}{2\pi} \left[\overline{\xi(z)}' \xi(z) \right]^{-1}$$

where $z = \exp(-i\omega)$ and $\xi(z) = A_0^{-1} \widehat{\Phi}(z)$. Then, imposing that the second structural shock has no impact on the first variable over a given frequency interval, say $[\underline{\omega}, \overline{\omega}]$, is equivalent to say that the $(1,2)$-element of ξ, defined by:

$$\begin{aligned}
\xi(z) &= \frac{1}{\det(A_0)} \begin{pmatrix} a_{22}(0) & -a_{12}(0) \\ -a_{21}(0) & a_{11}(0) \end{pmatrix} \begin{pmatrix} \widehat{\Phi}_{11}(z) & \widehat{\Phi}_{12}(z) \\ \widehat{\Phi}_{21}(z) & \widehat{\Phi}_{22}(z) \end{pmatrix} \\
&= \frac{1}{\det(A_0)} \begin{pmatrix} a_{22}(0)\widehat{\Phi}_{11}(z) - a_{12}(0)\widehat{\Phi}_{21}(z) & a_{22}(0)\widehat{\Phi}_{12}(z) - a_{12}(0)\widehat{\Phi}_{22}(z) \\ a_{11}(0)\widehat{\Phi}_{21}(z) - a_{21}(0)\widehat{\Phi}_{11}(z) & a_{11}(0)\widehat{\Phi}_{22}(z) - a_{21}(0)\widehat{\Phi}_{12}(z) \end{pmatrix},
\end{aligned}$$

equals zero on this frequency interval. Accordingly, the identifying restrictions are defined by:

$$a_{22}(0)\widehat{\Phi}_{12}(z) - a_{12}(0)\widehat{\Phi}_{22}(z) = 0,$$

which is equivalent to (using the structural VMA representation)

$$\widehat{c}_{11}(z)a_{12}(0) + \widehat{c}_{12}(z)a_{22}(0) = 0 \quad \forall \omega \in [\underline{\omega}, \overline{\omega}]\,.$$

Using the identity operator as a kernel operator and minimizing the corresponding objective function, it is straightforward to show that the first-step consistent estimator of $\tilde{a}_{12}^0$ is given by:

$$\widehat{\tilde{a}}_{12,T}^1 = \frac{\sum_{k=0}^{p}\sum_{j=0}^{p} \widehat{\Phi}_{12,k}\widehat{\Phi}_{22,j} \int_{\underline{\omega}}^{\overline{\omega}} \cos(\omega(k-j))d\omega}{\sum_{k=0}^{p}\sum_{j=0}^{p} \widehat{\Phi}_{22,k}\widehat{\Phi}_{22,j} \int_{\underline{\omega}}^{\overline{\omega}} \cos(\omega(k-j))d\omega}.$$

where $\widehat{\Phi}_{22,0} = 1$, $\widehat{\Phi}_{22,j} = -\widehat{\phi}_{22,j}$, $\widehat{\Phi}_{12,0} = 0$, and $\widehat{\Phi}_{12,j} = -\widehat{\phi}_{12,j}$ for $j \geq 1$.

Finally, the optimal C-ALS estimator can be expressed as follows:

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

where $\widehat{A}_T$, $\widehat{B}_T$ and $\widehat{W}_T$ are given by:

$$\widehat{A}_T = -\sum_{k=0}^{p}\sum_{j=0}^{p}\left[\left(\left(\frac{\partial\widehat{\Phi}_{22,k}}{\partial\beta'}\widehat{\tilde{a}}_{12,T}^1 - \frac{\partial\widehat{\Phi}_{12,k}}{\partial\beta'}\right)\widehat{\Omega}^{1/2}\right)'\widehat{\Phi}_{22,j}\right]\int_{\underline{\omega}}^{\overline{\omega}} \cos((k-j)\omega)d\omega$$

$$\widehat{B}_T = -\sum_{k=0}^{p}\sum_{j=0}^{p}\left[\left(\left(\frac{\partial\widehat{\Phi}_{22,k}}{\partial\beta'}\widehat{\tilde{a}}_{12,T}^1 - \frac{\partial\widehat{\Phi}_{12,k}}{\partial\beta'}\right)\widehat{\Omega}^{1/2}\right)'\widehat{\Phi}_{12,j}\right]\int_{\underline{\omega}}^{\overline{\omega}} \cos((k-j)\omega)d\omega,$$

and

$$\widehat{W}_T = \sum_{k=0}^{p}\sum_{j=0}^{p}\left[\left(\left(\frac{\partial\widehat{\Phi}_{22,k}}{\partial\beta'}\widehat{\tilde{a}}_{12,T}^1 - \frac{\partial\widehat{\Phi}_{12,k}}{\partial\beta'}\right)\widehat{\Omega}^{1/2}\right)'\left(\left(\frac{\partial\widehat{\Phi}_{22,j}}{\partial\beta'}\widehat{\tilde{a}}_{12,T}^1 - \frac{\partial\widehat{\Phi}_{12,j}}{\partial\beta'}\right)\widehat{\Omega}^{1/2}\right)\right]\int_{\underline{\omega}}^{\overline{\omega}} \cos((k-j)\omega)d\omega.$$

2 Asymptotic least squares estimator using a discretization of the frequency interval

One possible approach is to apply the standard asymptotic least squares procedure using a discretization of the frequency band, and thus evaluating $g(a_0, \widehat{\beta}_T, \omega_\tau) = 0$ at different points/frequencies, say for $\tau = 1, \cdots, n$.

Proposition 2.1. *Consider a discretization of the frequency band*

$$\underline{\omega} = \omega_1 < \omega_2 < \cdots < \omega_n = \overline{w}.$$

Suppose that (X_t) is described by a bivariate VAR(p) model and that the identifying restriction is given by:

$$\widehat{c}_{11}(e^{-i\omega})\tilde{a}_{12}(0) + \widehat{c}_{12}(e^{-i\omega}) = 0.$$

Then, the first-step discretized ALS estimator, denoted $\widehat{\widetilde{a}}_{12}^{1,d}$*, is:*

$$\widehat{\widetilde{a}}_{12}^{1,d} = -\frac{\sum\limits_{k=0}^{\infty}\sum\limits_{j=0}^{\infty}\left(\widehat{c}_{11,k}\widehat{c}_{12,j}\sum\limits_{i=1}^{n}\cos(\omega_i(k-j))\right)}{\sum\limits_{k=0}^{\infty}\sum\limits_{j=0}^{\infty}\left(\widehat{c}_{11,k}\widehat{c}_{11,j}\sum\limits_{i=1}^{n}\cos(\omega_i(k-j))\right)},$$

and the second-step discretized ALS-estimator, $\widehat{\widetilde{a}}_{12}^{2,d}$*, is:*

$$\widehat{\widetilde{a}}_{12}^{2,d} = -\frac{\sum\limits_{k=0}^{\infty}\sum\limits_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{12,j}\Lambda'(\omega_{1:n},k)\left[\sum\limits_{k=0}^{\infty}\sum\limits_{j=0}^{\infty}\alpha_{k,j}\left(\widehat{\widetilde{a}}_{12}^{1,d}\right)\Lambda^{\star}(\omega_{1:n},k,j)\right]^{-1}\Lambda(\omega_{1:n},j)}{\sum\limits_{k=0}^{\infty}\sum\limits_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{11,j}\Lambda'(\omega_{1:n},k)\left[\sum\limits_{k=0}^{\infty}\sum\limits_{j=0}^{\infty}\alpha_{k,j}\left(\widehat{\widetilde{a}}_{12}^{1,d}\right)\Lambda^{\star}(\omega_{1:n},k,j)\right]^{-1}\Lambda(\omega_{1:n},j)}.$$

where $\alpha_{k,j}\left(\widehat{\widetilde{a}}_{12}^{1,d}\right) = \left(\frac{\partial \widehat{c}_{11,k}}{\partial \beta'}\widehat{\widetilde{a}}_{12}^{1,d} + \frac{\partial \widehat{c}_{12,k}}{\partial \beta'}\right)\widehat{\Omega}_T\left(\frac{\partial \widehat{c}_{11,j}}{\partial \beta}\widehat{\widetilde{a}}_{12}^{1,d} + \frac{\partial \widehat{c}_{12,j}}{\partial \beta}\right)$ *is a scalar,* $\Lambda^{\star}(\omega_{1:n},k,j) := \Lambda(\omega_{1:n},k)\Lambda'(\omega_{1:n},j)$*, and* $\Lambda(\omega_{1:n},k) = \begin{pmatrix}\cos(\omega_1 k) & \cdots & \cos(\omega_n k) & \sin(\omega_1 k) & \cdots & \sin(\omega_n k)\end{pmatrix}'$.

Proof: Following the approach of Feuerverger and McDunnough (1981), Singleton (2001) and Chacko and Viceira (2003), we distinguish the real part and the imaginary part of the identifying restrictions:

$$\sum_{k=0}^{\infty}\widehat{c}_{11,k}\cos(\omega_i k)\tilde{a}_{12} + \sum_{k=0}^{\infty}\widehat{c}_{12,k}\cos(\omega_i k) = 0$$
$$\sum_{k=0}^{\infty}\widehat{c}_{11,k}\sin(\omega_i k)\tilde{a}_{12} + \sum_{k=0}^{\infty}\widehat{c}_{12,k}\sin(\omega_i k) = 0.$$

pour $i = 1, \cdots, n$. Accordingly, the moment conditions are given by:

$$g(\tilde{a}_{12}(0), \widehat{\beta}_T, \omega_{1:n}) = \begin{pmatrix} g_1(\tilde{a}_{12}(0), \widehat{\beta}_T, \omega_{1:n}) \\ g_2(\tilde{a}_{12}(0), \widehat{\beta}_T, \omega_{1:n}) \end{pmatrix}$$

where

$$g_1(\tilde{a}_{12}(0), \widehat{\beta}_T, \omega_{1:n}) = \begin{pmatrix} \sum\limits_{k=0}^{\infty}\widehat{c}_{11,k}\cos(\omega_1 k)\tilde{a}_{12}(0) + \sum\limits_{k=0}^{\infty}\widehat{c}_{12,k}\cos(\omega_1 k) \\ \vdots \\ \sum\limits_{k=0}^{\infty}\widehat{c}_{11,k}\cos(\omega_n k)\tilde{a}_{12}(0) + \sum\limits_{k=0}^{\infty}\widehat{c}_{12,k}\cos(\omega_n k) \end{pmatrix}$$

and

$$g_2(\tilde{a}_{12}(0), \widehat{\beta}_T, \omega_{1:n}) = \begin{pmatrix} \sum\limits_{k=0}^{\infty}\widehat{c}_{11,k}\sin(\omega_1 k)\tilde{a}_{12}(0) + \sum\limits_{k=0}^{\infty}\widehat{c}_{12,k}\sin(\omega_1 k) \\ \vdots \\ \sum\limits_{k=0}^{\infty}\widehat{c}_{11,k}\sin(\omega_n k)\tilde{a}_{12}(0) + \sum\limits_{k=0}^{\infty}\widehat{c}_{12,k}\sin(\omega_n k) \end{pmatrix}.$$

A first-step consistent estimator of $\tilde{a}_{12}(0)$ solves the following minimization problem (using the identity matrix of order $2n$):

$$\widehat{\tilde{a}}_{12}^{1} = \underset{\tilde{a}_{12}}{\operatorname{argmin}} \quad g'(\tilde{a}_{12}, \widehat{\beta}_T, \omega) g(\tilde{a}_{12}, \widehat{\beta}_T, \omega)$$

or

$$\widehat{\tilde{a}}_{12,T}^{1,\mathbf{d}} = \underset{\tilde{a}_{12}}{\operatorname{argmin}} \quad \left\{ \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \left[\left(\widehat{c}_{11,k}\widehat{c}_{11,j}\tilde{a}_{12}^2 + 2\widehat{c}_{11,k}\widehat{c}_{12,j}\tilde{a}_{12} + \widehat{c}_{12,k}\widehat{c}_{12,j} \right) \sum_{i=1}^{n} \cos(\omega_i(k-j)) \right] \right\}.$$

Therefore,

$$\widehat{\tilde{a}}_{12}^{1,\mathbf{d}} = -\frac{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \left(\widehat{c}_{11,k}\widehat{c}_{12,j} \sum_{i=1}^{n} \cos(\omega_i(k-j)) \right)}{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \left(\widehat{c}_{11,k}\widehat{c}_{11,j} \sum_{i=1}^{n} \cos(\omega_i(k-j)) \right)}.$$

Accordingly, the (discretized) second-step ALS estimator, denoted $\widehat{\tilde{a}}_{12}^{\mathbf{d}}$, solves:

$$\widehat{\tilde{a}}_{12}^{\mathrm{d}} = \underset{\tilde{a}_{12}}{\operatorname{argmin}} \quad g'\left(\tilde{a}_{12}, \widehat{\beta}_T, \omega\right) S_0^{-1}(\widehat{\tilde{a}}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n}) g\left(\tilde{a}_{12}, \widehat{\beta}_T, \omega\right)$$

where $S_0^{-1}(\widehat{\tilde{a}}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})$ is the $2n \times 2n$ efficient weighting matrix defined by:

$$S_0^{-1}(\widehat{\tilde{a}}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n}) : = \left[\frac{\partial g(\widehat{\tilde{a}}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial \beta'} \widehat{\Omega}_T \frac{\partial g'(\widehat{\tilde{a}}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial \beta} \right]^{-1}$$

with

$$\begin{aligned}
\frac{\partial g(\widehat{\tilde{a}}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial \beta'} \widehat{\Omega}_T \frac{\partial g'(\widehat{\tilde{a}}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial \beta} &= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \Lambda(\omega_{1:n}, k) \left(\frac{\partial \widehat{c}_{11,k}}{\partial \beta'} \widehat{\tilde{a}}_{12}^{1,\mathrm{d}} + \frac{\partial \widehat{c}_{12,k}}{\partial \beta'} \right) \widehat{\Omega}_T \left(\frac{\partial \widehat{c}_{11,j}}{\partial \beta} \widehat{\tilde{a}}_{12}^{1,\mathrm{d}} + \frac{\partial \widehat{c}_{12,j}}{\partial \beta} \right) \Lambda'(\omega_{1:n}, j \\
&= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \left(\frac{\partial \widehat{c}_{11,k}}{\partial \beta'} \widehat{\tilde{a}}_{12}^{1,\mathrm{d}} + \frac{\partial \widehat{c}_{12,k}}{\partial \beta'} \right) \widehat{\Omega}_T \left(\frac{\partial \widehat{c}_{11,j}}{\partial \beta} \widehat{\tilde{a}}_{12}^{1,\mathrm{d}} + \frac{\partial \widehat{c}_{12,j}}{\partial \beta} \right) \Lambda^\star(\omega_{1:n}, k, j) \\
&= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \alpha_{k,j} \left(\widehat{\tilde{a}}_{12}^{1,\mathrm{d}} \right) \Lambda^\star(\omega_{1:n}, k, j)
\end{aligned}$$

where

$$\Lambda^\star(\omega_{1:n}, k, j) := \left(\begin{array}{ccc|ccc}
\cos(\omega_1 k)\cos(\omega_1 j) & \cdots & \cos(\omega_1 k)\cos(\omega_n j) & \cos(\omega_1 k)\sin(\omega_1 j) & \cdots & \cos(\omega_1 k)\sin(\omega_n j) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\cos(\omega_n k)\cos(\omega_1 j) & \cdots & \cos(\omega_n k)\cos(\omega_n j) & \cos(\omega_n k)\sin(\omega_1 j) & \cdots & \cos(\omega_n k)\sin(\omega_n j) \\
\hline
\sin(\omega_1 k)\cos(\omega_1 j) & \cdots & \sin(\omega_1 k)\cos(\omega_n j) & \sin(\omega_1 k)\sin(\omega_1 j) & \cdots & \sin(\omega_1 k)\sin(\omega_n j) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\sin(\omega_n k)\cos(\omega_1 j) & \cdots & \sin(\omega_n k)\cos(\omega_n j) & \sin(\omega_n k)\sin(\omega_1 j) & \cdots & \sin(\omega_n k)\sin(\omega_n j)
\end{array} \right).$$

Note that the analytical expression of $\frac{\partial \widehat{c}_{11,j}}{\partial \beta}'$ and $\frac{\partial \widehat{c}_{12,j}}{\partial \beta'}$ is provided in Appendix 4. In addition, it also worth noting that the rank of the matrix $\Lambda^\star(\omega_{1:n}, k, j)$ is one.

Under suitable regularity conditions, the discretized second-step ALS estimator, denoted $\widehat{\widetilde{a}}_{12,T}^{2,\mathtt{d}}$, solves the first-order condition:

$$g'(\widehat{\widetilde{a}}_{12}^{2,\mathtt{d}}, \widehat{\beta}_T, \omega_{1:n}) S_0^{-1}(\widehat{\widetilde{a}}_{12}^{1,\mathtt{d}}, \widehat{\beta}_T, \omega_{1:n}) \frac{\partial g(\widehat{\widetilde{a}}_{12}^{2,\mathtt{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial \tilde{a}_{12}'} = 0$$

that is

$$\left[\sum_{k=0}^{\infty}\left\{\left(\widehat{c}_{11,k}\widehat{\widetilde{a}}_{12}^{2,\mathtt{d}} + \widehat{c}_{12,k}\right)\Lambda'(\omega_{1:n},k)\right\}\right] S_0^{-1}(\widehat{\widetilde{a}}_{12}^{1,\mathtt{d}}, \widehat{\beta}_T, \omega_{1:n}) \left[\sum_{j=0}^{\infty}\widehat{c}_{11,j}\Lambda(\omega_{1:n},j)\right] = 0.$$

Finally

$$\widehat{\widetilde{a}}_{12}^{2,\mathrm{d}} = -\frac{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{12,j}\Lambda'(\omega_{1:n},k)\left[\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\alpha_{k,j}\left(\widehat{\widetilde{a}}_{12}^{1,\mathrm{d}}\right)\Lambda^{\star}(\omega_{1:n},k,j)\right]^{-1}\Lambda(\omega_{1:n},j)}{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{11,j}\Lambda'(\omega_{1:n},k)\left[\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\alpha_{k,j}\left(\widehat{\widetilde{a}}_{12}^{1,\mathrm{d}}\right)\Lambda^{\star}(\omega_{1:n},k,j)\right]^{-1}\Lambda(\omega_{1:n},j)}.$$

3 The discretization issue

We now discuss the comparison between the C-ALS estimator and the discretized ALS estimator. In so doing, and for sake of tractability, we compare the first-step C-ALS estimator and the discretized first-step ALS estimator:

$$\widehat{\widetilde{a}}_{12}^{1,\text{c-als}} = -\frac{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{12,j}\int_{\omega_1}^{\omega_n}\cos(\omega(k-j))d\omega}{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{11,j}\int_{\omega_1}^{\omega_n}\cos(\omega(k-j))d\omega} \quad \text{and} \quad \widehat{\widetilde{a}}_{12}^{1,\mathrm{d}} = -\frac{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\left(\widehat{c}_{11,k}\widehat{c}_{12,j}\sum_{i=1}^{n}\cos(\omega_i(k-j))\right)}{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\left(\widehat{c}_{11,k}\widehat{c}_{11,j}\sum_{i=1}^{n}\cos(\omega_i(k-j))\right)}.$$

Whereas the first-step C-ALS makes use of weights given by $\int_{\omega_1}^{\omega_n}\cos(\omega(k-j))d\omega$ for all (j,k), the discretrized first-step ALS estimator uses $\sum_{i=1}^{n}\cos(\omega_i(k-j))$ with $\omega_1 \leq ... \leq \omega_n$. It can be easily seen that the weights $\int_{\omega_1}^{\omega_n}\cos(\omega(k-j)d\omega$ are bounded and belongs to $(-1,1)$ (assuming that $\omega_n - \omega_1 \in (-1,1)$ when $k=j$) whereas the sum $\sum_{i=1}^{n}\cos(\omega_i(k-j))$ belongs to $(-n,n)$. To simplify the analysis, we now assume that the increment of the discretization is constant, i.e. $\Delta = \frac{\omega_n-\omega_1}{n-1}$ and

$$\omega_l = \omega_1 + (\ell-1)\,\Delta$$

for $\ell = 1, \cdots, n$. Before comparing the two estimator, we first need an intermediate result.

Corollary 3.1. *Suppose that the increment of the discretization is constant, $\Delta = \frac{\omega_n-\omega_1}{n-1}$. Then, the first-step discretized ALS estimator is given by:*

$$\widehat{\widetilde{a}}_{12,T}^{1,\mathtt{d}} = -\frac{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{12,j}\cos\left(\frac{\omega_1+\omega_n}{n}(k-j)\right)\frac{\sin\left(\frac{\omega_n-\omega_1}{2}\frac{n}{n-1}(k-j)\right)}{\sin\left(\frac{\omega_n-\omega_1}{2}\frac{1}{n-1}(k-j)\right)}}{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{11,j}\cos\left(\frac{\omega_1+\omega_n}{n}(k-j)\right)\frac{\sin\left(\frac{\omega_n-\omega_1}{2}\frac{n}{n-1}(k-j)\right)}{\sin\left(\frac{\omega_n-\omega_1}{2}\frac{1}{n-1}(k-j)\right)}}$$

When $\omega_1 = -\omega_n$,

$$\widehat{\widehat{a}}^{1,d}_{12,T} = -\frac{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{12,j}\frac{\sin\left(\frac{n}{n-1}(k-j)\omega_n\right)}{\sin\left(\frac{1}{n-1}(k-j)\omega_n\right)}}{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\widehat{c}_{11,k}\widehat{c}_{11,j}\frac{\sin\left(\frac{n}{n-1}(k-j)\omega_n\right)}{\sin\left(\frac{1}{n-1}(k-j)\omega_n\right)}}$$

Proof: Using the Moivre representation and the notation $\xi = k - j$,

$$\begin{aligned}
\sum_{\ell=1}^{n}\exp(i\xi\omega_l) &= \sum_{\ell=1}^{n}\exp\left(i\xi\left(\omega_1+(\ell-1)\Delta\right)\right)\\
&= \exp(i\xi\omega_1)\sum_{\ell=0}^{n-1}\exp(i\xi\ell\Delta) = \exp(i\xi\omega_1)\frac{1-\exp(i\xi n\Delta)}{1-\exp(i\xi\Delta)}\\
&= \exp(i\xi\omega_1)\frac{\exp\left(in\xi\frac{\Delta}{2}\right)}{\exp\left(i\xi\frac{\Delta}{2}\right)}\frac{\left(\exp\left(in\xi\frac{\Delta}{2}\right)-\exp\left(-in\xi\frac{\Delta}{2}\right)\right)/2i}{\left(\exp\left(i\xi\frac{\Delta}{2}\right)-\exp\left(-i\xi\frac{\Delta}{2}\right)\right)/2i}\\
&= \exp\left(i\xi\left(\omega_1+(n-1)\frac{\Delta}{2}\right)\right)\frac{\sin\left(n\xi\frac{\Delta}{2}\right)}{\sin\left(\xi\frac{\Delta}{2}\right)}\\
&= \exp\left(i\xi\frac{\omega_1+\omega_n}{2}\right)\frac{\sin\left(\xi\frac{\omega_n-\omega_1}{2}\frac{n}{n-1}\right)}{\sin\left(\xi\frac{\omega_n-\omega_1}{2}\frac{1}{n-1}\right)}
\end{aligned}$$

Since $\sum_{\ell=1}^{n}\exp(i\xi\omega_l) = \sum_{\ell=1}^{n}\cos(\omega_\ell\xi) + i\sum_{\ell=1}^{n}\sin(\omega_\ell\xi)$, the result follows by identification with the previous expression, and one can also deduce the corresponding expression when $\omega_1 = -\omega_n$ (symmetric interval). Note that $\sum_{\ell=1}^{n}\cos(\omega_\ell(k-j)) = n$ when $k = j$.□

Using the previous corollary, we are now in a position to characterize the normalized error $\frac{1}{\Delta}\int_{\omega_1}^{\omega_n}\cos(\omega(k-j))d\omega - \sum_{\ell=1}^{n}\cos(\omega_\ell(k-j))$.

Corollary 3.2. *Suppose that the increment of the discretization is constant,* $\Delta = \frac{\omega_n-\omega_1}{n-1}$. *Then,*

$$\frac{1}{\Delta}\int_{\omega_1}^{\omega_n}\cos(\omega(k-j))d\omega - \sum_{\ell=1}^{n}\cos(\omega_\ell(k-j)) =$$

$$\begin{cases} -1 \quad \text{when } k=j \\ \cos\left(\frac{\omega_n+\omega_1}{2}(k-j)\right)\left[\frac{2}{k-j}\frac{1}{\Delta}\sin\left((n-1)\frac{\Delta}{2}(k-j)\right) - \frac{\sin\left(n\frac{\Delta}{2}(k-j)\right)}{\sin\left(\frac{\Delta}{2}(k-j)\right)}\right] \quad \text{when } k\neq j. \end{cases}$$

In the case of a symmetric interval (for $k \neq j$*), one has*

$$\frac{1}{\Delta}\int_{-\omega_n}^{\omega_n}\cos(\omega(k-j))d\omega - \sum_{\ell=1}^{n}\cos(\omega_\ell(k-j)) = \frac{2}{k-j}\frac{1}{\Delta}\sin\left(\omega_n(k-j)\right) - \frac{\sin\left(\frac{n}{n-1}(k-j)\omega_n\right)}{\sin\left(\frac{1}{n-1}(k-j)\omega_n\right)}.$$

Proof: Noting that

$$\int_{\omega_1}^{\omega_n}\cos(\omega(k-j))d\omega = \begin{cases} \omega_n-\omega_1 \quad \text{when } k=j \\ \frac{1}{k-j}\left(\sin\left(\omega_n(k-j)\right)-\sin\left(\omega_1(k-j)\right)\right) \quad \text{when } k\neq j \end{cases}$$

with $\sin(\omega_n(k-j)) - \sin(\omega_1(k-j)) = 2\sin\left(\frac{\omega_n-\omega_1}{2}(k-j)\right)\cos\left(\frac{\omega_n+\omega_1}{2}(k-j)\right)$, and, in the case of a symmetric interval,

$$\int_{-\omega_n}^{\omega_n} \cos(\omega(k-j))d\omega = \begin{cases} 2\omega_n & \text{when } k=j \\ \frac{2}{k-j}\sin(\omega_n(k-j)) & \text{when } k\neq j \end{cases},$$

the result follows by virtue of Corollary 3.1.□

Using a simple trapezoidal rule for integration,

$$\begin{aligned} \int_{\omega_1}^{\omega_n} \cos(\omega(k-j)d\omega &= \Delta\sum_{i=1}^{n}\cos(\omega_i(k-j)) - \Delta\frac{\cos(\omega_n(k-j)) + \cos(\omega_1(k-j))}{2} + \mathtt{R}_n \\ &= \Delta\sum_{i=1}^{n}\cos(\omega_i(k-j)) - \Delta\cos\left(\left(\frac{\omega_n+\omega_1}{2}\right)(k-j)\right)\cos\left(\left(\frac{\omega_n-\omega_1}{2}\right)(k-j)\right) + \mathtt{R}_n \end{aligned}$$

where the reminder $\mathtt{R}_n$ is negligible for n large enough, it turns out that an approximation of the normalized error when $\omega_1 = -\omega_n$ is given by :

$$\frac{1}{\Delta}\int_{-\omega_n}^{\omega_n}\cos(\omega(k-j))d\omega - \sum_{\ell=1}^{n}\cos(\omega_\ell(k-j)) \simeq \cos(\omega_n(k-j)).$$

In the technical Appendix, Figure 1 reports the normalized error and its approximation.

From a numerical point of view, using the integral weights will reduce the round-off errors and some instability with respect to the first-step discretized ALS estimator. This argument is also true when comparing the two-step C-ALS and the two-step discretized ALS estimators. On top of the approximation of the integral by a sum (up to a constant term—inverse of the increment), one key issue is the existence of the inverse in the expression of $S_0^{-1}(.)$ (see above) :

$$\left(\frac{\partial g(\widehat{a}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial\beta'}\widehat{\Omega}_T\frac{\partial g'(\widehat{a}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial\beta}\right)^{-1}$$

and thus the rank of the matrix $\frac{\partial g(\widehat{a}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial\beta'}\widehat{\Omega}_T\frac{\partial g'(\widehat{a}_{12}^{1,\mathrm{d}}, \widehat{\beta}_T, \omega_{1:n})}{\partial\beta}$. This matrix is a linear combination of the matrices $\Lambda^*(w_{1:n}, k, j)$ for all k, j, which are singular—the rank of each matrix $\Lambda^*(w_{1:n}, k, j)$ being one. It turns out that for large n and thus a refined grid, the matrix is not invertible. Finally, as one refines and extends the grid, i.e. $\Delta \to 0$, the discrete set of estimating equations converges to the continuous estimating function, i.e. the identifying restriction on a frequency band, while the optimal weighting matrix will now converge to the covariance operator associated with that estimating function.

4 Unreliability of long-run identification scheme

We discuss the unreliability of the long-run identification scheme. It can be explained from these two fundamental relationships:

$$C(1)A(0) = A(1), \tag{4.1}$$

and

$$A(L) = C(L)A(0) = C(L)C(1)^{-1}A(1),. \tag{4.2}$$

It turns out that the long-run identification scheme conducts to reliable inference if and only if the $A(1)$ is consistently estimated in finite samples and especially the lag order p is not misspecified. Otherwise, any inconsistent estimate of $A(1)$ leads to unreliable long-run effects of shocks (in finite samples). This in turn is transferred to the estimates of the dynamic multipliers of the structural shocks by virtue of Eq. (4.2).[1] In particular, one cannot form asymptotically correct confidence intervals for impulse responses of each structural shock and there is no consistent test that an individual impulse response coefficient is zero (Faust and Leeper, 1997). The fundamental issue is that the true data generating process may have an infinite-ordered VAR representation with $\Phi_0(L) = \sum_{j=1}^{\infty} \Phi_{0,j} L^j$ and thus the infinite sequence $\Phi_0 = \{\Phi_{0,1}, \Phi_{0,2}, \cdots\}$ must be approximated by a finite sequence $\widetilde{\Phi}_p = \{\Phi_1, \cdots, \Phi_p\}$ (i.e., a misspecified VAR model). Such finite-parameter approximations to infinite lag distributions have been studied extensively by Sims (1971, 1972) and Pötscher (2002), especially for least-squares criterion.[2] An accurate approximation from the point of view of least-squares fit does not imply an accurate approximation of the long run effect.[3] This means that convergence of the sequence $\tilde{\Phi}_p$ is not sufficient to guarantee the convergence of some functions of those parameters (Sims, 1971,1972; Pötscher, 2002) as pointwise convergence does not imply (locally) uniform convergence. More specifically, functions of a lag distribution (e.g., the sum of coefficients) are in general discontinuous with respect to the metric implied be least-squares estimation.[4] Say differently, the best least-squares approximation of Φ_0, Φ_p, might be arbitrarily close (w.r.t. L_2-norm) whereas $\Phi(1)$ and $\tilde{\Phi}_p(1)$ are arbitrarily far apart and thus converge to different limits. This stems also from the fact that the least-squares criterion at a single frequency admits a zero Lebesgue measure. From a practical point of view, it turns out that standard errors of estimates or the coefficient of determination might approach their optimum values in arbitrarily large samples while the estimated sum of coefficients remains arbitrarily far from their true values. Inference based on the sum of coefficients is then highly unreliable unless Φ is in fact contained in Φ_p, and not only close to it (Pötscher, 2002).[5]

[1] Using Monte Carlo simulations, Erceg et al. (2005) and Chari et al. (2008) study the extent of these small-sample estimation problems.

[2] A similar argument can be found in Christiano et al. (2006a).

[3] See Faust (1996,1999) for an application of this result to unit root tests and confidence intervals for points on spectrum.

[4] The functional $S_{\Delta X} \to S_{\Delta X}(0)$, with $S_{\Delta X}$ the spectrum of the stochastic process (ΔX_t), is highly discontinuous w.r.t. L_2-distance. This makes the problem fall into the category of ill-posed problem (Sims, 1972; Pötscher, 2002).

[5] Note that it might occur regardless of how large the sample size is.

References

[1] Chacko, G. and L. Viceira (2003), "Spectral GMM estimation of continuous-time processes", *Journal of Econometrics*, vol. 116, 259-292.

[2] Chari, V., Kehoe, P., and E. McGrattan (2008), "Are Structural VARs with long-run Restrictions Useful in Developing Business Cycle Theory?", *Journal of Monetary Economics*, vol. 55(8), 1337-1352.

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

[4] Erceg, Guerrieri and Gust (2005), "Can Long-run Restrictions Identify Technology Shocks?", *Journal of the European Economic Association*, vol. 3(6), 1237-1278.

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

[6] Feuerverger, A. and P. McDunnough (1981), "On some Fourier Methods for Inference", *J. R. Statist. Assoc.*, vol. 76, 379-387.

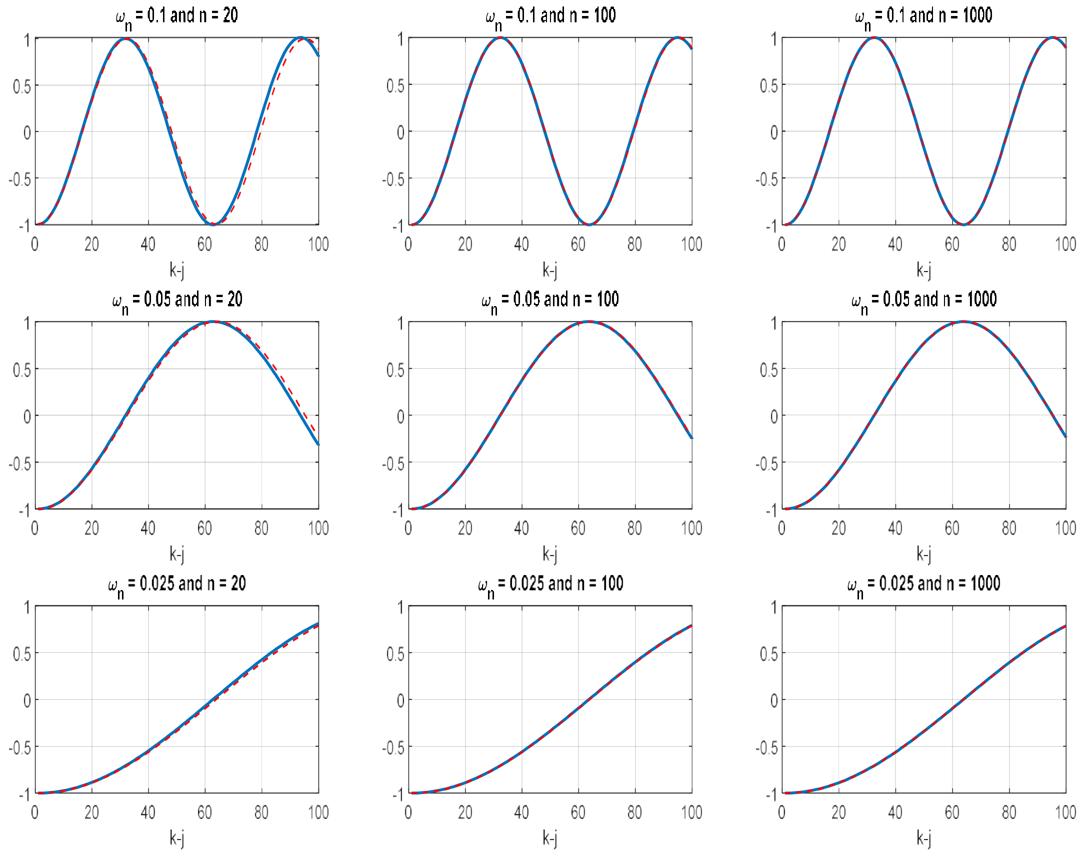
[7] Pötscher, B.M. (2002), "Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots, and Estimation of Long Memory Parameters", *Econometrica*, **70,** 1035-1065.

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

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

[10] Singleton, K. J. (2001), "Estimation of Affine Pricing Models Using the Empirical Characteristic Function", *Journal of Econometrics*, vol. 102, 111-141.

Figure 1: $\int_{-\omega_n}^{\omega_n} \cos(\omega(k-j))d\omega$ *versus* $\sum_{\ell=1}^{n} \cos(\omega_\ell(k-j))$



Note: The blue solid line and the red dotted line represent the normalized error $\frac{1}{\Delta}\int_{\omega_1}^{\omega_n} \cos(\omega(k-j))d\omega - \sum_{\ell=1}^{n} \cos(\omega_\ell(k-j))$ and the approximation of this error, $\cos(\omega_n(k-j))$.