Guay, Alain; Pelgrin, Florian; Surprenant, Stéphane

Working Paper

Max Share identification for structural VARs in levels: There is no free lunch!

Document de travail, No. 2024-06

Provided in Cooperation with:
Department of Economics, School of Management Sciences (ESG UQAM), University of Quebec in Montreal

Suggested Citation: Guay, Alain; Pelgrin, Florian; Surprenant, Stéphane (2024) : Max Share identification for structural VARs in levels: There is no free lunch!, Document de travail, No. 2024-06, Université du Québec à Montréal, École des sciences de la gestion (ESG UQAM), Département des sciences économiques, Montréal

This Version is available at:
https://hdl.handle.net/10419/311786

ESG RECHERCHE

DOCUMENT DE TRAVAIL / WORKING PAPER

No. 2024-06

Max Share Identification for Structural VARs in Levels: There Is No Free Lunch!

Alain Guay, Florian Pelgrin et Stéphane Surprenant

Novembre 2024

ESG UQÀM

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

Max Share Identification for Structural VARs in Levels: There Is No Free Lunch!

Alain Guay, Chaire en macroéconomie et prévisions ESG-UQAM, and CIREQ

Florian Pelgrin, EDHEC Business School

Stéphane Surprenant, Université du Québec à Montréal

Document de travail No. 2024-06
Novembre 2024

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

Max Share Identification for Structural VARs in Levels: There Is No Free Lunch![*]

Alain Guay[†]

Florian Pelgrin[‡]

Stéphane Surprenant[§]

This version: November 14, 2024

Abstract

This paper examines the implications of using VARs in levels under the Max Share identification approach when variables exhibit unit or near-unit roots. We derive the asymptotic distributions of the Max Share estimator, demonstrating that it converges to a random matrix, resulting in inconsistent reduced-form impulse responses and eigenvector estimates for structural shock identification. Monte Carlo simulations highlight that VAR models in levels can exhibit significant bias and higher RMSEs at intermediate and long horizons compared to stationary representations (e.g., first-difference VARs or VECMs), particularly in the presence of multiple permanent shocks. An empirical application focusing on investment-specific technology and TFP news shocks underscores the sensitivity of results to the nonstationarity of variables and the identification order of structural shocks when using VARs in levels.

Keywords: SVARs, Max Share identification and inference, Unit roots, Near-unit roots and Asymptotics.

JEL classification: E32, C32.

[*]The first author gratefully acknowledges financial support of SSHRC and FRQSC.

[†]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.

[§]Université du Québec à Montréal, *e-mail*: suprenant.stephane@uqam.ca.

1 Introduction

Structural VARs are now routinely applied in empirical macro research to assess and understand key mechanisms in macroeconomics, such as the impact of technology shocks or the primary drivers of business fluctuations. Building upon the seminal works of Sims (1980b; 1980a), moving from atheoretical/unrestricted VAR models to structural VAR models requires making identifying assumptions grounded in economic theory and related priors. This implies that VAR results cannot be interpreted independently of a more structural macroeconomic model (Cooley and Leroy, 1985; Bernanke, 1986).

Recent contributions have often concentrated on forecast error variance decompositions of some target variables, known as the *Max Share identification*, to pinpoint one structural shock (Faust, 1998; Uhlig, 2004) or multiple structural shocks (Ben Zeev and Khan, 2015; Carriero and Volpicella, 2024). For example, this approach identifies technology shocks as those explaining the most significant proportion of the forecast error variance decomposition of labor productivity over a 10-year period (Francis et al., 2014). Applications include identifying technology shocks (DiCecio and Owyang, 2010), news shocks (Barsky and Sims, 2011; Kurmann and Otrok, 2013; Forni et al., 2014; Kurmann and Sims, 2021; Bouakez and Kemoe, 2023; Kilian et al., 2024), neutral and investment-specific shocks (Chen and Wemy, 2015; Ben Zeev and Khan, 2015), credit shocks (Mumtaz et al., 2018), inflation target shocks (Mumtaz and Theodoridis, 2023), sentiment shocks (Fève and Guay, 2014; Levchenko and Pandalai-Nayar, 2020; Benhima and Cordonier, 2022) and a main business cycle shock (Angeletos et al., 2020).

A common practice in these contributions involves estimating unrestricted VARs *in levels* even when roots may be at or near unity. For instance, the structural identification of technological news shocks relies on a TFP (Total Factor Productivity) measure (e.g., Fernald, 2014), which inherently exhibits an exact unit root due to its construction.[1] Additionally, other macroeconomic variables of interest, such as the relative price of investment goods to consumption and real personal consumption expenditures per capita, may exhibit trending behavior and potentially near-unit root processes. Some of these variables may also be cointegrated, indicating the presence of common stochastic trends.

The rationale for specifying models *in levels* is that individual regression coefficients can be consistently estimated in any unrestricted VAR model *in levels*, regardless of the potential presence of unit roots and cointegration, as long as the model includes an intercept and sufficient lags, as shown by Sims et al. (1990). Kilian and Lütkepohl (2017) argue that the uncertainty regarding the presence of unit roots justifies this approach, as VAR models *in levels* encompass both integrated

[1] Starting from quarterly estimates of TFP growth (or the first-difference of the logarithm of TFP), the level (non-stationary) series can be derived, typically with initial levels normalized to zero.

VAR models and stationary models without a trend. Furthermore, uncertainty about the presence of unit roots in the variables and cointegration relationships between these unit root variables can lead to model misspecification and thus yield inconsistent estimates when pre-test procedures are used to transform some variables within the multivariate dynamic representation.

The application of the Max Share approach requires selecting a truncated forecast error variance horizon to capture short-to-medium or long-run cycles, which typically constitutes a substantial fraction of the sample size. For example, with quarterly observations spanning 60 years, a truncated horizon of 40, 60, or even 80 quarters represents a significant part of the sample size. Consequently, the rate at which the maximal truncated horizon increases relative to the sample size is crucial for both the small sample and asymptotic properties of the Max Share approach.

Unfortunately, combining the estimation of VARs *in levels* with a substantial horizon-to-sample size ratio can lead to undesirable statistical properties for impulse responses and forecast error variance decompositions. Phillips (1998) demonstrates that estimated impulse responses and forecast error decompositions are inconsistent at all horizons except the shortest ones when unit root processes or local-to-unity processes are present. These estimates tend to converge to random matrices rather than the true impulse responses, even though the individual autoregressive VAR parameters are estimated consistently (Sims et al., 1990).

The results of Phillips (1998) have several implications for the Max Share approach. Most notably, the Max Share identification relies on finding the largest eigenvalue(s) of the Max Share matrix derived from the forecast error variance decomposition, along with the associated eigenvector(s). Inconsistent estimates of the forecast error variance decomposition affect the eigendecomposition of this matrix, which, in turn, influences the distribution of the maximum eigenvalue and the corresponding eigenvector. The severity of this issue naturally varies with the forecast error variance horizon. Consequently, including nonstationary variables in unrestricted VARs *in levels* may result in the identification of a hybrid shock instead of a primitive shock, potentially causing a confounding effect.[2]

Therefore, using the Max Share approach involves a trade-off between using a *nonstationary* representation, like an unrestricted VAR *in levels* with some unit or near-unit roots, and a *stationary* representation, such as a VECM that accounts for common trends, or an unrestricted VAR with some first-differenced variables. In other words, there is no "free lunch"; estimating unrestricted VARs *in levels* can lead to inconsistent estimates of structural shocks and impulse responses. Conversely, estimating a VECM (or a VAR with differenced variables) may encounter misspecification

[2]See Dieppe et al. (2019) and Francis and Kindberg-Hanlon (2022) for discussions on confounding effects with the Max Share approach from other sources than the possible presence of unit root processes or local-to-unity processes.

issues arising from pre-test procedures.

It is therefore crucial to thoroughly understand the consequences of using variables *in levels* when applying the Max Share identification approach in the presence of unit-root or near-unit-root processes. This paper makes several key contributions: First, we derive the asymptotic results for the estimator of the Max Share matrix, including its eigenvalues and the corresponding eigenvectors, when (weakly) stationary variables are present. Notably, these asymptotic distributions have not been previously addressed in the literature. Second, building upon the seminal work of Phillips (1998), we derive the Max Share asymptotics when roots are at or near unity. Specifically, when the horizon represents a fixed fraction of the sample size, we show that the estimator of the Max Share matrix is inconsistent when the unrestricted VAR is estimated *in levels*, converging instead to a random matrix that represents a continuous average of a matrix quadratic form in the limiting (reduced-form) impulse responses. Consequently, the estimators of both the largest eigenvalue and the associated eigenvector are also inconsistent, converging towards a random variable/vector. We illustrate our results using a bivariate structural VAR in relevant cases for applied macroeconomics.

Third, we conduct Monte Carlo simulations using a flexible bivariate data-generating process (DGP) that accommodates a unit-root process, a highly persistent process, and a potential confounding effect involving two permanent structural shocks. Through these simulations, we compare the performance of a *stationary* representation, achieved by appropriately transforming the variables of the DGP, with a VAR *in levels*. We use OLS-based estimates for the stationary specification, along with OLS-based estimates, bias-corrected estimates (Pope, 1990), and bootstrapped estimates (Kilian, 1998; Inoue and Kilian, 2002) for the *level* specification, across various impulse response horizons and truncated forecast error variance horizons.[3]

In this respect, we highlight the following key insights:

1. **Structural impulse responses**: Structural impulse responses derived from VAR models in *levels* show a significant loss in terms of bias and RMSE properties at intermediate and long horizons compared to those from the stationary representation of VAR models (e.g., first-difference specification), despite performing similarly at (very) short horizons.

2. **Bias Correction and estimation methods**: Bias-corrected, bootstrap, or Bayesian methods can reduce the bias in OLS-based impulse response estimates in unrestricted VARs in *levels*. However, these methods may increase RMSEs and generally perform worse than estimates derived from a *stationary* representation (e.g., *first-difference* model).

3. **Confounding effects**: The presence of a potential confounding effect, such as two permanent

[3]Bayesian estimates using Minnesota priors and estimates from short-run identification are also available upon request.

shocks, further amplifies the discrepancies between *first-difference* estimates and *level*-based estimates.

Finally, we illustrate our theoretical and simulation results through an application identifying two permanent shocks, namely investment-specific technology and TFP news shocks, using the Max-share approach (see Fisher, 2006; Chen and Wemy, 2015; Ben Zeev and Khan, 2015; Kurmann and Sims, 2021). The results critically depend on the integration order of the variables, and thus on the chosen specification in levels or in first differences, as well as the identification order of structural shocks. Notably, in the specification using *level* variables, the impulse response functions differ substantially depending on whether the TFP measure or the relative price of consumption-to-investment is placed first. However, this discrepancy vanishes when stationary transformations of the variables are applied.

Both theoretical and empirical results underscore that the application of the Max-share approach using variables *in levels*, especially when some of the variables are characterized by unit or near-unit root processes (and possibly cointegration relationships), can adversely affect the identification of structural shocks and their corresponding impulse response functions. These issues are exacerbated when a long forecast error variance horizon is chosen and multiple permanent shocks are present. Therefore, it is strongly recommended to also report results using stationary transformations of the variables, such as with a VECM or a VAR with first-differenced variables.

The rest of the paper is organized as follows. Section 2 reviews notation and presents the Max-share identification strategy. Section 3 presents the asymptotic results in the presence of (weakly) stationary variables and extends the results to cases where unrestricted VAR models are estimated in levels and there are some roots at, or near, unity. Section 4 provides Monte Carlo simulations, while Section 6 discusses an empirical application regarding the identification of investment-specific and technology long-term shocks. The last section contains concluding comments and future extensions. Proofs are gathered in the Appendix.

2 Max share identification of structural VAR models

In this section, we first introduce preliminary notation and provide an overview of the Max Share approach.

2.1 Notation

Let $y_t = (y_{1,t}, \cdots, y_{N,t})'$ be a N-vector time series generated by the following pth order vector autoregressive model in *levels*:

$$y_t = \sum_{i=1}^{p} A_i y_{t-i} + u_t = A(L) y_{t-1} + u_t \tag{2.1}$$

where L is the lag operator, the $(N \times N)$ autoregressive matrices A_i are fixed, $u_t = (u_{1t}, \cdots, u_{Nt})'$ is a N-dimensional weak white noise with $\mathbb{E}[u_t] = 0_{N\times 1}$ and $\mathbb{E}[u_t u_t'] = \Sigma_u$. The reduced-form (2.1) is initialized at $t = -p + 1.....0$ and we let these initial values be any random vectors including constants.The presence of deterministic regressors does not affect our main results and thus we proceed without them to keep the derivations as simple as possible.

The reduced-form VAR can also be written in companion form as:

$$Y_t = AY_{t-1} + U_t \tag{2.2}$$

where $Y_t = \left(y_t', \ldots, y_{t-p}'\right)'$, $U_t = (u_t', 0, \ldots, 0)'$, and

$$A = \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I_K & 0 \end{bmatrix}. \tag{2.3}$$

Up to some initial conditions, the vector moving average (VMA) representation of the reduced-form VAR is then defined by:

$$Y_t = \sum_{i=0}^{t-1} A^i U_{t-i} \tag{2.4}$$

and one can retrieve the reduced-form `VMA` representation of y_t:

$$y_t = \sum_{i=0}^{t-1} \Phi_i u_{t-i}, \tag{2.5}$$

where Φ_0 is the identity matrix of order N, denoted by I_N, and $\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j$. Notably, equation 2.5 can be decomposed as follows:

$$y_t = \sum_{i=0}^{t-1} \Phi_i u_{t-i} = \sum_{i=0}^{i^*} \Phi_i u_{t-i} + \sum_{i=i^*+1}^{t-1} \Phi_i u_{t-i}. \tag{2.6}$$

This decomposition of the VMA representation is useful for studying the impulse response and forecast error variance asymptotics in nonstationary VARs (Phillips, 1998). In particular, for a

small fixed i^*, it is worth emphasizing that the estimates of the impulse responses matrices $\hat{\Phi}_i$ have asymptotic normal distributions (as in the stationary case) even in the presence of unit root or near-unit root processes. In contrast, the estimates of impulse-response matrices in the second term on the right-hand side, which are those associated to a lead time i that can be written formally as $i = fT$, where $f > 0$ is a fixed fraction of the sample, are inconsistent for all $i > i^\star$. More specifically, the limits are random variables with unit root or local-to-unity distributions.

The structural VAR model can be written as:

$$B_0 y_t = \sum_{i=1}^{p} B_i y_{t-i} + w_t = B(L) y_{t-1} + w_t \tag{2.7}$$

where w_t is an $N \times 1$ random vector of structural shocks with $\mathbb{E}[w_t] = 0$ and $\mathbb{E}[w_t w_t'] = \Gamma$. A common identification assumption is $\Gamma = I_N$. Taking equations (2.4) and (2.7), the error terms of the reduced-form model are a linear combination of structural shocks:[4]

$$u_t = B_0^{-1} w_t = A_0 w_t, \tag{2.8}$$

with $B_0^{-1} B(L) = A(L)$. The structural infinite VMA representation is then defined by:

$$y_t = \sum_{i=0}^{\infty} \Phi_i B_0^{-1} w_{t-i} = \sum_{i=0}^{\infty} \Theta_i w_{t-i}$$

where $\Theta_i = \Phi_i B_0^{-1} = \Phi_i A_0$. Using equation (2.8), one has $\Sigma_u = B_0^{-1} B_0^{-1\prime} = A_0 A_0'$. Let Σ_{tr} denote the lower triangular Cholesky decomposition of Σ_u (with the diagonal elements normalized to be positive), and let Q be a $N \times N$ orthogonal matrix. Since $Q'Q = QQ' = \mathrm{I}_N$ and hence $(\Sigma_{\text{tr}} Q)(\Sigma_{\text{tr}} Q)' = \Sigma_{\text{tr}} \Sigma_{\text{tr}}'$, the set of possible solutions for B_0^{-1} can be written as $\Sigma_{\text{tr}} Q$. Then identification involves pinning down some or all columns of Q.

Finally, Equation (2.1) can also be equivalently written in the levels and differences format as:

$$y_t \;=\; \Pi y_{t-1} + \Upsilon(L) \Delta y_{t-1} + u_t \tag{2.9}$$

where $\Pi = A(1)$, $\Upsilon = \sum_{i=1}^{p-1} \Upsilon_i L^{i-1}$ with $\Upsilon_i = -\sum_{m=i+1}^{p} A_m$. Assumptions regarding nonstationary components and the presence of cointegration (i.e., the dimension of the cointegrating space and the form of the cointegration vectors) will be specified in Assumption 3.2. Furthermore, the formulation (2.9) proves to be useful when deriving the asymptotic properties of the Max Share matrix estimator.

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

2.2 Max share approach

Starting with the seminal contributions of Faust (1998) and Uhlig (2004; 2003), the Max Share identification scheme focuses on maximizing the contribution of a (structural) shock to the forecast-error variance of a given variable at a long but finite horizon, say h. To illustrate it, consider the bivariate structural VAR model of Galí (1999) that attributes variation in U.S. labour productivity and hours worked to a technology shock and a non-technology shock. The first structural shock, labelled a technology shock, can be identified by maximizing its contribution to the forecast-error variance of labor productivity (Francis et al., 2014).

Using the `VMA` representation of the reduced-form `VAR`, the starting point is to define the h-step-ahead forecast error for y_t as a function of realized reduced-form errors:

$$y_{t+h} - \mathbb{E}_t\left[y_{t+h}\right] = \sum_{\ell=0}^{h-1} \Phi_\ell u_{t+h-\ell}. \tag{2.10}$$

Accordingly, the h-step-ahead forecast-error variance matrix is given by:

$$\texttt{MSE}(h) = \sum_{\ell=0}^{h-1} \Phi_\ell \Sigma_u \Phi_\ell' = \sum_{\ell=0}^{h-1} \Phi_\ell \Sigma_{\text{tr}} Q Q' \Sigma_{\text{tr}}' \Phi_\ell'. \tag{2.11}$$

Then the share of forecast-error variance of a given variable k that is attributed to a given shock j at horizon h is:

$$\tau_{kj}(h) = \frac{q_j' S_k(h) q_j}{e_k' \texttt{MSE}(h) e_k} = \frac{e_j' Q' S_k(h) Q e_j}{e_k' \texttt{MSE}(h) e_k} \tag{2.12}$$

where e_k is the k-th column vector of the identity matrix, $q_j = Q e_j$ is the j-th column of the orthogonal matrix Q, and $S_k(h)$ is the Max Share matrix at horizon h for the variable k:

$$S_k(h) = \sum_{\ell=0}^{h-1} \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k' \Phi_\ell \Sigma_{\text{tr}} = \sum_{\ell=0}^{h-1} \Psi_{k.,\ell}' \Psi_{k.,\ell} \tag{2.13}$$

with $\Psi_{k.,\ell} = e_k' \Phi_\ell \Sigma_{\text{tr}}$ the k-th row of $\Phi_\ell \Sigma_{\text{tr}}$. According to the decomposition of the VMA representation (2.6), the Max Share matrix depends not only on the impulse responses at short-run horizons but also on those at longer horizons when h constitutes a substantial fraction of the sample size.

We consider the first structural shock $j = 1$, which is identified by solving, for a given horizon h, the following maximization of the *Max Share statistic* with respect to q_1:

$$q_{1,k}(h) = \underset{q_1}{\texttt{argmax}} \quad \frac{q_1' S_k(h) q_1}{e_k' \texttt{MSE}(h) e_k} \tag{2.14}$$

subject to $q_1'q_1 = 1$.[5] Note that the solution, denoted as $q_{1,k}(h)$, depends on the choice of a truncated forecast error variance horizon h. Following Faust (1998) and Uhlig (2004; 2003), it can be shown that q_1 is the eigenvector associated to the largest eigenvalue of the Max Share matrix or, equivalently, is the first principal component:

$$S_k(h)q_{1,k}(h) = \lambda_{\max}q_{1,k}(h). \tag{2.15}$$

Thus, the structural IRFs from the identified shock are given by:

$$\Theta_{\cdot 1,i}(h) = \Phi_i\Sigma_{\text{tr}}q_{1,k}(h) \tag{2.16}$$

where $\Theta_{\cdot 1,i}$ is the first column of the impulse response matrix Θ_i. The identified shocks and the corresponding IFRs then depend on the finite sample and the asymptotic properties of the Max Share matrix for a given horizon h (i.e., $S_k(h)$), as well as the largest eigenvalue $\lambda_{\max}$ and the associated eigenvector $q_{1,k}$. As aforementioned and according to the VMA decomposition (2.6), the Max Share matrix depends not only on the impulse responses at short-run horizons but also on those at longer horizons when h constitutes a substantial fraction of the sample size. The next section examines the asymptotic properties of these elements.

As a final remark, other Max Share matrices have been considered in the literature. On the one hand, as stated in Uhlig (2003) and Barsky and Sims (2011), an *accumulated* Max Share approch can be employed, i.e., the (partial) sum of the contributions of a given structural shock to the forecast-error variance of a given variable between two finite horizons, say $\underline{h}$ and $\overline{h}$ (with $\overline{h} \geq \underline{h}$). Notably, the accumulated Max Share matrix, denoted by $S_k(\underline{h}, \overline{h})$, is then given by:

$$S_k(\underline{h}, \overline{h}) = \sum_{h=\underline{h}}^{\overline{h}} \frac{S_k(h)}{e_k'\texttt{MSE}(h)e_k} = \sum_{h=\underline{h}}^{\overline{h}} \frac{\sum_{\ell=0}^{h-1}\Psi_{k.,\ell}'\Psi_{k.,\ell}}{\sum_{\ell=0}^{h-1}\Psi_{k.,\ell}\Psi_{k.,\ell}'}.$$

In this expression, the weight decreases for $\ell = \underline{h}, \cdots, \overline{h}$, and thus the accumulated Max Share matrix places more weight on short horizons than long horizons. Similarly to the non-accumulated Max Share approach, the first structural shock $j = 1$ is identified by maximizing, for a given horizon interval $[\underline{h}; \overline{h}]$, the following Max Share statistics with respect to q_1:

$$q_1^{(k)} = \underset{q_1}{\texttt{argmax}} \quad q_1'\left(\sum_{h=\underline{h}}^{\overline{h}} \frac{S_k(h)}{e_k'\texttt{MSE}(h)e_k}\right)q_1 \tag{2.17}$$

subject to $q_1'q_1 = 1$.

On the other hand, building on the contribution of DiCecio and Owyang (2010), Francis et al.

[5] Without loss of generality, note that further structural constraints can be added, such as the absence of contemporaneous effect of a structural shock (e.g., Ben Zeev and Khan, 2015; Bouakez and Kemoe, 2023).

(2014), and Angeletos et al. (2020), the Max Share approach in the frequency domain aims to maximize the contribution of a given structural shock to the spectral density of a given variable over a frequency interval, say $[\underline{\omega};\overline{\omega}]$. Provided the multivariate spectral density, denoted by f, is well-defined $\forall\omega \in [-\pi;\pi]$, one has:

$$f(\omega) = \frac{1}{2\pi}\overline{\Phi\left(exp(-i\omega)\right)}\Sigma_u\Phi\left(exp(-i\omega)\right)$$

where $\Phi\left(exp(-i\omega) = \sum_{\ell=0}^{\infty}\Phi_\ell(\exp(-i\omega\ell)\right)$, and $\overline{\Phi}$ denotes the complex conjugate transpose of Φ. Therefore, the Max share statistics in the frequency domain, representing the contribution to the spectral density of a given variable k attributable to a given shock j, say $j = 1$, over a frequency band $[\underline{\omega};\overline{\omega}]$ is defined by:

$$\frac{q_1' S_k\left(\underline{\omega};\overline{\omega}\right) q_1}{e_k'(\underline{\omega},\overline{\omega})e_k}, \tag{2.18}$$

where the frequency Max Share matrix over the frequency band $[\underline{\omega};\overline{\omega}]$ is:

$$S_k\left(\underline{\omega};\overline{\omega}\right) = 2\mathtt{Re}\int_{\underline{\omega}}^{\overline{\omega}}\overline{\Psi_{k.}(\exp(-i\omega))}\Psi_{k.}(\exp(-i\omega))d\omega. \tag{2.19}$$

where $\Psi_{k.}(\exp(-i\omega)) = [\Phi(\exp(-i\omega))\Sigma_{\text{tr}}]_{k.} = \left[\sum_{\ell=0}^{\infty}\Phi_\ell(\exp(-i\omega\ell))\Sigma_{\text{tr}}\right]_{k.}$ and $\mathtt{Re}$ is the real part of any complex. The identification and interpretation of the first structural shock then follows the same procedure as in the case of the *non-accumulated* Max Share approach at a given horizon h. In the frequency domain, the Max Share matrix relies on the infinite sum of the (reduced-form) impulse responses irrespective of the frequency interval. A truncated sum may weaken the statistical performances in the presence of persistent stochastic processes.[6]

3 Max Share asymptotics

This section delves into the asymptotic properties of the Max Share estimator. We start by examining the asymptotic distribution of the estimator for the Max Share matrix in the context of a weakly stationary multivariate process y_t. These results are also valid for VAR models *in levels* with a fixed i in the VMA decomposition (2.6), i.e., for short horizons. Next, we characterize the asymptotic distribution of the (maximal) eigenvalues and their corresponding eigenvectors. We then provide the asymptotic distribution of the Max Share estimator in the presence of roots at or near unity. As highlighted by the decomposition (2.6), the small sample properties of the Max Share approach based on a VAR *in levels* stem from the finite-sample approximations associated with these two asymptotic behaviors.

[6]Note that the multivariate spectral density is no longer defined at $\omega = 0$ in the presence of unit or near-unit root processes.

3.1 Max Share asymptotics with weakly stationary processes

We suppose that:

Assumption 3.1.

(a) u_t is an i.i.d. process with zero mean, covariance matrix $\Sigma_u > 0$ and finite fourth cumulants;

(b) The determinantal equation $|I_K - \sum_{i=1}^{p} A_i L^i| = 0$ has roots outside the unit circle.

Following Lütkepohl (2007) and using the Delta method, the asymptotic distribution of the estimator of the Max Share matrix at a fixed horizon is then given in Theorem 3.1.

Theorem 3.1. *Let Assumption 3.1 hold and let $\tilde{\alpha} = vec([A_1, \ldots, A_p])$, $\sigma = vech(\Sigma_u)$, $Y = (Y_{p+1}, \ldots, Y_T) \in \mathbb{R}^{Np \times (T-p)}$, and $\Gamma_Y(0) := E\left[(Y - E(Y))(Y - E(Y))'\right]$. Then, as $T \to \infty$, the estimator of the Max Share matrix $S_k(h)$, denoted $\widehat{S}_{k,T}(h)$, at a fixed and finite forecast error variance horizon h is weakly consistent and is asymptotically normally distributed:*

$$\sqrt{T} vec\left(\widehat{S}_{k,T}(h) - S_k(h)\right) \xrightarrow{d} \mathcal{N}(0, \Omega(h))$$

with

$$\Omega(h) = \begin{bmatrix} D_{\tilde{\alpha}}(h) & D_{\sigma}(h) \end{bmatrix} \begin{bmatrix} \Gamma_Y(0)^{-1} \otimes \Sigma_u & 0 \\ 0 & 2D_N^{+}(\Sigma_u \otimes \Sigma_u) D_N^{+\prime} \end{bmatrix} \begin{bmatrix} D'_{\tilde{\alpha}}(h) \\ D'_{\sigma}(h) \end{bmatrix}$$

where $D_N^{+} := (D'_N D_N)^{-1} D'_N$ is the Moore-Penrose generalized inverse of an appropriate $N^2 \times N(N+1)/2$ duplication matrix, and the gradients $D_{\tilde{\alpha}}(h)$ and $D_{\sigma}(h)$ are defined in Appendix 1.

Proof: See Appendix 1.

This theorem can be readily adapted to apply to the *accumulated* Max Share approach (Uhlig, 2003; Barsky and Sims, 2011), which sums the contributions of the jth structural shock to the forecast error variance of the kth variable between two finite horizons. It can also be extended to the Max Share approach in the *frequency* domain (DiCecio and Owyang, 2010; Francis et al., 2014; Angeletos et al., 2020).[7]

We can now provide the asymptotic distribution of the eigenvalues of $S_k(h)$. We begin with the spectral decomposition of $S_k(h)$:

$$S_k(h) = Q(h)\Lambda(h)Q(h)', \tag{3.20}$$

[7]The online Appendix provides the results for the accumulated Max Share matrix and the frequency-based Max Share approach.

where $\Lambda(h)$ is the diagonal matrix associated with the N ordered eigenvalues $\lambda_i(h)$, $i = 1, \ldots, N$, and $Q(h)$ is the corresponding matrix of (orthonormal) eigenvectors for a given forecast error variance horizon h. By convention, we assume that the eigenvalues are always arranged in algebraically non-increasing order:

$$\lambda_{\max}(h) \equiv \lambda_1(h) \geq \lambda_2(h) \geq \cdots \geq \lambda_N(h) \equiv \lambda_{\min}(h).$$

Since $S_k(h)$ is not necessarily of full rank, suppose that the first r eigenvalues are different from zero, and thus the last $N - r$ eigenvalues are equal to zero. Accordingly, the orthonormal matrix $Q(h)$ can be partitioned as $Q(h) = \begin{bmatrix} Q_r(h) & Q_{N-r}(h) \end{bmatrix}$, where $Q_r(h) = \begin{bmatrix} q_1(h) & Q_{2:r}(h) \end{bmatrix}$, with $q_1(h)$ being the eigenvector associated with $\lambda_{\max}(h)$, $Q_{2:r}(h)$ the matrix of eigenvectors associated with $\lambda_2(h), \ldots, \lambda_r(h)$, and $Q_{N-r}(h)$ the matrix of eigenvectors associated with the $N - r$ smallest eigenvalues $\lambda_{r+1}(h), \ldots, \lambda_{\min}(h)$. Notably,

$$\begin{aligned} \lambda_{\max}(h) &= q_{1,k}'(h) S_k(h) q_{1,k}(h), \\ \lambda_{2:r}(h) &= \operatorname{vec}\left(Q_{2:r}'(h) S_k(h) Q_{2:r}(h)\right). \end{aligned}$$

Combining this decomposition with Theorem 3.1, the asymptotic distribution of the eigenvalues follows.

Theorem 3.2. *Let Assumption 3.1 hold. Then, the (ordered) eigenvalue estimators $\hat{\lambda}_{i,T}(h)$, which solve the spectral decomposition (equation 3.20) for $\widehat{S}_{k,T}(h)$, are weakly consistent estimators of $\lambda_i(h)$, $i = 1, \ldots, N$. Furthermore, the asymptotic distribution of $\widehat{\lambda}_{\max,T}(h)$ at a fixed and finite forecast error variance horizon h is given by:*

$$\sqrt{T}\left(\widehat{\lambda}_{\max,T}(h) - \lambda_{\max}(h)\right) \xrightarrow{d} N\left(0, (q_{1,k}(h) \otimes q_{1,k}(h))' \Omega(h) (q_{1,k}(h) \otimes q_{1,k}(h))\right),$$

where $\Omega(h)$ is the asymptotic variance-covariance matrix of $\sqrt{T} vec\left(\widehat{S}_{k,T}(h) - S_k(h)\right)$ as given in Theorem 3.1. Additionally, the asymptotic distribution of $\widehat{\lambda}_{2:r,T} = \left(\widehat{\lambda}_{2,T}(h), \cdots, \widehat{\lambda}_{r,T}(h)\right)'$ is:

$$\sqrt{T}\left(\widehat{\lambda}_{2:r,T}(h) - \lambda_{2:r}(h)\right) \xrightarrow{d} N\left(0, (Q_{2:r}(h) \otimes Q_{2:r}(h))' \Omega(h) (Q_{2:r}(h) \otimes Q_{2:r}(h))\right).$$

Proof: See Appendix 1.

There are three points worth noting. First, the weak convergence of the eigenvalue estimator stems from the continuity property. Specifically, this implies that $\hat{\lambda}_i(h) \xrightarrow{p} \lambda_i(h)$ for $i = 1, \ldots, r$ and $\hat{\lambda}_i(h) \xrightarrow{p} 0$ for $i = r + 1, \ldots, N$. Second, a consistent estimate of the asymptotic variance-covariance matrix of the largest eigenvalue relies on a consistent estimate of both $\Omega(h)$ and the eigenvector associated with $\lambda_{\max}(h)$. This creates two sources of uncertainty in finite samples. Third, it is straightforward to show that the asymptotic distribution of the largest eigenvalue (and

of $\lambda_{2:r}(h)$, respectively) in the case of the *accumulated* or frequency-based Max Share approach has the same expression as in Theorem 3.2, except for the appropriate asymptotic variance-covariance matrix Ω defined in the *online Appendix*.

Finally, we derive the asymptotic distribution of the eigenvector associated with the maximal eigenvalue, as well as the joint distribution of the $r-1$ eigenvectors associated with the remaining $r-1$ largest (nonzero) eigenvalues, denoted by $\lambda_{2:r}(h)$. For simplicity, $q_{1,k}(h)$ is abbreviated as $q_1(h)$, and the results apply to any variable $k = 1, \cdots, N$.

Theorem 3.3. *Let Assumption 3.1 hold. Suppose that $\lambda_{\max}(h) > \lambda_2(h) + \epsilon$ for $\epsilon > 0$, i.e., the maximum eigenvalue of $S_k(h)$ is well-separated from the second highest eigenvalue. Then,*

i) $\widehat{q}_{1,T}(h) \xrightarrow{p} q_1(h)$ and the asymptotic distribution of $\widehat{q}_{1,T}(h)$ is:

$$\sqrt{T}\left(\widehat{q}_{1,T}(h) - q_1(h)\right) \xrightarrow{d} N\left(0, \Sigma_{q_1}(h)\right),$$

where $\Sigma_{q_1}(h) = (q_1'(h) \otimes I_N)F_1'(h)\Omega(h)F_1(h)(q_1(h) \otimes I_N)$, with $F_1(h) = \sum_{j=2}^{N}(\lambda_{\max}(h) - \lambda_j(h))^{-1}P_{\lambda_1}(h) \otimes P_{\lambda_j}'(h)$, and $P_{\lambda_j}(h) = q_j(h)q_j'(h)$ is the eigenprojection associated with $\lambda_j(h)$;

ii) $\widehat{Q}_{2:r,T}(h) \xrightarrow{p} Q_{2:r}(h)$ and the asymptotic distribution of $\mathrm{vec}(\widehat{Q}_{2:r,T})(h)$ is:

$$\sqrt{T}\left(\mathrm{vec}(\widehat{Q}_{2:r,T}(h)) - \mathrm{vec}(Q_{2:r}(h))\right) \xrightarrow{d} N\left(0, \Sigma_{Q_{2:r}}(h)\right),$$

where $\Sigma_{Q_{2:r}}(h) = (Q_{2:r}'(h) \otimes I_N)F_2'(h)\Omega(h)F_2(h)(Q_{2:r}(h) \otimes I_N)$, with $F_2(h) = \sum_{j=2}^{N}\sum_{i \neq j}(\lambda_j(h) - \lambda_i(h))^{-1}P_{\lambda_j}(h) \otimes P_{\lambda_i}'(h)$.

Proof: See Appendix 1.

As a consequence of the above results, considering the expression of the structural IRFs using the first (identified) structural shock associated with the largest eigenvalue (equation 2.16), the estimates of the structural IRFs depend on the reduced-form estimates $\hat{\Phi}_i$, the lower-triangular factor from the Cholesky decomposition, and the eigenvector $\hat{q}_{1,T}(h)$, which are nonlinear functions of the estimates of the autoregressive parameters. In the weakly stationary case, all these estimates converge in probability to their respective true values, implying that the structural IRFs associated with the largest eigenvalue are weakly consistent.[8]

In the special case where the error terms of the reduced-form VAR are normally distributed and the eigenvalues are all distinct, according to Anderson (1963), the expression of $\Sigma_{q_1}(h)$ is given by:

$$\Sigma_{q_1}(h) = Q_{2:r}(h)\left(\widetilde{\Lambda}(h)\right)^2 Q_{2:r}'(h),$$

[8]See Lütkepohl (2007) for the asymptotic properties of the reduced-form moving average coefficients in the stationary case.

where

$$\widetilde{\Lambda}(h) = \begin{bmatrix} (\lambda_1(h)\lambda_2(h))^{1/2}/(\lambda_1(h) - \lambda_2(h)) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (\lambda_1(h)\lambda_N(h))^{1/2}/(\lambda_1(h) - \lambda_N(h)) \end{bmatrix}.$$

In the general case, the expression of $\Sigma_{q_1}(h)$ (respectively, $\Sigma_{Q_{2:r}}(h)$) depends on $F_1(h)$ (respectively, F_2). Specifically, $F_1(h)$ is a linear combination of the Kronecker products between the eigenprojection associated with the maximal eigenvalue $\lambda_{\max}(h)$, denoted by $P_{\lambda_1}(h)$, and those associated with each other eigenvalue $\lambda_j(h)$, denoted by $P_{\lambda_j}(h)$. The weight of each Kronecker product is determined by the discrepancy between $\lambda_{\max}(h)$ and $\lambda_j(h)$, specifically by $\lambda_{\max}(h) - \lambda_2(h)$, that is, the difference between the two largest eigenvalues. When these two eigenvalues are roughly of the same magnitude, the presence of at least two driving structural shocks cannot be ruled out, causing $F_1(h)$ to become arbitrarily large. Thus, an examination of the empirical eigenvalues is necessary.

3.2 Max Share asymptotics with some roots at, or near, unity

We now discuss the asymptotics of the Max Share estimator in the presence of nearly unit root and/or nonstationary processes when the VAR model is estimated *in levels*. In the spirit of Phillips (1998), our primary interest is in the behavior of $\hat{S}_{k,T}$ when the sample size T goes to infinity and the horizon h is a fixed fraction f of T, that is, $h = fT$.[9]

Consider the general specification in levels and differences:

$$y_t = \Pi y_{t-1} + \Upsilon(L)\Delta y_{t-1} + u_t.$$

We construct the orthogonal matrix $B = [\beta_\perp \ \beta]$, where $\beta_\perp$ is an $N \times (N - r)$ orthogonal full-rank matrix containing the unit roots or near unit roots linear combinations of y_t, and β is an $N \times r$ orthogonal full-rank matrix containing the stationary linear combinations of y_t.

Following Phillips (1998), we assume that:

Assumption 3.2.

(a) u_t *is an i.i.d. process with zero mean, covariance matrix* $\Sigma_u > 0$*, and finite fourth cumulants;*

(b) The determinantal equation $|I_K - \sum_{i=1}^{p} A_i L^i| = 0$ *has roots on or outside the unit circle;*

(c) $\Pi = \beta_\perp \exp(T^{-1}\Gamma)\beta_\perp' + \beta\beta' + \alpha\beta'$*, where* $\alpha, \beta \in \mathbb{R}^{N\times r}$ *and* $0 \leq \operatorname{rank}(\alpha) = \operatorname{rank}(\beta) = r \leq m$*. Without loss of generality,* β *is orthonormal, and* $\Gamma \in \mathbb{R}^{s\times s}$ *is a constant matrix;*

[9] One could also consider the case where both h and T go to infinity such that $h/T \to 0$. However, this case is less relevant from a macroeconomic perspective.

(d) The matrix $\alpha_{\perp}' \left(I_N - \Upsilon(1)\right) \beta_{\perp}$ *is nonsingular, and* $\alpha_{\perp}, \beta_{\perp} \in \mathbb{R}^{N \times s}$*, with* $s = N - r$*, are the orthogonal complements of* α *and* β*, respectively.*

The standard condition **(a)** is necessary for deriving the asymptotic variance matrix. Condition **(b)** allows for the inclusion of both stationary and nonstationary components. Condition **(c)** encompasses the unit root and local-to-unity cases. Specifically, the matrix Γ can be interpreted as a noncentrality parameter matrix (see Phillips 1998).[10] Moreover, note that $\Pi = \alpha\beta'$ in the presence of unit roots and cointegration. Lastly, condition **(d)** specifies that the stochastic process y_t is driven by s random walks and/or nearly integrated processes. Consequently, the linear combinations $\beta_{\perp}' y_t$ exhibit unit roots or near unit roots (or a mixture of both), while $\beta' y_t$ remains stationary.

Interestingly, Assumption 3.2 covers several cases of interest. Notably, empirical macroeconomic applications often focus on one of the following four cases.

- **Case 1:** Some variables have a unit root while other variables are weakly stationary. For instance, in the bivariate case, one variable possesses a unit root (e.g., a TFP measure) and the other is stationary with an autoregressive coefficient (e.g., a financial spread), say $\rho = 0.9$. In this case, $s = 1$, $\Gamma = 1$, and $\alpha = [0 \;\; \rho - 1]'$.

- **Case 2:** All variables in the vector y_t possess a unit root without cointegration (Lütkepohl and Velinov, 2016). Accordingly, $\beta_{\perp} = I$, β is the null matrix, and $\exp(T^{-1}\Gamma) = I$.

- **Case 3:** All variables possess a unit root, but there are r cointegration relationships (e.g., the baseline quarterly model of King et al. (1991), $\exp(T^{-1}\Gamma) = I$, and $\Pi = \alpha\beta'$.

- **Case 4:** Some variables have a unit root (e.g., a TFP measure), while other variables have near unit roots (e.g., hours worked). In particular, Γ can be a diagonal matrix where some series may be I(1) processes corresponding to the components with $\Gamma_{ii} = 0$, and some series may be stable processes with near unit roots (that is, $\Gamma_{ii} < 0$).[11] The matrix Γ can be partitioned such that the first diagonal elements correspond to the I(1) variables, and the remaining elements correspond to the nearly integrated variables.

We are now in a position to present the asymptotics of the impulse responses and the Max Share matrix in the presence of unit or near-unit roots when the unrestricted (reduced-form) VAR is estimated *in levels*. To avoid any confusion, note that in the sequel, we use the index i (respectively, the notation h) to denote the impulse response horizon or lead time (respectively, the forecast error variance horizon).

[10] An alternative and asymptotically equivalent approach is to replace the matrix exponential representation with deviations from I_s of the form $I_s + T^{-1}\Gamma$.

[11] Note that if Γ has some nonzero off-diagonal elements, one can have series that are near integrated of different orders.

Theorem 3.4. *Consider the reduced-form VAR in levels (equation 2.1). Let Assumption 3.2 hold, and let* $f, g \in [0, 1]$. *Then,*

i) If the lead time $i = gT$, *where* $g > 0$ *is a fixed fraction of the sample, the limiting reduced-form impulse response matrix* Φ_i *is nonzero as* $T \to \infty$*:*

$$\Phi_i \Rightarrow \bar{\Phi} = \beta_\perp \exp(gU_\Gamma)\beta'_E, \tag{3.21}$$

ii) If $h = fT$, *where* $f > 0$ *is a fixed fraction of the sample, then the limiting non-accumulated Max Share matrix and the h-step-ahead forecast-error variance matrix at horizon h are random as* $T \to \infty$*:*

$$h^{-1}\widehat{S}_{k,T}(h) \Rightarrow f^{-1} \int_0^f \Sigma'_{tr}\beta_E \exp(sU'_\Gamma)\beta'_\perp e_k e'_k \beta_\perp \exp(sU_\Gamma)\beta'_E \Sigma_{tr} ds \equiv \bar{S}_k(U_\Gamma), \tag{3.22}$$

$$h^{-1}\texttt{MSE}(h) \Rightarrow f^{-1} \int_0^f \beta_\perp \exp(sU_\Gamma)\beta'_E \Sigma_u \beta_E \exp(sU'_\Gamma)\beta'_\perp ds, \tag{3.23}$$

where $\Rightarrow$ *denotes weak convergence. The formal definitions of the matrices* $\beta_\perp$, β_E, *and* U_Γ, *which is a matrix function of a mixture of unit-root or local-to-unity distributions (or a mixture of both distributions), are given in Appendix 2.*

Proof: These results follow directly from Lemma 2.2 and Theorem 3.1 in Phillips (1998).

Part (i) of Theorem 3.4 asserts that the limiting response matrices of the moving average (reduced-form) representation lie in the range of $\beta_\perp$ in the presence of roots at or near unity. This implies that the limiting impulse responses, denoted as $\bar{\Phi}$, are nonzero exclusively for nonstationary variables possessing unit roots or near unit roots, particularly for $\beta'_\perp y_t$, especially when the lead time i represents a significant fraction of the sample size. Moreover, the matrix β_E captures the permanent impact of the reduced-form innovations on $\beta'_\perp y_t$.

Importantly, result (i) shows that for $i = gT$, where $g > 0$ is a fixed fraction of the sample, the impulse response matrices in the moving average representation for the VAR *in levels* are inconsistent except at the very shortest horizons. More specifically, the limits of elements of the impulse response matrices become random variables rather than true values. The presence of roots at or near unity accelerates the convergence of OLS estimates and leads to (super-)consistency in OLS regressions in levels (see Sims et al., 1990). As explained by Phillips (1998), impulse response functions do not converge faster in some directions, defined by the range of $\beta_\perp$, but rather carry the effects of (near) unit roots indefinitely as the lead time i increases. It is important to note that (near) unit roots are estimated with some degree of error, and this error not only persists but also accumulates as $T \to \infty$, with the impulse response horizon constituting a non-negligible fraction of the sample size.

The second result (ii) of Theorem 3.4 establishes that the estimator of the Max Share matrix $\widehat{S}_{k,T}(h)$ becomes inconsistent and converges to a random matrix when the unrestricted VAR is estimated in levels. This inconsistency arises because $\widehat{S}_{k,T}(h)$ depends on the reduced-form impulse response estimates at medium-to-long horizons (equations 2.6 and 2.13). Specifically, this random matrix represents a continuous average of a (matrix) quadratic form derived from the limiting (reduced-form) impulse responses (equation 3.22). As a result, the estimators of the corresponding eigenvalues and eigenvectors are also inconsistent and fail to converge to their true values.

Similarly, the mean squared error converges to a random variable, which is a continuous average of a quadratic form derived from the limiting reduced-form impulse responses. Interestingly, the Max Share statistic can weakly converge in probability to a non-random matrix when the forecast error variance horizon is a fixed fraction of the sample size and T diverges to infinity, meaning that the mixture of unit root or local-to-unity distributions does not affect the limiting Max Share statistic. For example, this occurs in the first experiment of our Monte Carlo simulations. However, finite-sample approximations can be severely distorted relative to the limiting distribution.

Given that structural IRFs from the identified Max Share shock are given by equation (2.16), the presence of some roots at, or near, unity has three significant implications. Firstly, according to Theorem 3.4(i), structural impulse responses, which are fundamentally functions of the reduced-form impulse responses, are inconsistent when roots are at or near unity. Their limits are altered by the distribution of the unit root or near unit root processes. Secondly, as stated in Theorem 3.4(ii), structural impulse response functions are also inconsistent due to the estimation of the eigenvector $q_{1,k}(h)$. Specifically, with a medium- to long-term Max Share identification scheme, the (inconsistent) estimate of $q_{1,k}(h)$, derived from an inconsistent estimate of the Max Share matrix, affects all structural impulse response matrices. This impact is not limited to those with a lead time extending beyond a fixed fraction of the sample size but also contaminates the entire structural IRF matrices. Thirdly, in combination with the inherent inconsistency of the reduced-form impulse response matrices, non-normal asymptotics generally prevail. This results in non-normal random limits, even in the presence of stationary components within the VAR specification. Therefore, the structural IRFs are influenced by the stochastic nature of the eigenvector estimates and the nonstationarity embedded within the unrestricted VAR model *in levels*.

4 Monte Carlo simulations

This section provides some Monte Carlo simulations to study the performances of the Max Share procedure in the presence of misspecification regarding the integration order. We assume that the data generating process (DGP) is a bivariate VAR model:

$$\begin{pmatrix} \Delta X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12}+\delta \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \Delta X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} 0 & -a_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta X_{1,t-2} \\ X_{2,t-2} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \tag{4.24}$$

with

$$u_t = \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix} \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix},$$

where $w_t \sim N(0, I_2)$ is a bivariate vector of structural shocks.

The parameter δ controls the number of permanent structural shocks and the magnitude of the permanent effect of the second shock $\epsilon_{2,t}$ on the first variable $X_{1,t}$. When $\delta = 0$, only the first structural shock has a permanent impact on the first variable. To some extent, the corresponding specification can be viewed as the one often encountered in the macroeconomic literature to identify a permanent shock, for example, the identification of a technology shock with some measures of (labor or total) productivity and hours worked (see Section 6).[12] When $\delta \neq 0$, the two structural shocks have a permanent effect on the first variable (e.g., Fisher, 2006). In other words, the identification of the first structural shock can be contaminated by the second permanent structural shock, meaning the two permanent shocks can be confounded. By taking the transformation of the first variable, this specification is labeled the *first-difference* model.

On the other hand, the corresponding specification *in levels* is given by:

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 1+a_{11} & a_{12}+\delta \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} -a_{11} & -a_{12} \\ -a_{21} & 0 \end{pmatrix} \begin{pmatrix} X_{1,t-2} \\ X_{2,t-2} \end{pmatrix} + u_t. \tag{4.25}$$

In both cases, we can also consider a situation in which the second variable $X_{2,t}$ is nearly integrated, that is, $a_{22} = \exp(-c/T)$ with $c > 0$. To summarize, $X_{1,t}$ is integrated of order one and is either specified in first-difference or in level, and $X_{2,t}$ is either weakly stationary or nearly integrated in our Monte Carlo simulations. In the sequel, we assume that $a_{11} = 0$. Appendix 2 provides the derivation of the asymptotic distribution of the Max Share matrix $\widehat{S}_{1,T}(h)$ for different configurations of this DGP.

Using equation (4.24), we generate 10,000 samples of size $T = 240$ observations, a common sample

[12] It is worth emphasizing that the VAR(1) specification (the first set of experiments) is the DGP of Gospodinov et al. (2012) and Chevillon et al. (2020), whereas the VAR(2) (the second set of experiments) corresponds to that of Gospodinov (2010) and Gospodinov et al. (2011).

size in applied macroeconomic research. To control for initial condition effects, we include 200 pre-sampled observations that are subsequently discarded during estimation. In each replication, we set the lag order to its true value, whether considering $(\Delta X_{1,t}, X_{2,t})'$ or $(X_{1,t}, X_{2,t})'$, ensuring that our results are free from lag order misspecification issues. [13] For each replication, we perform OLS estimation for both the *first difference* (equation 4.24) and *level* (equation 4.25) VAR specifications. Additionally, for the level-based specification, we apply the analytical correction proposed by Pope (1990) and a bootstrap procedure (Kilian, 1998; Inoue and Kilian, 2002).[14] Subsequently, we identify two structural shocks using the Max Share approach, which involves maximizing the contribution of the first structural shock to the h-step ahead forecast error variance of the first variable $\Delta X_{1,t}$ or $X_{1,t}$ (equation 2.14). We explore different truncated forecast error variance horizons for the Max Share criterion, including $h = 0$, 40, and 80 quarters. Notably, when $h = 0$, the Max Share approach simplifies to a Cholesky decomposition of the variance-covariance matrix of the innovations u_t.

The results are evaluated across three dimensions. Firstly, after computing the (cumulative) mean bias and root mean squared error (RMSE) for selected lead times ($i = 0$, 4, 8, and 40 quarters), we analyze the average impulse response functions of the j-th variable due to the ℓ-th structural shock at each lead time i, using a forecast error variance horizon of $h = 0, 40$, or 80 quarters for the Max Share matrix. These average impulse response functions are denoted as $\overline{\texttt{IRF}}_{j\ell,i}(h)$ and are compared against the true impulse response function $\texttt{IRF}_{j\ell,i}$. Note that we only report the impulse responses for the first structural shock for sake of conciseness: detailed tables regarding the bias and RMSE, along with further evidence for the second structural shock, are provided in the *online appendix*. Secondly, we calculate the contemporaneous correlation between the estimated structural shocks and true structural shocks, denoted by $\texttt{corr}\left(\hat{w}_{i,t}, w_{i,t}\right)$ for $i = 1, 2$, as well as the contemporaneous correlation between the estimated structural shocks and true complementary structural shocks, denoted by $\texttt{corr}\left(\hat{w}_{i,t}, w_{j,t}\right)$ for $i \neq j$. Lastly, we analyze the empirical distribution of the first (and second) element of the eigenvector v_1, denoted by $v_{1,1}(h)$ (and $v_{2,1}(h)$), associated with the maximal eigenvalue of the Max Share matrix.

In our *initial experiment*, we assume that $(\Delta X_{1,t}, X_{2,t})'$ is modeled as a VAR(1) system with parameters $(a_{11}, a_{12}, a_{21}, a_{22}, b_{12}, b_{21}, \delta) = (0, 0, 0.2, 0.96, 0, 0.5, 0)$. Since $a_{11} = a_{12} = \delta = 0$, $X_{1,t}$ follows a random walk, while the second variable is a (persistent) stationary process driven by ρ. This configuration corresponds to **Case 1** as described in Section 4. Furthermore, with $\delta = 0$, only the first structural shock has a lasting impact on the first variable. As depicted in Figure 1

[13] Several robustness exercises, available upon request, were conducted to control for lag order selection, all of which confirm the consistency of our results.

[14] Bayesian estimation with Minnesota unit root priors and consideration of short-run restrictions were also conducted, although detailed results are not presented here, but are available upon request.

for $h = 0$, there is no contemporaneous bias observed in the average structural impulse response function (IRF) estimates, denoted by $\overline{\texttt{IRF}}_{11,0}(0)$ and $\overline{\texttt{IRF}}_{21,0}(0)$, regardless of how the nonstationary variable X_1 is handled. This absence of bias is consistent with the fact that the Max Share identification method is here essentially equivalent to a recursive Cholesky identification approach. As demonstrated by Phillips (1998), IRFs are then consistently estimated at short horizons i, where $i = fT$ represents a small fraction of the sample size.

As the lead time i of structural impulse response functions (IRFs) increases, the bias and root mean squared error (RMSE) of $\overline{\texttt{IRF}}_{11,i}(0)$ increase significantly when the reduced-form VAR is estimated *in levels*. Specifically, the (average) bias of the impulse response function of the first variable to the first structural shock, $\overline{\texttt{IRF}}_{11,i}(0)$, is approximately 0.05 at $i = 20$ and 0.07 at $i = 40$ for the *first-difference* VAR, whereas these biases are notably higher at 0.35 and 0.55, respectively, for the VAR *in levels*. Meanwhile, with the exception of the shortest horizons, the RMSE of the level-based specifications rises rapidly compared to the first-difference specification, showing a multiplication factor of two or even three at medium-to-long horizons.

Interestingly, both Pope's correction and the bootstrap method exhibit similar bias reduction performances, halving the bias compared to the (uncorrected) VAR *in levels*. However, the (average) bias remains substantial, around 0.15 and 0.3 at $i = 20$ and 40, respectively. This bias reduction comes at the cost of a slight RMSE increase at the shortest horizons ($i \leq 4$), followed by a much larger RMSE at medium-to-long horizons compared to the corresponding performances of the *first-difference* specification.

Furthermore, similar patterns are observed when analyzing the (average) impulse response function of the second variable to the first structural shock, as well as the corresponding RMSE at each horizon. Starting from $i > 4$, a notable discrepancy in bias performances between the *first-difference* and the *level*-based specifications is observed regarding $\overline{\texttt{IRF}}_{21}$. This relative performance is even more pronounced when examining the RMSE. Indeed, using Pope's correction or the bootstrap procedure effectively reduces the (average) bias to levels comparable to the *first-difference* specification but comes with a multiplication factor (for the RMSE) greater than two at medium-to-long horizons.

As illustrated in Figure 2 and Figure 3, an increase in the forecast error variance horizon h within the Max Share procedure unveils three main features. First, consistent with the findings of Theorem 3.4, a contemporaneous bias in the impulse response function of the first variable, resulting from the first structural shock, emerges when *level*-based methods are used. Additionally, neither Pope's correction nor standard bootstrap techniques fully mitigate this bias, particularly at the shortest impulse response horizons i. When analyzing the effect of the first structural shock on the

second variable, $\overline{\texttt{IRF}}_{21}$, both bias-correction methods display minimal (average) bias and perform comparably to the *first-difference* method, albeit at the cost of lower efficiency. The uncertainty associated with level-based structural IRF estimates for the second variable increases with the forecast error variance horizon h. Specifically, the RMSE for bias-corrected methods is higher than that inherited from ordinary least squares estimation of the VAR *in levels* when considering the IRF of the second variable due to the first structural shock. Conversely, for the IRF of the first variable, the RMSE from bias-corrected methods is lower.

Examining the eigenvector corresponding to the maximal eigenvalue, Figure 4 displays the distributions of its two elements when the forecast error variance horizon h is 40 or 80 quarters. When employing the *first-difference* specification, the distribution of the first element of the eigenvector exhibits a pronounced peak around the true value of the first unit vector element. In contrast, all estimation methods using the *level* specification lead to a significantly greater dispersion for the first element, with values ranging between 0.6 and 1. The distributions for the second element of the eigenvector, while approximately symmetric around the true value of 0, span a broad interval from -1 to 1.

These results can be rationalized by analyzing the asymptotic distribution of the Max Share statistic. According to the derivations presented in Appendix 2, the asymptotic distribution of the Max Share matrix is characterized as a random matrix expressed by:

$$h^{-1}\widehat{S}_{1,T}(h) \Rightarrow \frac{1}{f}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\int_0^f \exp(2Us)\,ds \tag{4.26}$$

with $\int_0^f \exp(2Us)\,ds = \frac{1}{2U}\left[\exp(2Uf)-1\right]$ where the real random variable U has a unit root distribution. Moreover, the asymptotic distribution of the mean squared error is given by:

$$h^{-1}e_1'\texttt{MSE}(h)e_1 \Rightarrow \frac{1}{f}\int_0^f \exp(2Us)\,ds.$$

These two results imply that the Max Share statistic is weakly convergent, i.e.

$$\frac{q_1'S_k(h)q_1}{e_1'\texttt{MSE}(h)e_1} \xrightarrow{p} \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}. \tag{4.27}$$

In this case, the limit of the Max Share statistic is consistent, and the limits in (weak) probability of the eigenvalue estimators are one and zero, respectively. Moreover, the limit of the eigenvector estimator associated with the maximal eigenvalue is the vector $[1\ 0]'$. This aligns with the simulation results depicted in Figure 4. Additionally, given the definition of the structural impulse response functions outlined in equation (2.16), the impact on these functions also depends on the distribution

of the reduced-form impulse responses Φ_i. As outlined in Theorem 3.4, the latter is given by:

$$\Phi_i \Rightarrow \begin{bmatrix} \exp(gU) & 0 \\ 0 & 0 \end{bmatrix}.$$

This distribution has a random limit, characterized by the exponential of the scalar unit root distribution, which notably exhibits left-skewness that intensifies with increasing values of g.[15] For instance, with $g = \frac{40}{240}$, our Monte Carlo simulations indicate that the resulting asymmetry in the distributions of the reduced-form impulse response functions is pronounced, featuring a significant negative skewness coefficient. This asymmetry mirrors that typically observed in unit-root distributions.[16]

For our *second experiment*, we maintain the same parameter vector as in the initial experiment. However, we now assume that both structural shocks have a permanent effect on the first variable ($\delta = -0.025 \neq 0$), potentially leading to a confounding effect. Several noteworthy observations arise from Figure 5. Firstly, consistent with Experiment 1, impulse response estimates derived from the *first-difference* method consistently outperform those from the *level* specification across all lead times, demonstrating superior bias and RMSE properties.[17] Secondly, as the forecast error variance horizon h increases, we observe significant differences, particularly regarding the impact of the first structural shock on the second variable. This suggests that the Max Share identification method may partially confound the two permanent structural shocks.

This interpretation is further supported by the correlation analysis between each estimated structural shock and the true complementary structural shock (see *online Appendix 2*). Specifically, we note that these (absolute) correlations hover around 25% for *level*-based impulse response estimates, whereas they are negligible when using the *first-difference* specification. Additionally, we observe an average 10% decrease in the correlation between each estimated structural shock and the true one for OLS-based, bias-corrected, and bootstrapped estimates derived from the *level* specification. In contrast, these correlations remain unchanged and close to 100% in the case of the *first-difference* specification.

Thirdly, we observe that the RMSE generally increases as the forecast error variance horizon extends in the Max Share procedure, particularly noticeable at the shortest impulse response horizons for *level*-based estimates. Fourthly, consistent with the results detailed in Appendix 2, two main features emerge regarding the distribution of the eigenvector elements (see Figure 6). On the one hand, using the *first-difference* specification in the presence of two persistent structural shocks

[15] See also Phillips (1998, Figure 1(a)).

[16] Note that the asymmetry results from the nonnormal limit theory (Phillips, 1998).

[17] Figures for $h = 0$ and $h = 80$ are reported in the *online Appendix*.

results in distributions that remain nearly invariant compared to those in the first experiment. On the other hand, employing the specification *in levels* broadens the support of the two distributions. Notably, the distribution of the second eigenvector element exhibits a right-skewed pattern, with a negative mean and median estimate of $v_{2,1}$ around -0.5, significantly deviating from the true value of 0. Increasing the forecast error variance horizon from 40 to 80 quarters further exacerbates this issue. This can be understood by examining the asymptotic distribution of $\widehat{S}_{1,T}(h)$ given by:

$$h^{-1}\widehat{S}_{1,T}(h) \Rightarrow \frac{1}{f}\begin{bmatrix} 0.3735 & -0.3395 \\ -0.3395 & 0.3086 \end{bmatrix} \int_0^f \exp(2Us)\, ds.$$

Comparing with the expression of the asymptotic distribution of the Max Share matrix estimator in equation 4.26, one main difference is that the matrix $\Sigma'_{\text{tr}}\beta_E\beta'_\perp e_1 e_1 \beta_\perp \beta'_E \Sigma_{Tr}$ given in the right-hand side now possesses four nonzero elements due to the presence of two permanent structural shocks, thus $\delta \neq 0$. It turns out that the finite sample estimation of these elements further contributes to increased uncertainty, compounding the finite sample approximation of nonnormal, asymmetric asymptotics associated with the unit-root distribution.[18]

For our last two reported experiments, we focus on **Case 4** (Section 3), where $(\Delta X_{1,t}, X_{2,t})'$ is modeled as a VAR(1) system with parameter $(a_{11}, a_{12}, a_{21}, a_{22}, b_{12}, b_{21}) = (0, -0.2, 0.2, 0.99, 0, 0.5)$ and $\delta = 0$ (*experiment 3*) or -0.025 (*experiment 4*). The second variable is modeled as a near-unit root process. In the case of a forecasting error variance horizon of 40 quarters, Figures 7 and 8 report the average impulse responses and the RMSE for $\delta = 0$ and $\delta = -0.025$, respectively.[19] Several points are worth noting. The presence of a near-unit root second variable substantially increases the bias of impulse response estimates, even in the case of the *first-difference* specification. In particular, the bias is more pronounced for greater impulse response horizons and $h = 0$ or 40 quarters when studying the effect of the first structural shock on the second variable (lower panels in Figure 7 and 8) in the presence or absence of a confounding effect. Meanwhile, the *level*-based estimates of $\overline{\texttt{IRF}}_{21}$ exhibit, as in our second experiment, a significant contemporaneous bias along with large RMSEs. With only one permanent structural shock ($\delta = 0$), the occurrence of a near-unit root for X_2 leads to contemporaneous correlations (in absolute value) $\texttt{corr}\,(\hat{w}_{i,t}, w_{j,t})$ of 15% between the two structural shocks. In the case of two permanent structural shocks, these correlations increase to around 60% for $h = 40$, while those between the j-th estimated *level*-based structural shock and the true one, $\texttt{corr}\,(\hat{w}_{i,t}, w_{i,t})$, drop to 74% ($h = 40$) and 60% ($h = 80$). Consistent with previous results, in the case of *first-difference* estimates, the correlations $\texttt{corr}\,(\hat{w}_{i,t}, w_{i,t})$ for $i = 1, 2$ remain close to 100%, and those between $\hat{w}_{i,t}$ and $w_{j,t}$ for $i \neq j$ are close to zero.

[18] Interestingly, when we apply the Max Share matrix estimator with a starting horizon of 20 quarters instead of zero and a maximum horizon of 80 quarters, the medians of the first eigenvector closely match those of the asymptotic matrix mentioned above.

[19] Results for $h = 0$ and 80 are also available in the supplementary material.

The rationale behind the structural IRF results remains consistent with the findings of the first two experiments. On the one hand, using the derivations detailed in Appendix 2, the asymptotic distribution of the Max Share matrix is given by:

$$h^{-1}\widehat{S}_{1,T}(h) \Rightarrow \frac{1}{f}\int_0^f \begin{bmatrix} 0.8615 & 0.6808 \\ -0.2 & 0.9615 \end{bmatrix} \exp(sU_\Gamma')e_1e_1' \exp(sU_\Gamma) \begin{bmatrix} 0.8615 & -0.2 \\ 0.6808 & 0.9615 \end{bmatrix} ds$$

where U_Γ denotes a matrix function representing a mixture of unit root and local-to-unity distributions and $\Gamma = \begin{bmatrix} 0 & c_1 \\ 0 & c_2 \end{bmatrix}$, with $\delta = 0$ (*experiment 3*) or $\delta = -.025$ (*experiment 4*), $\delta = \frac{c_1}{T}$, and c_1 characterizes the root near unity (see Assumption 3.2). The finite sample approximation of this more complex asymptotic distribution, which is nonnormal and asymmetric, significantly impacts the eigenvector associated to the largest eigenvalue.

On the other hand, this effect is compounded with the estimation of the reduced-form impulse responses whose asymptotic distribution, as indicated by Theorem 3.4, is given in both cases by:

$$\Phi_i \Rightarrow \exp(gU_\Gamma) \begin{bmatrix} 0.9615 & -0.2 \\ 0.2 & 0.9615 \end{bmatrix},$$

where $\exp(gU_\Gamma)$ is a 2×2 matrix. Consequently, the elements of the structural IRFs of the first variable, as identified by the Max Share approach, are adversely affected by the relationship (2.16).

Regarding the finite sample distribution of the two elements of the eigenvector associated with the maximal eigenvalue, the simulation results (Figures 9 and 10) from the first two experiments are further exacerbated in the context of both a near-unit root for the second variable and a possible confounding effect ($\delta \neq 0$). Specifically, the distributions of *level*-based estimates of the eigenvector elements exhibit either a left-skewed shape (for the first eigenvector element) or a right-skewed shape (for the second eigenvector element) when $h = 40$, with only minor concentration around the true value. As the forecast error variance horizon increases, both distributions undergo significant distortions in the presence of both a unit root and a near-unit root in the unrestricted VAR *in levels*. In particular, when $\delta = -0.025$, the distribution of the first eigenvector element displays an inverted U-shape with a broad range, while the distribution of the second eigenvector element is bimodal, with values predominantly clustered around either -1 or 1.[20]

[20]The bimodal distribution can be understood with the following argument: let $A = \begin{bmatrix} 1+\varepsilon & 0 \\ 0 & 1-\varepsilon \end{bmatrix}$, where $\varepsilon > 0$ is small enough. A has two eigenvalues $\lambda_{\max} = 1+\varepsilon$ and $\lambda_{\min} = 1-\varepsilon$. One eigenvector associated with the largest (respectively, smallest) eigenvalue is the first (respectively, second) basis vector of $\mathbb{R}^2$, $v_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}'$ (respectively,

In conclusion, given the prevalence of such a data-generating process (DGP) in macroeconomic applications, these simulation experiments highlight several interesting insights. First, structural impulse responses derived from VAR models in *levels* show a substantial loss in terms of bias and RMSE properties at intermediate and long horizons. This contrasts with those obtained from the *first-difference* specification. Second, while bias-corrected, bootstrap and Bayesian methods mitigate the bias issue, they still perform worse than estimates from a *stationary* representation, especially in terms of RMSE. Third, the presence of a potential confounding effect, such as two permanent shocks, exacerbates the discrepancies between *first-difference* estimates and *level*-based estimates. This further outlines the need for caution when interpreting results from unrestricted VAR models *in levels*.

5 Empirical application

Our empirical application highlights the potential issues of relying exclusively on VAR models *in levels*, aligning with our theoretical and simulation results. We draw upon the study of Ben Zeev and Khan (2015), who used Max Share identification to investigate the nature of investment-specific technology (IST) news shocks. Interestingly, their unrestricted VAR specification, which includes IST and TFP variables *in levels*, highlights two possible sources of long-run fluctuations, potentially leading to a confounding effect.

We consider a reduced-form VAR with five (log-) variables for the US economy from 1959Q1 to 2019Q4:

$$y_t = (\log \text{TFP}_t, \log \text{IST}_t, \log C_t, \log H_t, \Delta \log P_t)'.$$

The first variable is the real-time, quarterly series on total factor productivity (TFP) adjusted for variations in factor utilization, constructed by Fernald (2014). Our benchmark measure of IST is the inverse of the real price of investment, which is defined as the ratio of the investment deflator to the consumption deflator. The consumption deflator encompasses nondurable and service consumption from the National Income and Product Account (NIPA) whereas the investment deflator corresponds to private fixed investment and durable consumption. Following Whelan (2002), we use a Fisher index to obtain chain-aggregated data.[21] Consumption is measured as the sum of nondurables and services and is converted to per capita terms by dividing by the civilian noninstitutionalized population aged 16 and over. The real series is then obtained by using the corresponding

$v_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}'$). Consider a small perturbation of A, $A_\varepsilon = A + \epsilon \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$. The two eigenvalues remain unchanged, while (using the same normalization as in the initial matrix) one eigenvector associated with $\lambda_{\max}$ is the sum vector of $\mathbb{R}^2$, $v_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}'$ and $v_2 = \begin{bmatrix} 1 & -1 \end{bmatrix}'$.

[21] The IST series is nearly identical when using the Törnqvist-Theil discrete time approximation to a Divisia index.

chain-weighted deflator. The hours series is the log of total hours worked in the nonfarm business sector adjusted for the civilian noninstitutionalized population aged 16 and over. Finally, inflation is measured as the percentage change in the GDP deflator.

Our identification strategy assumes two sources of persistent fluctuations in the system, which we define as TFP and IST news shocks. In line with the approach of Kurmann and Sims (2021), we sequentially apply the Max Share approach from Francis et al. (2014) to identify two structural permanent shocks, setting the truncated forecast error variance horizon to $h = 80$ quarters. Let $\tilde{A}_0$ represent the lower triangular Cholesky factor of the reduced-form covariance matrix Σ_u, and Q be an orthonormal matrix such that all impact matrices are given by $A_0 = \tilde{A}_0 Q$. The first structural shock is identified by solving:

$$q_1(h) = \underset{q_1}{\text{argmax}} \quad \frac{q_1' S_1(h) q_1}{e_1' \texttt{MSE}(h) e_1} \quad \text{s.t.} \quad q_1' q_1 = 1$$

where q_1 is the first column of the matrix Q. This vector q_1 represents the linear combination that maximally contributes to the future forecast error variance of TFP over a given horizon h, indicating the maximum share of TFP's future forecast error variance explained by this shock.[22] The second structural shock is identified similarly, under the additional condition that this shock is orthogonal to the first structural shock:

$$q_2(h) = \underset{q_2}{\text{argmax}} \quad \frac{q_2' S_2(h) q_2}{e_2' \texttt{MSE}(h) e_2} \quad \text{s.t.} \quad q_2' q_2 = 1 \quad \text{and} \quad q_2' q_1 = 0$$

where q_2 is the second column of the matrix Q. This vector q_2 primarily accounts for the long-term fluctuations in IST. Consequently, the first two columns of $\tilde{A}_0 Q$ encompass the TFP and IST news shocks. To ensure the robustness of our identification strategy, we reverse the order of identification assigning q_1 to IST and q_2 to TFP.

In our empirical analysis, we employ two specifications. First, we estimate an unrestricted reduced-form VAR *in levels* with four lags, which is standard practice for quarterly data. Second, we estimate a VECM to account for potential unit roots and long-run relationships. Unit root tests provide evidence that the first three variables—TFP, IST, and consumption—are non-stationary. Moreover, economic theory suggests that consumption shares a stochastic trend with both TFP and IST.

More specifically, Johansen's (1995) cointegration tests, using both the trace and maximum eigenvalue test-statistics, reject the null hypothesis of a cointegration rank of two or less, but not of three

[22] As pointed out by Kurmann and Sims (2021) for the TFP shock, imposing short-run restrictions, such as zero impact restrictions used by Barsky and Sims (2011), may lead to misleading outcomes due to the imperfect measurement of factor utilization. Therefore, we refrain from imposing such identifying restrictions on TFP and IST permanent shocks.

or less. This result implies that the data is consistent with the presence of two stochastic trends, suggesting no more than one cointegrating relationship among TFP, IST, and consumption, assuming total hours worked and inflation are covariance stationary. In contrast, the Engle-Granger test rejects the null hypothesis of cointegration between any combinations of these three variables. To reconcile these conflicting pre-test results, we report results based on a single cointegration vector, although our findings remain robust even under the assumption of no cointegration, such as when estimating a VAR with the three non-stationary variables in first differences. This dual approach, using both a VECM and a first-differenced VAR, aims to provide a comprehensive understanding of the underlying dynamics while enhancing the robustness of our results against varying assumptions regarding the cointegration structure.

Figure 11 illustrates the structural impulse response functions (top panel) and the forecast error variance shares (bottom panel) due to the structural TFP shock on each variable under both identification strategies. Figure 12 provides a similar view for the structural IST shock. Notably, the IRFs and FEVD shares for the TFP shocks differ significantly depending on the identification ordering. For instance, while both orderings agree that a TFP shock increases hours worked (except at impact), the VAR *in levels* yields a distinct response and attributes a substantially larger share in the variance decomposition of hours worked to the TFP shock. Specifically, TFP shocks account for nearly 60% of the variance decomposition of hours worked after 10 quarters when identified first, but only 20% when identified second. Thus, when TFP shocks are identified first, they are seen as the main driver of fluctuations in hours worked; this conclusion, however, should be moderated when these shocks are identified second.

The differences are even more pronounced with IST shocks, as depicted in Figure 12. When the VAR *in levels* is employed, hours worked decline for several quarters following the impact of IST shocks when these shocks are identified conditional on TFP shocks. If IST shocks are identified first, they account for a significant portion of fluctuations in hours worked and consumption. However, when IST shocks are identified second, their impact is considerably less significant. This suggests that, in the second identification scheme, IST shocks may not be a primary driver of business cycles.

This analysis provides evidence that applying the Max Share approach to VAR *in levels* at a distant horizon can lead to conflicting results, likely due to the confounding of the two permanent shocks, as seen in our simulation experiments. By explicitly accounting for the stochastic trends using a VECM, the impact of the identification order is substantially reduced. As detailed in the *online Appendix*, this impact becomes almost negligible when employing a reduced-form VAR model with the first three variables in first differences. Regardless of the stationary transformation or the identification order, the impulse response functions and the forecast error variance shares remain consistent for the two structural shocks. Importantly, neither TFP nor IST shocks emerge

as the primary drivers of fluctuations in hours worked.

Finally, this empirical application underscores the potential pitfalls of relying solely on VAR *in levels*. While this application does not settle the debate on whether TFP or IST shocks are pivotal for business cycles, it highlights the sensitivity of results when using a VAR *in levels* in the presence of persistent processes with roots at or near unity. This emphasizes the importance of also estimating stationary representations, such as a reduced-form VECM or a reduced-form VAR with certain variables in first differences, to accurately capture the dynamics of structural shocks and enhance the robustness of macroeconomic analysis.

6 Conclusion

This paper critically examines the implications of using VAR models *in levels* with the Max Share identification approach, particularly in the presence of unit or near-unit root processes. Our theoretical and empirical analyses provide several key insights. First, structural impulse responses from level-based VARs exhibit significant bias and higher RMSE at intermediate and long horizons compared to those from stationary representations, although they perform similarly at very short horizons. Second, although bias-corrected, bootstrap, and Bayesian methods reduce some bias, they tend to increase RMSE and do not consistently outperform stationary specifications, such as first-difference models. Third, the presence of multiple permanent shocks exacerbates discrepancies between estimates from level-based and differenced VARs, potentially leading to confounding effects and unreliable identification of structural shocks.

These findings emphasize the importance of using stationary transformations, such as VECMs or differencing, and reporting the corresponding results to ensure reliable identification of structural shocks and impulse responses. Such transformations help mitigate the risk of identifying hybrid shocks instead of primitive shocks. While unrestricted VARs *in levels* can be useful when there is uncertainty about unit roots and cointegration, it is advisable to complement this approach with *stationary* model estimates, such as the adaptive automated VECM estimation procedure proposed by Liao and Phillips (2015), which effectively handles unknown cointegrating rank structures and transient lag dynamic orders. Alternatively, a thorough VECM robustness analysis can be conducted in a stepwise manner.

Additional supporting information may be found in the supplementary materiel of this paper:
Appendix 1: The accumulated Max Share approach and the Max Share approach in the frequency domain.
Appendix 2: Monte Carlo simulations.
Appendix 3: Empirical application.

References

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

Anderson, T. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.*, 34(1):122–148.

Angeletos, G.-M., Collard, F., and Dellas, H. (2020). Business-cycle anatomy. *American Economic Review*, 110(10):3030–3070.

Barsky, R. B. and Sims, E. R. (2011). News shocks and business cycles. *Journal of Monetary Economics*, 58(3):273–289.

Ben Zeev, N. and Khan, H. (2015). Investment-specific news shocks and u.s. business cycles. *Journal of Money, Credit, and Banking*, 47(7):1443–1464.

Benhima, K. and Cordonier, R. (2022). News, sentimens and capital flows. *Journal of International Economics*, 137.

Bernanke, B. (1986). Alternative explanations of the money-income correlation. *Carnegie Rochester Conference Series on Public Policy*, 25(0):49–99.

Bouakez, H. and Kemoe, L. (2023). News shocks, business cycles, and the disinflation puzzle. *Journal of Money, Credit, and Banking*, 59:2115–2151.

Bura, E. and Pfeiffer, R. (2008). On the distribution of the left singular vectors of a random matrix and its applications. *Statistics and Probability Letters*, 78:2275–2280.

Carriero, A. and Volpicella, A. (2024). Max share identification of multiple shocks: An application to uncertainty and financial conditions. *Journal of Business & Economic Statistics*. forthcoming.

Chen, K. and Wemy, E. (2015). Investment-specific technological changes: The source of long-run tfp fluctuations. *European Economic Review*, 80:230–252.

Chevillon, G., Mavroeidis, S., and Zhan, Z. (2020). Robust inference in structural vector autoregressions with long-run restrictions. *Econometric Theory*, 36:86–121.

Cooley, T. and LeRoy, S. (1985). Atheoretical macroeconomics: A critique. *Journal of Monetary Economics*, 16(3).

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

Dieppe, A., Francis, N., and Kindberg-Hanlon, G. (2019). The identification of dominant macroeconomic drivers: Coping with confounding shocks. ECB Working Paper 2534, European Central Bank.

Faust, J. (1998). The robustness of identified var conclusions about money. *Carnegie-Rochester Conference Series ob Public Policy*, 49(0):207–244.

Fernald, J. (2014). A quarterly, utilization-adjusted series on total factor productivity. Working Paper Series 2012-19, Federal Reserve Bank of San Francisco.

Fève, P. and Guay, A. (2014). Sentiments in svars. *Economic Journal*, 129:877–896.

Fisher, J. (2006). The dynamic effects of neutral and investment-specific technology shocks. *Journal of Political Economy*, 114(3):413–451.

Forni, M., Gambetti, L., and Sala, L. (2014). No news in business cycles. *Economic Journal*, 124:1168–1191.

Francis, N. and Kindberg-Hanlon, G. (2022). Signing out confounding shocks in variance-maximizing identification methods. *American Economic Association papers and proceedings*, 112:476–480.

Francis, N., Owyang, M., Roush, J., and DiCecio, R. (2014). A flexible finite-horizon alternative to long-run restrictions with an application to technology shocks. *The Review of Economics and Statistics*, 96(3):638–647.

Galí, J. (1999). Technology, employment and the business cycle: Do technology shocks explain aggregate productivity? *American Economic Review*, 41:1201–1249.

Gospodinov, N. (2010). Inference in nearly nonstationary svar models with long-run identifying restrictions. *Journal of Business & Economic Statistics*, 28(1):1–12.

Gospodinov, N., Herrera, A., and Pesavento, E. (2012). Unit roots, cointegration and pretesting in var models. *Advances in Econometrics*, 32:81–115.

Gospodinov, N., Maynard, A., and Pesavento, E. (2011). Sensitivity of impulse responses to small low-frequency comovements: Reconciling the evidence on the effects of technology shocks. *Journal of Business & Economic Statistics*, 29(4):455–467.

Inoue, A. and Kilian, L. (2002). Bootstrapping autoregressive processes with possible unit roots. *Econometrica*, 70:377–391.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, 2nd edition.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics*, 80:218–230.

Kilian, L. (2013). Structural vector autoregressions. In Hashimzade, N. and Thornton, M., editors, *Handbook of Research Methods and Applications in Empirical Macroeconomics*, pages 515–554. Edward Elgar, Cheltenham, UK.

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

Kilian, L., Plante, M., and Richter, A. (2024). Estimating macroeconomic news and surprise shocks. Technical report, Federal Reserve Bank of Dallas.

King, R., Plosser, C., Stock, J., and Watson, M. (1991). Stochastic trends and economic fluctuations. *American Economic Review*, 81(4):819–840.

Kurmann, A. and Otrok, C. (2013). News shocks and the slope of the term structure of interest rates. *American Economic Review*, 103(6):2612–2632.

Kurmann, A. and Sims, E. (2021). Revisions in utilization-adjusted tfp and robust identification of news shocks. *The Review of Economics and Statistics*, 103(2):216–235.

Levchenko, A. and Pandalai-Nayar, N. (2020). Tfp, news and sentiments: the international transmission of business cycles. *Journal of the European Economic Association*, 18(1):302–341.

Liao, Z. and Phillips, P. (2015). Automated estimation of vector error correction models. *Econometric Theory*, 31(3):581–646.

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

Lütkepohl, H. and Velinov, A. (2016). Structural vector autoregressions: Checking identifying long-run restrictions via heteroskedasticity. *Journal of Economic Surveys*, 30(2):377–392.

Mumtaz, H., Pinter, G., and Theodoridis, K. (2018). What do vars tell us about the impact of a credit supply shock? *International Economic Review*, 59(2):625–646.

Mumtaz, H. and Theodoridis, K. (2023). The federal reserve implicit inflation target and macroeconomic dynamics: A svar analysis. *International Economic Review*, 64(4):1749–1775.

Phillips, P. (1998). Impulse response and forecast error variance asymptotics in nonstationary vars. *Journal of Econometrics*, 83:21–56.

Pope, A. (1990). Biases of estimators in multivariate non-gaussian autoregressions. *Journal of Time Series Analysis*, 11:249–258.

Sims, C. (1980a). Comparison of interwar and postwar business cycles: Monetarism reconsidered. *American Economic Review*, 70(2):250–257.

Sims, C. (1980b). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Sims, C., Stock, J., and Watson, M. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58:113–144.

Tyler, D. (1981). Asymptotic inference for eigenvectors. *The Annals of Statistics*, 9(4):725–736.

Uhlig, H. (2003). What moves real gnp? Mimeo, Humbold University Berlin.

Uhlig, H. (2004). Do technology shocks lead to a fall in total hours worked? *Journal of the European Economic Association*, 2(2-3):361–371.

Whelan, K. (2002). A guide to us chain aggregated nipa data. *Review of Income and Wealth*, 48:21–233.

Appendix 1: Proofs of asymptotic results in the stationary case

Proof of Theorem 3.1

Let Assumption (3.1) hold and consider the behavior of $S_{k,T}(h)$ as $T \to \infty$ for some finite, fixed h. The asymptotic distribution can be derived from the delta method. Let $\tilde{\alpha} = \text{vec}([A_1, \ldots, A_p])$, one has

$$D_{\tilde{\alpha}}(h) := \frac{\partial \text{vec}(S_k(h))}{\partial \tilde{\alpha}'} = \sum_{\ell=0}^{h-1} \left(I_N \otimes \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k'\right) \frac{\partial \text{vec}(\Phi_\ell \Sigma_{\text{tr}})}{\partial \tilde{\alpha}'} + \sum_{\ell=0}^{h-1} \left(\Sigma_{\text{tr}}' \Phi_\ell' e_k e_k' \otimes I_N\right) \frac{\partial \text{vec}(\Sigma_{\text{tr}}' \Phi_\ell')}{\partial \tilde{\alpha}'}$$

with (see Lütkepohl Lütkepohl (2007) p.668, rules 7 and 6, respectively)

$$\begin{aligned}
\frac{\partial \text{vec}(\Phi_\ell \Sigma_{\text{tr}})}{\partial \tilde{\alpha}'} &= \left(\Sigma_{\text{tr}}' \otimes I_N\right) \frac{\partial \text{vec}(\Phi_\ell)}{\partial \tilde{\alpha}'} \\
\frac{\partial \text{vec}(\Sigma_{\text{tr}}' \Phi_l')}{\partial \tilde{\alpha}'} &= \left(I_N \otimes \Sigma_{\text{tr}}'\right) \frac{\partial \text{vec}(\Phi_\ell')}{\partial \tilde{\alpha}'} \\
&= \left(I_N \otimes \Sigma_{\text{tr}}'\right) K_{NN} \frac{\partial \text{vec}(\Phi_\ell)}{\partial \tilde{\alpha}'}.
\end{aligned}$$

where K_{NN} is the $N^2 \times N^2$ commutation matrix such that $\text{vec}(A') = K_{NN}\text{vec}(A)$ for any $N \times N$ matrix. Therefore, using the Kronecker product rules and the properties of the commutation matrix, one has:[23]

$$\begin{aligned}
\frac{\partial \text{vec}(S_k(h))}{\partial \tilde{\alpha}'} &= \sum_{\ell=0}^{h-1} \left(I_K \otimes \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k'\right) \left(\Sigma_{\text{tr}}' \otimes I_N\right) \frac{\partial \text{vec}(\Phi_\ell)}{\partial \tilde{\alpha}'} + \sum_{\ell=0}^{h-1} \left(\Sigma_{\text{tr}}' \Phi_\ell' e_k e_k' \otimes I_N\right) \left(I_N \otimes \Sigma_{\text{tr}}'\right) K_{NN} \frac{\partial \text{vec}(\Phi_\ell)}{\partial \tilde{\alpha}'} \\
&= \sum_{\ell=0}^{h-1} \left(\Sigma_{\text{tr}}' \otimes \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k'\right) \frac{\partial \text{vec}(\Phi_\ell)}{\partial \tilde{\alpha}'} + \sum_{\ell=0}^{h-1} \left(\Sigma_{\text{tr}}' \Phi_\ell' e_k e_k' \otimes \Sigma_{\text{tr}}'\right) K_{NN} \frac{\partial \text{vec}(\Phi_\ell)}{\partial \tilde{\alpha}'} \\
&= \sum_{\ell=0}^{h-1} \left(\Sigma_{\text{tr}}' \otimes \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k'\right) \frac{\partial \text{vec}(\Phi_\ell)}{\partial \tilde{\alpha}'} + \sum_{\ell=0}^{h-1} K_{NN} \left(\Sigma_{\text{tr}}' \otimes \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k'\right) \frac{\partial \text{vec}(\Phi_\ell)}{\partial \tilde{\alpha}'} \\
&= \sum_{\ell=0}^{h-1} \left(I_{N^2} + K_{NN}\right) \left(\Sigma_{\text{tr}}' \otimes \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k'\right) \frac{\partial \Phi_\ell}{\partial \tilde{\alpha}'}.
\end{aligned}$$

Likewise, defining $\sigma := \text{vech}(\Sigma_u)$,

$$\begin{aligned}
D_\sigma(h) := \frac{\partial \text{vec}(S_k(h))}{\partial \sigma'} &= \sum_{\ell=0}^{h-1} \left(I_N \otimes \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k' \Phi_\ell\right) \frac{\partial \text{vec}(\Sigma_{\text{tr}})}{\partial \sigma'} + \sum_{\ell=0}^{h-1} \left(\Sigma_{\text{tr}}' \Phi_\ell' e_k e_k' \Phi_\ell \otimes I_N\right) \frac{\partial \text{vec}(\Sigma_{\text{tr}}')}{\partial \sigma'} \\
&= \sum_{\ell=0}^{h-1} \left(I_{N^2} + K_{NN}\right) \left(I_N \otimes \Sigma_{\text{tr}}' \Phi_\ell' e_k e_k' \Phi_\ell\right) \frac{\partial \text{vec}(\Sigma_{\text{tr}})}{\partial \sigma'}.
\end{aligned}$$

We define the $Np \times (T-p)$ matrix $Y := (Y_{p+1}, \ldots, Y_T)$ and $\Gamma_Y(0) := E\left((Y - E(Y))\left(Y' - E(Y)\right)'\right)$, the contemporaneous covariance matrix of the vector process y_t. Under Assumption (3.1),

$$\sqrt{T} \begin{bmatrix} \tilde{\alpha}_T - \tilde{\alpha} \\ \sigma_T - \sigma \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} \Gamma_Y(0)^{-1} \otimes \Sigma_u & 0 \\ 0 & 2D_N^+ \left(\Sigma_u \otimes \Sigma_u\right) D_N^{+'} \end{bmatrix}\right)$$

where $D_N^+ := \left(D_N' D_N\right)^{-1} D_N'$ is Moore-Penrose generalized inverse of the $N^2 \times N(N+1)/2$ duplication matrix. Using the Delta method,

$$\sqrt{T}\,\text{vec}\left(\widehat{S}_{k,T}(h) - S_k(h)\right) \xrightarrow{d} N\left(0, \begin{bmatrix} D_{\tilde{\alpha}}(h) & D_\sigma(h) \end{bmatrix} \begin{bmatrix} \Gamma_Y(0)^{-1} \otimes \Sigma_u & 0 \\ 0 & 2D_N^+ \left(\Sigma_u \otimes \Sigma_u\right) D_N^{+'} \end{bmatrix} \begin{bmatrix} D_{\tilde{\alpha}}'(h) \\ D_\sigma'(h) \end{bmatrix}\right)$$

[23] Let G be $(m \times n)$ and F $(p \times q)$. Then $K_{pm}(G \otimes F) = (F \otimes G)K_{qn}$, with K_{pm} and K_{qn} some commutation matrices.

as stated.

Proof of Theorem 3.2

Suppose that the Max Share matrix $S_k(h)$ is of rank $r \leq N$. The spectral decomposition of $S_k(h)$ is given by:

$$S_k(h) = Q(h)\Lambda(h)Q'(h),$$

where $\Lambda(h)$ is the diagonal matrix whose diagonal elements are the eigenvalues arranged in algebraically nonincreasing order, and $Q(h)$ is the orthogonal matrix with $Q(h)Q'(h) = \mathrm{I}_N$ containing the associated eigenvectors. Given that the Max Share matrix is not necessarily of full rank (e.g., Case 1, Section 4), we assume that the first r eigenvalues are different from zero and the last $N-r$ eigenvalues are equal to zero. Accordingly, the submatrix $Q_r(h)$ contains the eigenvectors associated with the first r eigenvalues, and the submatrix $Q_{N-r}(h)$ contains the eigenvectors associated with the last $N-r$ eigenvalues, so that $Q(h)Q'(h) = Q_r(h)Q_r'(h) + Q_{N-r}(h)Q_{N-r}'(h)$. Additionally, $Q_r(h)$ can be partitioned as $Q_r(h) = [q_1(h) \quad Q_{2:r}(h)]$, where $q_1(h) = q_k(h)$ denotes the eigenvector associated with the maximal eigenvalue when the target is the kth variable. Therefore, the matrix $\Lambda(h)$ can be written as:

$$\begin{aligned}\Lambda(h) &\equiv \begin{bmatrix} \Lambda_r(h) & 0 \\ 0 & \Lambda_{N-r}(h) \end{bmatrix} \\ &= Q'(h)S_k(h)Q(h) = \begin{bmatrix} Q_r'(h)S_k(h)Q_r(h) & Q_r'(h)S_k(h)Q_{N-r}(h) \\ Q_{N-r}'(h)S_k(h)Q_r(h) & Q_{N-r}'(h)S_k(h)Q_{N-r}(h) \end{bmatrix},\end{aligned}$$

where $Q_r'(h)S_k(h)Q_r(h) = \begin{bmatrix} q_1'(h)S_k(h)q_1(h) & q_1'(h)S_k(h)Q_{2:r}(h) \\ Q_{2:r}'(h)S_k(h)q_1(h) & Q_{2:r}'(h)S_k(h)Q_{2:r}(h) \end{bmatrix}$. Using the vectorization of the matrix $\Lambda_r(h)$, we have:

$$\mathrm{vec}\left(\Lambda_r(h)\right) = \mathrm{vec}\left(Q_r'(h)S_k(h)Q_r(h)\right) = \left(Q_r'(h) \otimes Q_r'(h)\right)\mathrm{vec}\left(S_k(h)\right).$$

Using Theorem 3.1, $T^{1/2}\mathrm{vec}\left(\widehat{S}_{k,T}(h) - S_k(h)\right) \xrightarrow{d} \mathcal{N}(0,\Omega)(h)$, and assuming a weakly consistent estimate of the submatrix of eigenvectors $Q_r(h)$, $\widehat{Q}_{r,T}(h) \xrightarrow{p} Q_r(h)$, the two results of Theorem 3.2 follow by virtue of the Slutsky theorem.

Proof of Theorem 3.3

Theorem 3.3 is an application of Theorem 4.1 and Theorem 4.2. (p. 729) of Tyler (1981) and Theorem 1 in Bura and Pfeiffer (2008). The expression of the variance-covariance matrix of the eigenvector associated to the maximal eigenvalue results from Anderson Anderson (1963).

Appendix 2: Asymptotic properties of the Max Share matrix for the bivariate DGP

We first provide some preliminary notations.

Rotating the System in Separate I(1) and I(0) Components

Some of the results in the paper rely on rotating the system into separate I(1) and I(0) components following Phillips (1998). Define $C := (\beta_\perp, \beta)$, $z_t := C'y_t$, $\eta_t := C'u_t$ and $\tilde{A}_i := C'A_iC$. Then,

$$z_t = \sum_{i=1}^{p} \tilde{A}_i z_{t-i} + \eta_t.$$

Letting $\tilde{\Pi} := C'(\Pi)C$, we get

$$z_t = \tilde{\Pi} z_{t-1} + C'\Gamma(L)C\Delta z_{t-1} + \eta_t.$$

Furthermore, we define $W_t := \left(\Delta z'_{t-1}, \ldots, \Delta z'_{t-p+1}\right)' = \left(\Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}\right)(I_{p-1} \otimes C)$ and $F := C'\left(\Gamma_1, \ldots, \Gamma_{p-1}\right)(I_{p-1} \otimes C)$ whence

$$\begin{aligned} FW_t &= C'\Gamma(L)C\Delta z_t \\ \Rightarrow z_t &= \tilde{\Pi} z_{t-1} + FW_t + \eta_t. \end{aligned} \tag{6.28}$$

We now obtain a partition of the matrix $\tilde{\Pi}$. Under **(c)**, one gets

$$\tilde{\Pi} = \begin{bmatrix} \exp(n^{-1}\Gamma) & \beta'_\perp\alpha \\ 0_{r\times s} & I_r + \beta'\alpha \end{bmatrix}.$$

Two Random Matrix distributions

Following Phillips (1998) we define the random matrices U and U_Γ. Recall that $\eta_t := C'u_t$ are the residuals in the separated system and define $\bar{S}(.)$ as a vector of uncorrelated Brownian motions. Under assumption **(c)**, we further define $\tilde{\eta}_{1t} := \beta'_\perp\alpha z_{2,t-1} + F_1W_t + \eta_{1t}$, its long-run covariance matrix as Λ, the long-run covariance matrix of η_{1t} as P and the correlated Brownian motions $S_\eta(.) = P\bar{S}(.)$ and $S_1(.) = \Lambda\bar{S}(.)$. Then, the so-called unit root matrix is given by

$$U := \int_0^1 dS_\eta(\nu)S_1(\nu)' \left(\int_0^1 S_1(\nu)S_1(\nu)'d\nu\right)^{-1}. \tag{6.29}$$

Now, under assumption **(c')**, we must also introduce the following Ornstein-Uhlenbeck process $J_\Gamma(\nu) := \int_0^\nu \exp\left((\nu - s)\Gamma\right) dS_1(s)$ and the so-called local-to-unit matrix

$$U_\Gamma := \int_0^1 dS_\eta(\nu)J_\Gamma(\nu)' \left(\int_0^1 J_\Gamma(\nu)J_\Gamma(\nu)'d\nu\right)^{-1}. \tag{6.30}$$

Specific derivations for the experiments

We discuss part **(i)** of Theorem 3.4 (i.e., the limiting distribution of IRFs) in the special case of the DGP used in Section 4:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 + a_{11} & a_{12} + \delta \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} -a_{11} & -a_{12} \\ -a_{21} & 0 \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ y_{2t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

Equivalently, the reduced-form VECM form can be written as:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & \delta \\ 0 & a_{22} - 1 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

More generally,

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & \delta \\ 0 & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

Following Phillips (1998), the system of equations admits the following alternate companion form such that $\tilde{x}_t = E\tilde{x}_{t-1} + U_t$ with $\tilde{x}_t = (y_{1t}, y_{2t}, \Delta y_{1t}, \Delta y_{2t})'$:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & \delta & a_{11} & a_{12} \\ 0 & a_{22} & a_{21} & 0 \\ 0 & \delta & a_{11} & a_{12} \\ 0 & a_{22}-1 & a_{21} & 0 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ \Delta y_{1t-1} \\ \Delta y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{1t} \\ u_{2t} \end{bmatrix}.$$

where $U_t = \tilde{J}'u_t$ with $\tilde{J} = [I_2, I_2]$.

The (modified) companion matrix E can be partitioned as follows:

$$\begin{bmatrix} \exp(T^{-1}\Gamma) & E_{12} \\ \begin{bmatrix} 0_{r\times s} \\ \exp(T^{-1}\Gamma) - I_s \\ 0_{(p-1)\times s} \end{bmatrix} & E_{22} \end{bmatrix}. \tag{6.31}$$

where s is the number of unit root or near unit root variables, r the number of stationary variable and p the number of lags in the level-based specification. Note that the DGP imposes that the matrix Γ can be either a scalar or a 2×2 matrix depending on the number of unit roots or near-unit roots. In particular, when Γ is a matrix, the exponential function $\exp(\cdot)$ is understood as the matrix exponential.

As shown in Phillips (1998, p.28), the impulse response matrices can be rewritten in terms of this companion form as follows

$$\begin{aligned} \Phi_i &= JE^i\tilde{J}' \\ &= J \begin{bmatrix} \exp(iT^{-1}\Gamma) & \exp(iT^{-1}\Gamma)E_{12}(I + \cdots + E_{22}^{i-1}) \\ O(T^{-1}) & E_{22}^i + O(T^{-1}) \end{bmatrix} \tilde{J}'. \end{aligned}$$

and $J = [I_2, 0_{2\times 2}]$ and the moving average (MA) representation of y_t is

$$y_t = \sum_{i=0}^{t-1} JE^i\tilde{J}'u_{t-i} = \sum_{i=0}^{t-1} \Phi_i u_{t-i}.$$

For $i = fT$ for a fixed fraction $f > 0$, then $\Phi_i \to \bar{\Phi}$ where

$$\bar{\Phi} = \begin{bmatrix} \exp(f\Gamma) & \exp(f\Gamma)E_{12}\left(I - E_{22}\right)^{-1} \\ 0 & 0 \end{bmatrix} \tilde{J}'$$

as $T \to \infty$ because E_{22} has stable roots. Using an OLS-based estimation of the unrestricted VAR in levels, the ith impulse response matrix estimate is given by:

$$\widehat{\Phi}_i = J\widehat{E}^i\tilde{J}'.$$

where

$$\widehat{E} = \begin{bmatrix} \widehat{B} & \widehat{F}_{12} \\ \widehat{B} - I_2 & \widehat{F}_{22} \end{bmatrix}. \tag{6.32}$$

with

$$B = \begin{bmatrix} 1 & \delta \\ 0 & a_{22} \end{bmatrix} \quad \text{and} \quad F_{12} = F_{22} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & 0 \end{bmatrix}.$$

Let B_1 denote the submatrix containing the first s columns of B (i.e., the nonstationary components of B), the limit distribution is given by:

$$T(\widehat{B}_1 - B_1) \Rightarrow \left(\int_0^1 dS_\eta J_\Gamma' \right) \left(J_\Gamma J_\Gamma' \right)^{-1} = U_\Gamma$$

where $J_\Gamma(r) = \int_0^r \exp\{(r-s)\Gamma\} dS_1$ is a vector diffusion process and S_1 is vector of Brownian motion (see Phillips, 1988). In particular, when the diagonal elements of Γ are equal to zero, then $J_\Gamma(r)$ reduces to $S_1(r)$.
It follows that the expression of $\widehat{E}^i$ is:

$$\widehat{E}^i = \begin{bmatrix} \widehat{B}_1^i + O_p(T^{-1}) & \sum_{h=0}^i \widehat{B}_1^{i-h} \widehat{E}_{12} \widehat{E}_{22}^h + O_p(T^{-1}) \\ O_p(T^{-1}) & \widehat{E}_{22}^i + O_p(T^{-1}). \end{bmatrix} \tag{6.33}$$

and, as shown by Phillips Phillips (1998),

$$\widehat{B}_1^i = [I_s + (\widehat{B}_1 - I_s)]^i = [I_s + T(\widehat{B}_1 - I_s)/T]^{fT} \Rightarrow \exp(fU_\Gamma)$$

as $T \to \infty$ and $i = fT$. Finally,

$$\widehat{E}^i \Rightarrow \begin{bmatrix} \exp(fU_\Gamma) & \exp(fU_\Gamma) E_{12} (I - E_{22})^{-1} \\ 0 & 0 \end{bmatrix}$$

and

$$\widehat{\Phi}_i \Rightarrow J \begin{bmatrix} \exp(fU_\Gamma) & \exp(fU_\Gamma) E_{12} \left(I - E_{22}\right)^{-1} \\ 0 & 0 \end{bmatrix} \tilde{J}'.$$

.

We now provide the complete derivations of different configurations.

- y_{1t} **is** $I(1)$ **and** y_{2t} **is** $I(0)$ **with possibly** $\delta \neq 0$

  In this case $s = 1$, $\Gamma = 0$ and the matrix E is given by (assuming that $a_{11} = 0$ as in Section 5):

$$E = \begin{bmatrix} 1 & E_{12} \\ 0_{3\times 1} & E_{22} \end{bmatrix} = \left[\begin{array}{c|ccc} 1 & \delta & 0 & a_{12} \\ \hline 0 & a_{22} & a_{21} & 0 \\ 0 & \delta & 0 & a_{12} \\ 0 & a_{22} - 1 & a_{21} & 0 \end{array}\right].$$

  where $E_{12} = [\delta \quad 0 \quad a_{12}]$ and

$$E_{22} = \begin{bmatrix} a_{22} & a_{21} & 0 \\ \delta & 0 & a_{12} \\ a_{22} - 1 & a_{21} & 0 \end{bmatrix}$$

To determine the (random) limit, we need to compute $E_{12}\left(I_3 - E_{22}\right)^{-1}$. Using

$$\left(I_3 - E_{22}\right)^{-1} = \begin{bmatrix} 1-a_{22} & -a_{21} & 0 \\ -\delta & 1 & -a_{12} \\ 1-a_{22} & -a_{21} & 1 \end{bmatrix}^{-1} = \frac{1}{d_{22}} \begin{bmatrix} 1-a_{12}a_{21} & a_{21} & a_{12}a_{21} \\ \delta - a_{12}(1-a_{22}) & (1-a_{22}) & a_{12}(1-a_{22}) \\ a_{21}\delta - (1-a_{22}) & 0 & (1-a_{22}) - a_{21}\delta \end{bmatrix},$$

we have

$$\begin{aligned} E_{12}(I_3 - E_{22})^{-1} &= \frac{1}{d_{22}} \begin{bmatrix} (\delta(1-a_{21}a_{21}) - a_{12}((1-a_{22}) - a_{21}\delta)) & \delta a_{21} & (\delta a_{12}a_{21} + a_{12}((1-a_{22}) - a_{21}\delta)) \end{bmatrix} \\ &= \frac{1}{d_{22}} \begin{bmatrix} \delta - a_{12}(1-a_{22}) & a_{21}\delta & a_{12}(1-a_{22}) \end{bmatrix} \end{aligned}$$

where $d_{22} = (1 - a_{22} - a_{21}\delta)$. Therefore,

$$\lim_{i\to\infty} E^i = \bar{E} = \begin{bmatrix} 1 & E_{12}\left(I_3 - E_{22}\right)^{-1} \\ 0_{3\times 1} & 0_{3\times 3} \end{bmatrix}.$$

As the impulse response horizon $i \to \infty$, the impulse response matrix of the unrestricted (reduced-form) VAR lin levels converge to $\bar{\Phi}$:

$$\lim_{i\to\infty} \Phi_i = \bar{\Phi} = J\bar{E}\tilde{J}' = \begin{bmatrix} 1 + \frac{a_{21}\delta}{d_{22}} & \frac{\delta}{d_{22}} \\ 0 & 0 \end{bmatrix}$$

It means that the only permanent impact is from both shocks on the first variable. Turning to the limiting distribution of the Max Share matrix, we first define $\beta_\perp$, β and β_E. Taking that $\beta'_\perp = [1\ 0]$, $\beta' = [0\ 1]$ and $C = [\beta_\perp\ \beta]$, the permanent effect of the reduced-form innovations on the first variable is given by:

$$\begin{aligned} \beta'_E &= \beta'_\perp + E_{12}\left(I_3 - E_{22}\right)^{-1} \begin{bmatrix} \beta' \\ C' \end{bmatrix} \\ &= [1\ 0] + \begin{bmatrix} \frac{\delta - a_{12}(1-a_{22})}{d_{22}} & \frac{a_{21}\delta}{d_{22}} & \frac{a_{12}(1-a_{22})}{d_{22}} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 + \frac{a_{21}\delta}{d_{22}} & \frac{\delta}{d_{22}} \end{bmatrix}. \end{aligned}$$

Finally, it involves the expression of the matrix U, which is a scalar. In so doing, we derive the expression of η_{1t} from the specification of Δy_{1t}, namely

$$\begin{aligned} \Delta y_{1t} &= \delta y_{2t-1} + a_{12}\Delta y_{2t-1} + u_{1t} = \eta_{1t} \\ &= \sum_{i=0}^{\infty} \psi_i u_{t-i} \end{aligned}$$

where the last expression results from the moving average representation of the innovations $u_t = (u_{1t}, u_{2t})'$.Then,

$$U = \int_0^1 dS_\eta(\nu) S_1(\nu) \left(\int_0^1 S_1(\nu) S_1(\nu)' d\nu \right)^{-1} = \frac{\int_0^1 S_1(\nu) dS_\eta(\nu)}{\int_0^1 S_1(\nu)^2 d\nu}$$

where S_1 is a scalar Brownian motion with a variance given by the long-run variance of $\sum_{i=0}^{\infty} \psi_i u_{t-i}$ and S_η is a scalar Brownian motion with variance given by $var(u_{1t})$. The asymptotic distribution of $h^{-1}S_{k,T}(h)$ as $T \to \infty$ with $h = fT$ for $f \in]0, 1]$ is therefore:

$$\begin{aligned} h^{-1}S_{k,T}(h) &\Rightarrow \frac{1}{f} \int_0^f \Sigma'_{\text{tr}} \beta_E \exp\left(sU'\right) \beta'_\perp e_k e'_k \beta_\perp \exp\left(sU\right) \beta'_E \Sigma_{\text{tr}} ds \\ &\Rightarrow \frac{1}{f} \int_0^f \Sigma'_{\text{tr}} \begin{bmatrix} 1 + \frac{a_{21}\delta}{d_{22}} \\ \frac{\delta}{d_{22}} \end{bmatrix} \exp\left(sU\right) \begin{bmatrix} 1 & 0 \end{bmatrix} e_k e'_k \begin{bmatrix} 1 \\ 0 \end{bmatrix} \exp\left(sU\right) \begin{bmatrix} 1 + \frac{a_{21}\delta}{d_{22}} & \frac{\delta}{d_{22}} \end{bmatrix} \Sigma_{\text{tr}} ds \end{aligned}$$

where Σ_{tr} is the lower triangular Cholesky decomposition of Σ_u, $\Sigma_u = \Sigma'_{\text{tr}}\Sigma_{\text{tr}}$, and e_k is the first ($k = 1$) or second ($k = 2$) base vector of $\mathbb{R}^2$. When $k = 1$, one has:

$$
\begin{aligned}
h^{-1}S_{1,T}(h) &\Rightarrow \frac{1}{f}\int_0^f \exp(2sU)\,\Sigma'_{\text{tr}}\begin{bmatrix}1+\frac{a_{21}\delta}{d_{22}}\\ \frac{\delta}{d_{22}}\end{bmatrix}\begin{bmatrix}1+\frac{a_{21}\delta}{d_{22}} & \frac{\delta}{d_{22}}\end{bmatrix}\Sigma_{\text{tr}}ds \\
&\Rightarrow \frac{1}{f}\int_0^f \exp(2sU)\begin{bmatrix}\tilde{\sigma}_{11} & \tilde{\sigma}_{21}\\ 0 & \tilde{\sigma}_{22}\end{bmatrix}\begin{bmatrix}1+\frac{a_{21}\delta}{d_{22}}\\ \frac{\delta}{d_{22}}\end{bmatrix}\begin{bmatrix}1+\frac{a_{21}\delta}{d_{22}} & \frac{\delta}{d_{22}}\end{bmatrix}\begin{bmatrix}\tilde{\sigma}_{11} & 0\\ \tilde{\sigma}_{21} & \tilde{\sigma}_{22}\end{bmatrix}ds \\
&\Rightarrow \frac{1}{f}\int_0^f \exp(2sU)\begin{bmatrix}\tilde{\sigma}_{11}\left(1+\frac{a_{21}\delta}{d_{22}}\right)+\tilde{\sigma}_{21}\frac{\delta}{d_{22}}\\ \tilde{\sigma}_{22}\frac{\delta}{d_{22}}\end{bmatrix}\begin{bmatrix}\tilde{\sigma}_{11}\left(1+\frac{a_{21}\delta}{d_{22}}\right)+\tilde{\sigma}_{21}\frac{\delta}{d_{22}} & \tilde{\sigma}_{22}\frac{\delta}{d_{22}}\end{bmatrix}ds \\
&\Rightarrow \frac{1}{f}(2U)^{-1}\begin{bmatrix}\left(\tilde{\sigma}_{11}\left(1+\frac{a_{21}\delta}{d_{22}}\right)+\tilde{\sigma}_{21}\frac{\delta}{d_{22}}\right)^2 & \left(\tilde{\sigma}_{11}\left(1+\frac{a_{21}\delta}{d_{22}}\right)+\tilde{\sigma}_{21}\frac{\delta}{d_{22}}\right)\tilde{\sigma}_{22}\frac{\delta}{d_{22}}\\ \left(\tilde{\sigma}_{11}\left(1+\frac{a_{21}\delta}{d_{22}}\right)+\tilde{\sigma}_{21}\frac{\delta}{d_{22}}\right)\tilde{\sigma}_{22}\frac{\delta}{d_{22}} & \left(\tilde{\sigma}_{22}\frac{\delta}{d_{22}}\right)^2\end{bmatrix}(\exp(f)-1).
\end{aligned}
$$

where $\tilde{\sigma}_{ij}$ are the elements of lower triangular Cholesky decomposition of Σ_u. Accordingly, $\widehat{S}_{1,T}(h)$ is of rank one, the second eigenvalue converges to zero and the resulting eigenvector converges to $[1\ \ 0]'$. When $\delta = 0$,

$$
h^{-1}\widehat{S}_{1,T}(h) \Rightarrow \frac{1}{f}(2U)^{-1}\begin{bmatrix}\tilde{\sigma}_{11}^2 & 0\\ 0 & 0\end{bmatrix}(\exp(f)-1).
$$

- y_{1t} **is** $I(1)$ **and** y_{2t} **is** $I(1)$

  In this case, $s = 2$, Γ is a 2×2 null matrix and E is given by:

$$
E = \begin{bmatrix}I_{2\times 2} & E_{12}\\ 0_{2\times 2} & E_{22}\end{bmatrix} = \left[\begin{array}{cc|cc}1 & 0 & 0 & a_{12}\\ 0 & 1 & a_{21} & 0\\ \hline 0 & 0 & 0 & a_{12}\\ 0 & 0 & a_{21} & 0\end{array}\right].
$$

  Since assumption **(c)** is satisfied here, we first need to compute the expression of $E_{12}(I_2 - E_{22})^{-1}$:

$$
(I_2 - E_{22})^{-1} = \begin{bmatrix}1 & -a_{12}\\ -a_{21} & 1\end{bmatrix}^{-1} = \frac{1}{1-a_{12}a_{21}}\begin{bmatrix}1 & a_{12}\\ a_{21} & 1\end{bmatrix}.
$$

  It implies that:

$$
E_{12}(I_2 - E_{22})^{-1} = \frac{1}{1-a_{12}a_{21}}\begin{bmatrix}a_{12}a_{21} & a_{12}\\ a_{21} & a_{12}a_{21}\end{bmatrix}
$$

  and thus

$$
\lim_{i\to\infty} E^i = \begin{bmatrix}I_{2\times 2} & E_{12}(I_2 - E_{22})^{-1}\\ 0_{2\times 2} & 0_{2\times 2}\end{bmatrix}
$$

  Capitalizing on Lemma 2.2. of Phillips (1998), one gets:

$$
\lim_{i\to\infty}\Phi_i = \bar{\Phi} = \begin{bmatrix}I_2 & E_{12}(I_2 - E_{22})^{-1},\end{bmatrix}
$$

  meaning that the only permanent impact is from both shocks on the first variable. Regarding the limiting distribution of $S_k(h)$, since β is the empty matrix,

$$
\beta'_\perp = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} = C
$$

and the permanent effect of the reduced-form innovations on the two variables is then given by:

$$\beta_E' \quad = \beta_\perp' + E_{12}\left(I_2 - E_{22}\right)^{-1} C' = \begin{bmatrix} \frac{1}{1-a_{12}a_{21}} & a_{12} \\ a_{21} & \frac{1}{1-a_{12}a_{21}} \end{bmatrix}.$$

Finally, starting from the (infinite) moving average representation of $\tilde{\eta}_{1t}$

$$\begin{aligned} \tilde{\eta}_{1t} \quad = \quad & \begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & a_{12} \\ a_{21} & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \\ & = \sum_{i=0}^{\infty} \Upsilon_1^i u_{t-i} \end{aligned}$$

where $\Upsilon_1^i = E_{12}^i$, one can define the matrix unit root distribution

$$U := \int_0^1 dS(\nu) S'(\nu) \left(\int_0^1 S(\nu) S(\nu)' \right)$$

where $S(\nu)$ is a two dimensional Brownian motion vector with covariance matrix $\Upsilon(1)\Sigma_u\Upsilon(1)'$. Using Theorem , the asymptotic distribution of $h^{-1}\widehat{S}_{k,T}(h)$ as $T \to \infty$ with $h = fT$ for $f \in]0,1]$ is therefore

$$\begin{aligned} h^{-1}\widehat{S}_{k,T}(h) &\Rightarrow \frac{1}{f} \int_0^f \Sigma_{\text{tr}}' \beta_E \exp\left(sU'\right) \beta_\perp' e_k e_k' \beta_\perp \exp\left(sU\right) \beta_E' \Sigma_{\text{tr}} ds \\ &\Rightarrow \frac{1}{f} \int_0^f \Sigma_{\text{tr}}' \beta_E \exp\left(sU'\right) e_k e_k' \exp\left(sU\right) \beta_E' \Sigma_{\text{tr}} ds. \end{aligned}$$

- y_{1t} **is** $I(1)$ **and** y_{2t} **is a nearly unit-root process with possibly** $\delta \neq 0$

  The derivation is the same, with the exception that:

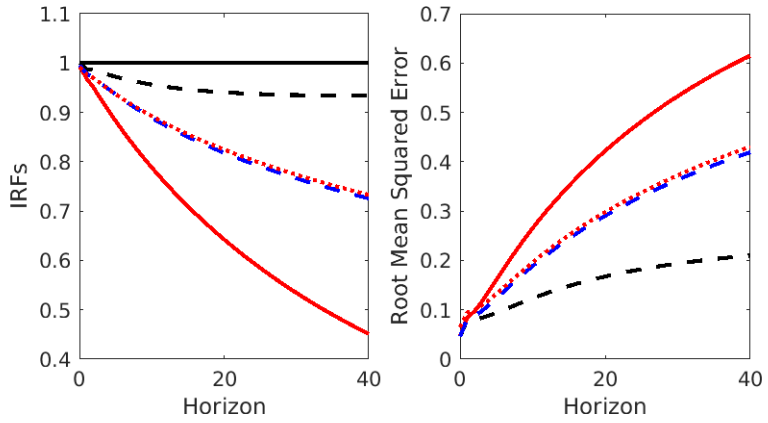
$$U_\Gamma := \int_0^1 dS_\eta(\nu) J_\Gamma(\nu)' \left(\int_0^1 J_\Gamma(\nu) J_\Gamma(\nu)' d\nu \right)^{-1}$$

  where

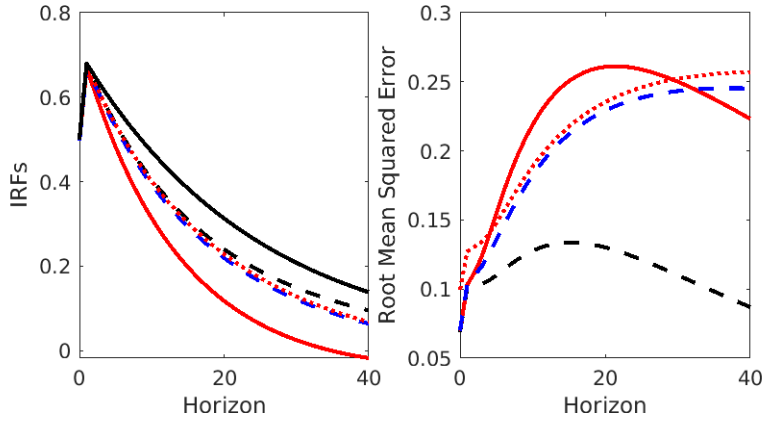
$$\Gamma = \begin{bmatrix} 0 & \delta \\ 0 & c \end{bmatrix}$$

  This is a mixture of correlated unit root distribution and a local-to-unit root distribution.

Figure 1: Impulse response effects of the first structural shock based on a contemporaneous (h = 0) Max-Share identification (experiment 1)



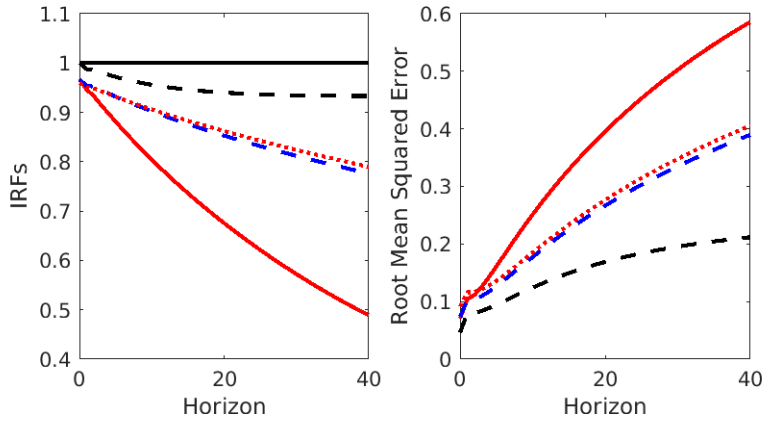
(a) Variable 1



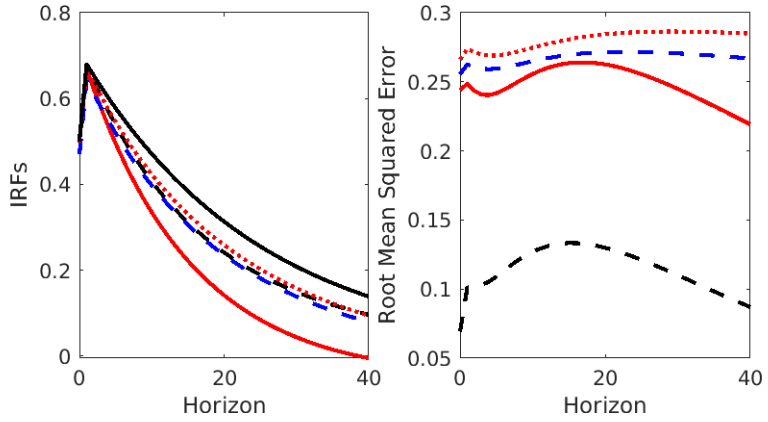
(b) Variable 2

Notes: (1) The black solid line represents the true IRF, $\texttt{IRF}_{11,i}$ and $\texttt{IRF}_{21,i}$ for $i = 0, \cdots, 40$, whereas the dashed line, the red solid line, the blue dashed line, and the red dotted line represent the average IRF estimates, $\overline{\texttt{IRF}}_{11,i}(0)$ and$\overline{\texttt{IRF}}_{11,i}(0)$, inherited from the first-difference specification, the level specification, the bias-correction method of Pope, and a bootstrap procedure, respectively.
(2) $X_{1,t}$, is a random walk, and the second variable is a persistent stationary process with an intrinsic persistence $\rho = 0.96$. (3) The top (respectively, bottom) left panel displays the average impulse responses of the first (respectively, the second) variable to the first structural shock. The top (respectively, bottom) right panel displays the RMSE of the corresponding structural IRF. (4) The lagged parameters are given by $a_{11} = a_{12} = 0$, $a_{12} = 0.2$, $a_{22} = 0.96$, $\delta = 0$.

Figure 2: Impulse response effects of the first structural shock based on a *non-accumulated* Max-Share identification with $h = 40$ (experiment 1)



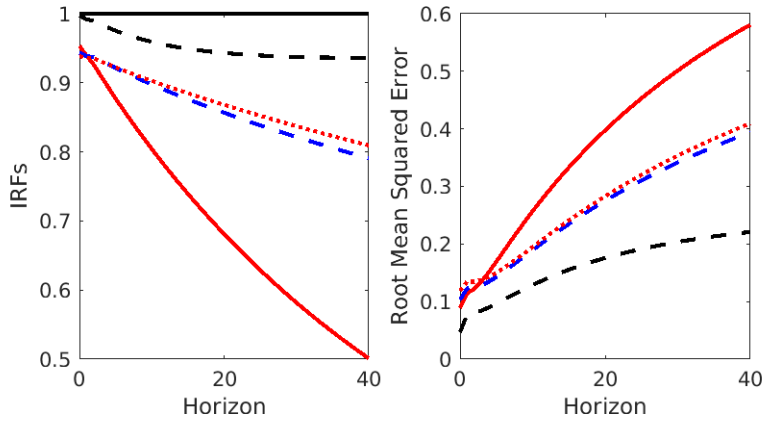
(a) Variable 1



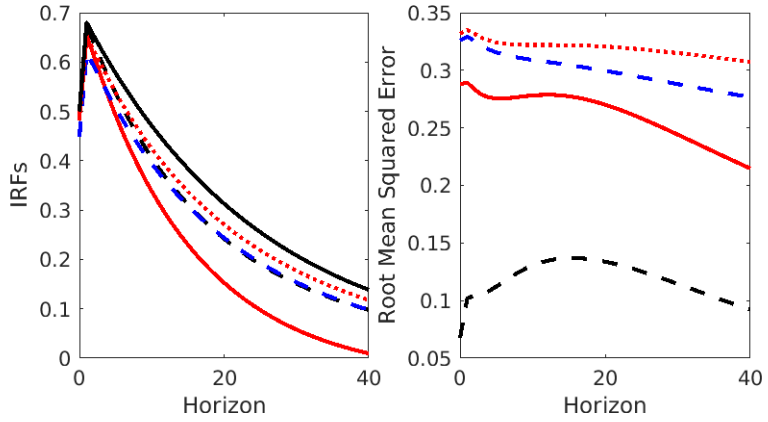
(b) Variable 2

Notes: The black solid line represents the true IRF, $\texttt{IRF}_{11,i}$ and $\texttt{IRF}_{21,i}$ for $i = 0, \cdots, 40$, whereas the dashed line, the red solid line, the blue dashed line, and the red dotted line represent the average IRF estimates, $\overline{\texttt{IRF}}_{11,i}(0)$ and$\overline{\texttt{IRF}}_{11,i}(0)$, inherited from the first-difference specification, the level specification, the bias-correction method of Pope, and a bootstrap procedure, respectively.

Figure 3: Impulse response effects of the first structural shock based on a *non-accumulated* Max-Share identification with $h = 80$ (experiment 1)



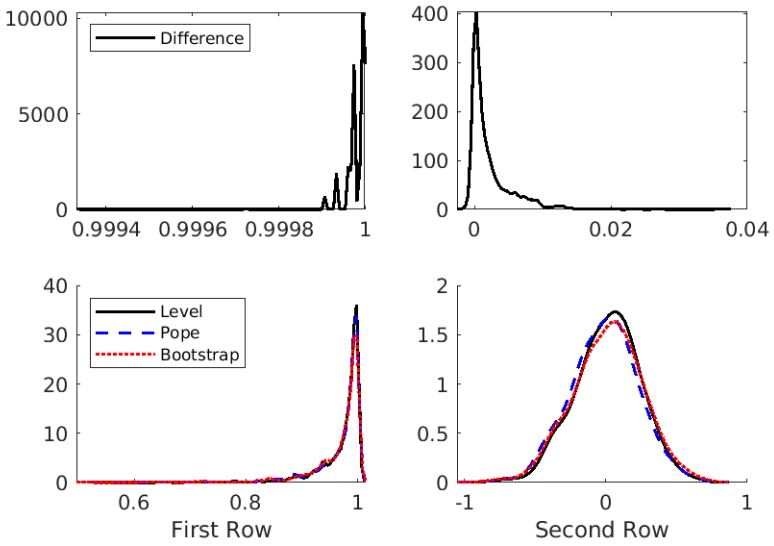
(a) Variable 1



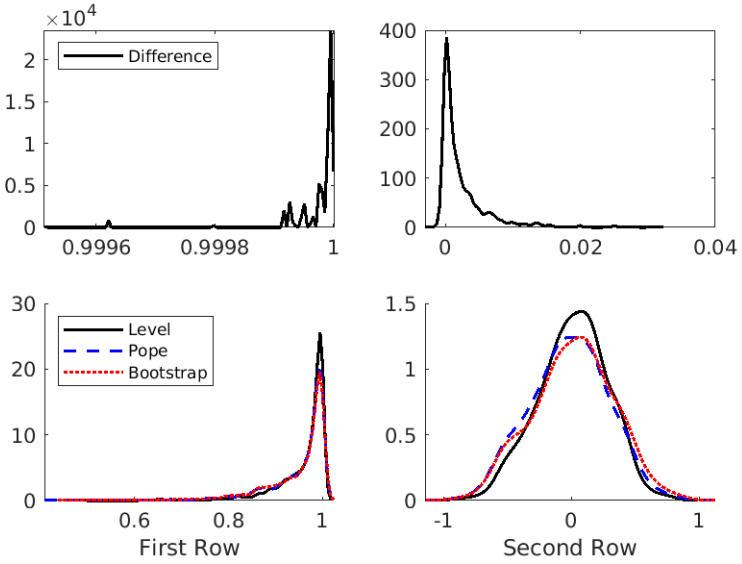
(b) Variable 2

Note: The black solid line represents the true IRF, $\texttt{IRF}_{11}$ and $\texttt{IRF}_{21}$, whereas the dashed line, the red solid line, the blue dashed line, and the red dotted line represent the average estimate of the IRF inherited from the first-difference specification, the level specification, the bias-correction method of Pope, and a bootstrap procedure, respectively.

Figure 4: Distributions of the eigenvector elements associated to the largest eigenvalue (experiment 1)



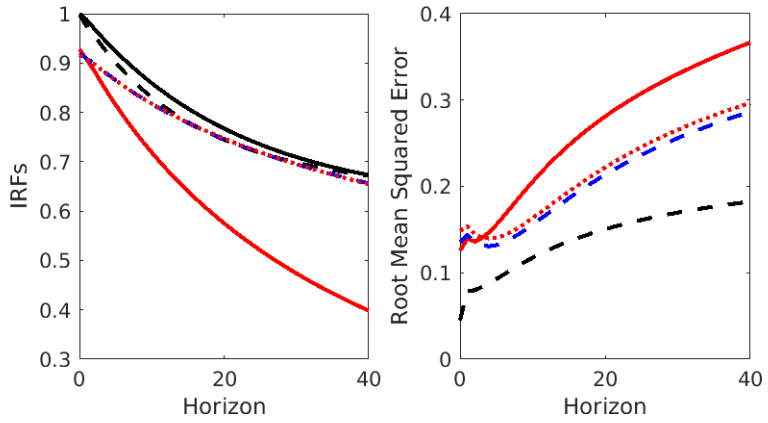
(a) h = 40



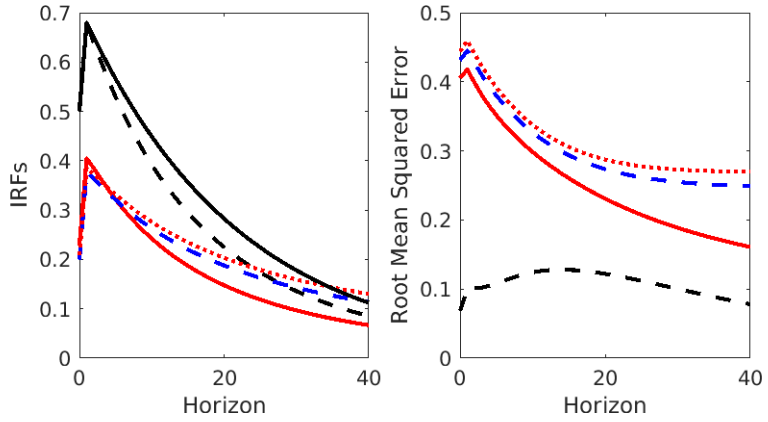
(b) h = 80

Notes: (1) The top (respectively, bottom) panel illustrates the distribution of the two elements $v_{1,1}(h)$ and $v_{2,1}(h)$ of the eigenvector associated with the largest eigenvalue when the forecast error variance horizon is set to $h = 40$ (respectively, $h = 80$). (2) For each horizon, the two upper subfigures depict the distributions of the eigenvector elements (black solid line) when considering the *first-difference* model. The two lower subfigures display the distributions of the OLS-based estimates (black solid line), bias-corrected estimates (blue dashed line), and the bootstrapped estimates (red solid line) of the *level* specification.

Figure 5: Impulse response effects of the first structural shock based on a *non-accumulated* Max-Share identification with $h = 40$ (experiment 2)



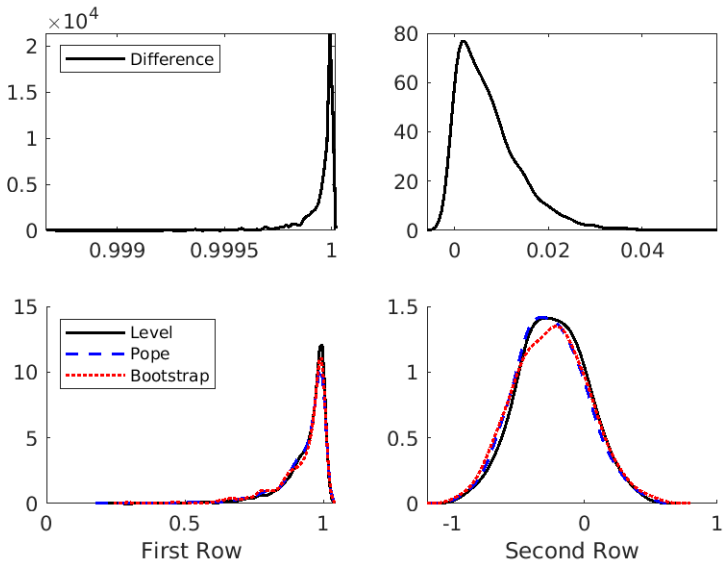
(a) Variable 1



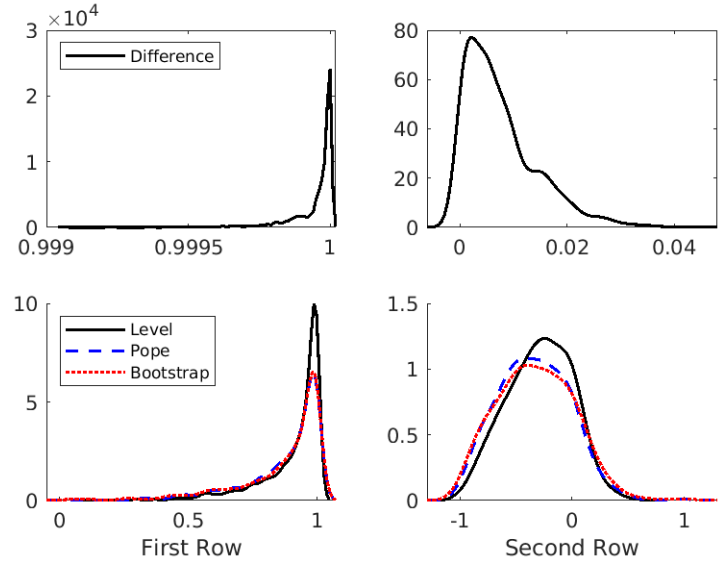
(b) Variable 2

Note: The black solid line represents the true IRF, $\texttt{IRF}_{11,i}$ and $\texttt{IRF}_{21,i}$ for $i = 0, \cdots, 40$, whereas the dashed line, the red solid line, the blue dashed line, and the red dotted line represent the average IRF estimates, $\overline{\texttt{IRF}}_{11,i}(0)$ and$\overline{\overline{\texttt{IRF}}}_{11,i}(0)$, inherited from the first-difference specification, the level specification, the bias-correction method of Pope, and a bootstrap procedure, respectively. (2) The lagged parameters are given by $a_{11} = a_{12} = 0$, $a_{12} = 0.2$, $a_{22} = 0.96$, $\delta = -0.025$.

Figure 6: Distributions of the eigenvector elements associated to the largest eigenvalue (experiment 1)



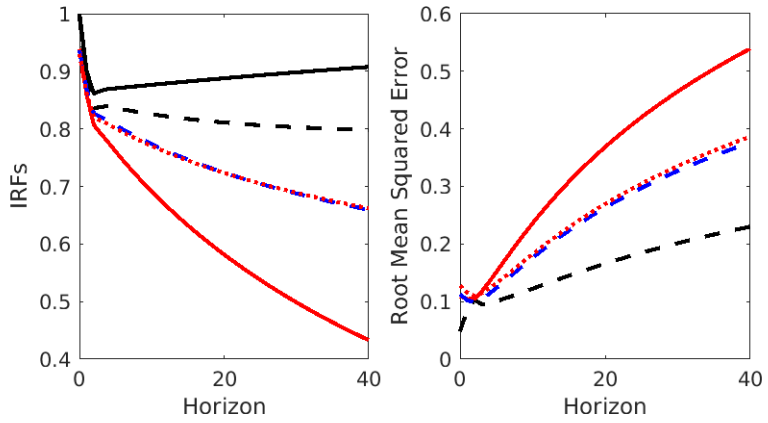
(a) h = 40



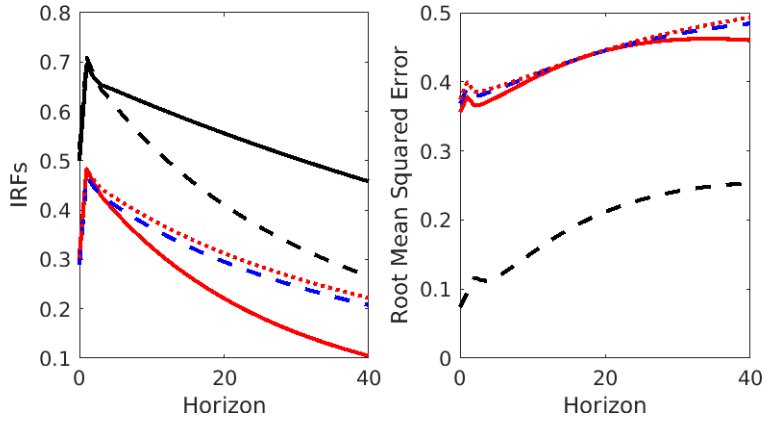
(b) h = 80

Notes: (1) The top (respectively, bottom) panel illustrates the distribution of the two elements $v_{1,1}(h)$ and $v_{2,1}(h)$ of the eigenvector associated with the largest eigenvalue when the forecast error variance horizon is set to $h = 40$ (respectively, $h = 80$). (2) For each horizon, the two upper subfigures depict the distributions of the eigenvector elements (black solid line) when considering the *first-difference* model. The two lower subfigures display the distributions of the OLS-based estimates (black solid line), bias-corrected estimates (blue dashed line), and the bootstrapped estimates (red solid line) of the *level* specification.

Figure 7: Impulse response effects of the first structural shock based on a *non-accumulated* Max-Share identification with $h = 40$ (experiment 3)



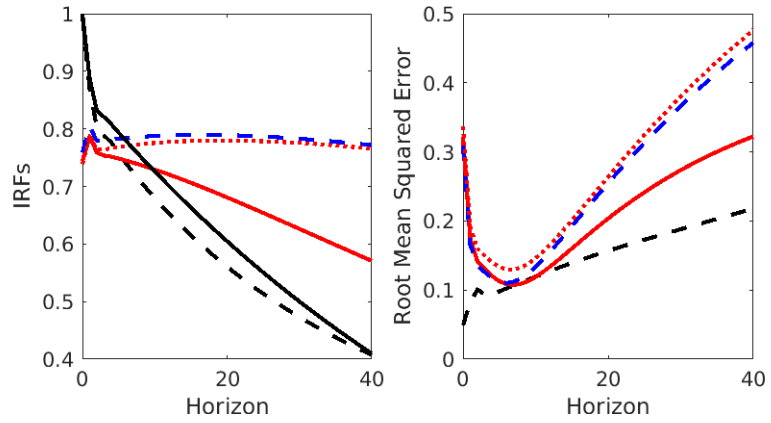
(a) Variable 1



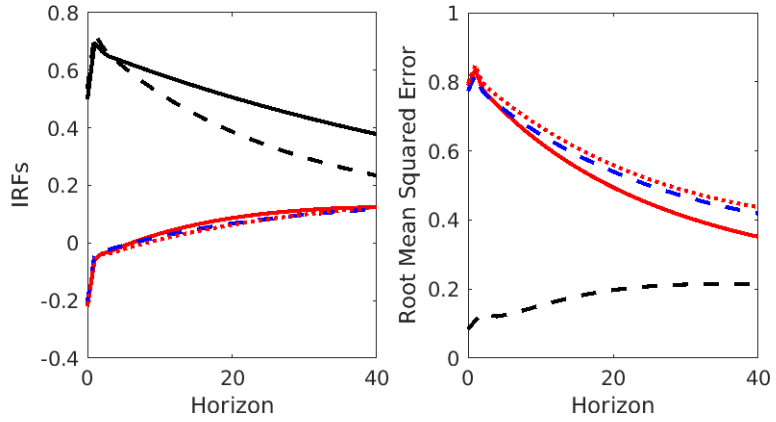
(b) Variable 2

Note: The black solid line represents the true IRF, $\texttt{IRF}_{11,i}$ and $\texttt{IRF}_{21,i}$ for $i = 0, \cdots, 40$, whereas the dashed line, the red solid line, the blue dashed line, and the red dotted line represent the average IRF estimates, $\overline{\texttt{IRF}}_{11,i}(0)$ and$\overline{\overline{\texttt{IRF}}}_{11,i}(0)$, inherited from the first-difference specification, the level specification, the bias-correction method of Pope, and a bootstrap procedure, respectively. (2) The lagged parameters are given by $a_{11} = 0$, $a_{12} = -0.2$, $a_{21} = 0.2$, $a_{22} = 0.99$, $\delta = 0$.

Figure 8: Impulse response effects of the first structural shock based on a *non-accumulated* Max-Share identification with $h = 40$ (experiment 4)



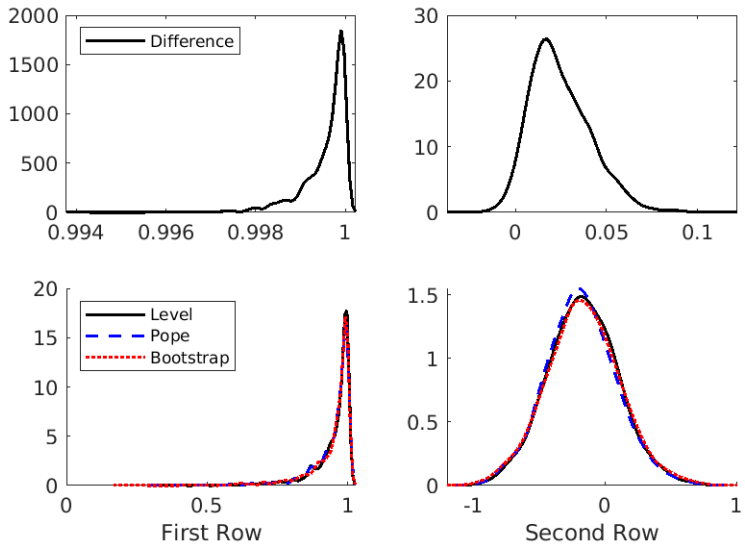
(a) Variable 1



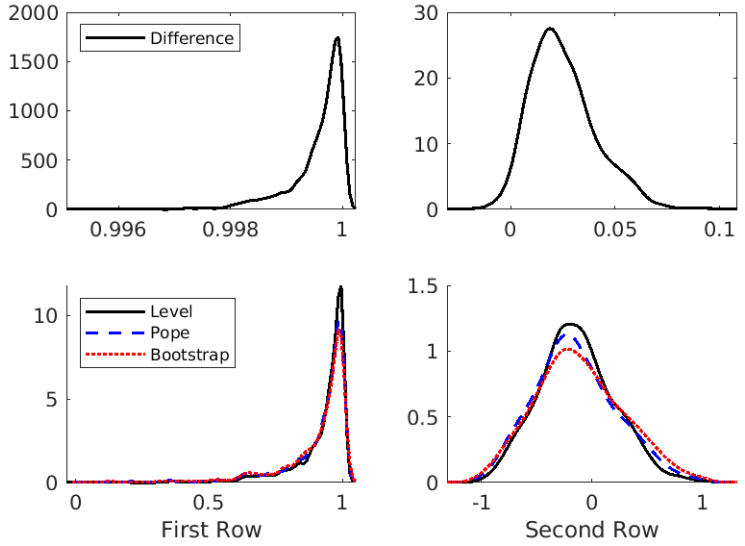
(b) Variable 2

Note: The black solid line represents the true IRF, $\texttt{IRF}_{11,i}$ and $\texttt{IRF}_{21,i}$ for $i = 0, \cdots, 40$, whereas the dashed line, the red solid line, the blue dashed line, and the red dotted line represent the average IRF estimates, $\overline{\texttt{IRF}}_{11,i}(0)$ and$\overline{\overline{\texttt{IRF}}}_{11,i}(0)$, inherited from the first-difference specification, the level specification, the bias-correction method of Pope, and a bootstrap procedure, respectively. (2) The lagged parameters are given by $a_{11} = 0$, $a_{12} = -0.2$, $a_{21} = 0.2$, $a_{22} = 0.99$, $\delta = -0.025$.

Figure 9: Distributions of the eigenvector elements associated to the largest eigenvalue (experiment 3)



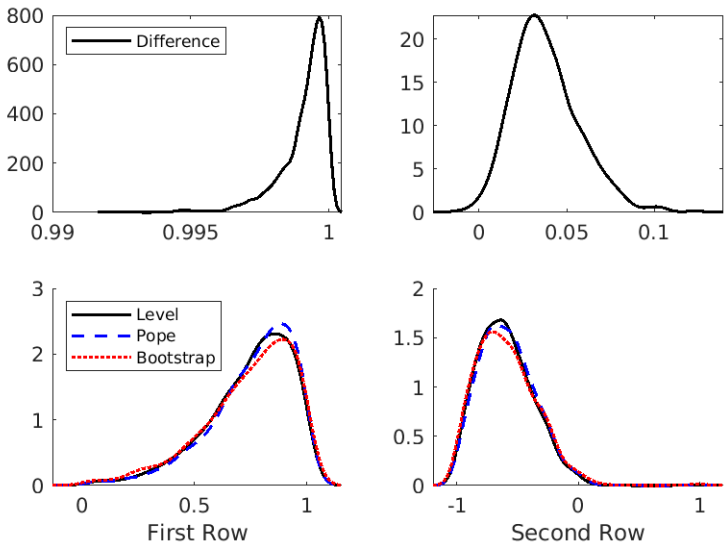
(a) h = 40



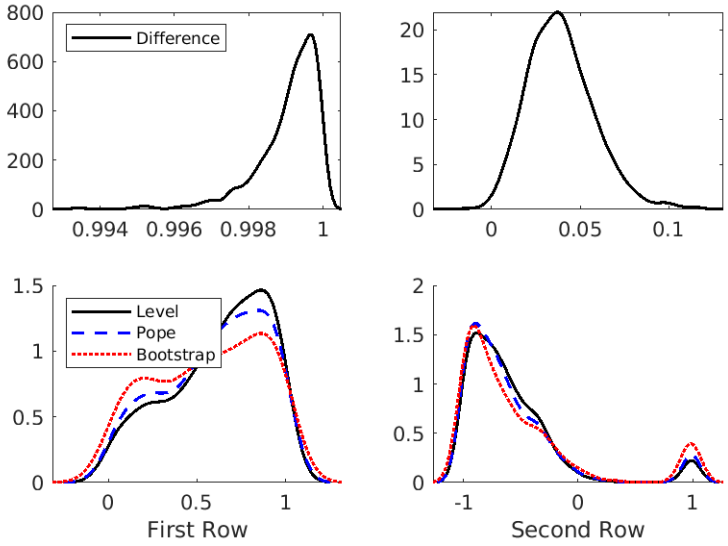
(b) h = 80

Notes: (1) The top (respectively, bottom) panel illustrates the distribution of the two elements $v_{1,1}(h)$ and $v_{2,1}(h)$ of the eigenvector associated with the largest eigenvalue when the forecast error variance horizon is set to $h = 40$ (respectively, $h = 80$). (2) For each horizon, the two upper subfigures depict the distributions of the eigenvector elements (black solid line) when considering the *first-difference* model. The two lower subfigures display the distributions of the OLS-based estimates (black solid line), bias-corrected estimates (blue dashed line), and the bootstrapped estimates (red solid line) of the *level* specification.

Figure 10: Distributions of the eigenvector elements associated to the largest eigenvalue (experiment 4)



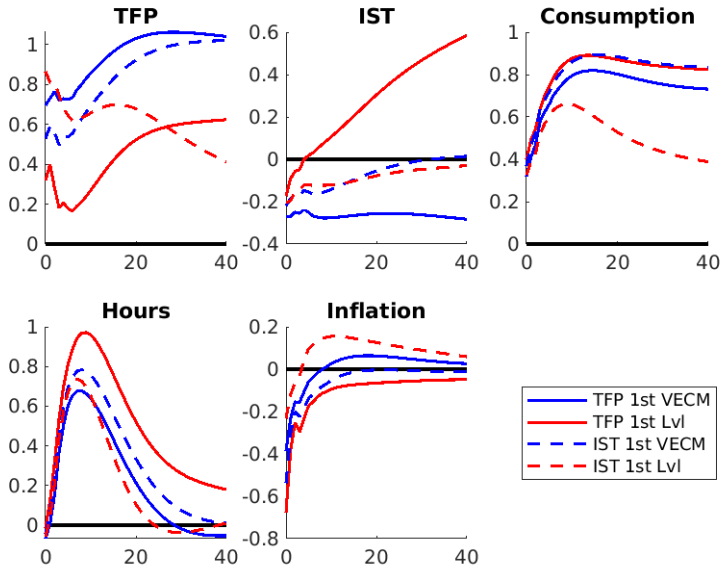
(a) h = 40



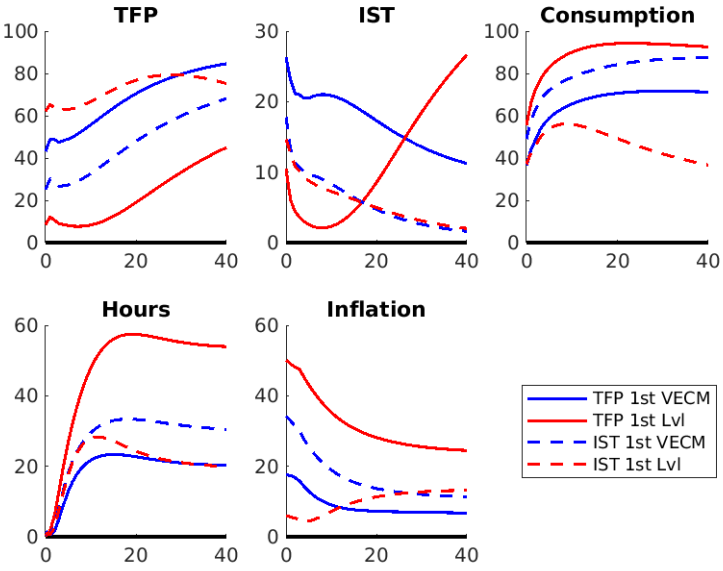
(b) h = 80

Notes: (1) The top (respectively, bottom) panel illustrates the distribution of the two elements $v_{1,1}(h)$ and $v_{2,1}(h)$ of the eigenvector associated with the largest eigenvalue when the forecast error variance horizon is set to $h = 40$ (respectively, $h = 80$). (2) For each horizon, the two upper subfigures depict the distributions of the eigenvector elements (black solid line) when considering the *first-difference* model. The two lower subfigures display the distributions of the OLS-based estimates (black solid line), bias-corrected estimates (blue dashed line), and the bootstrapped estimates (red solid line) of the *level* specification.

Figure 11: TFP shock



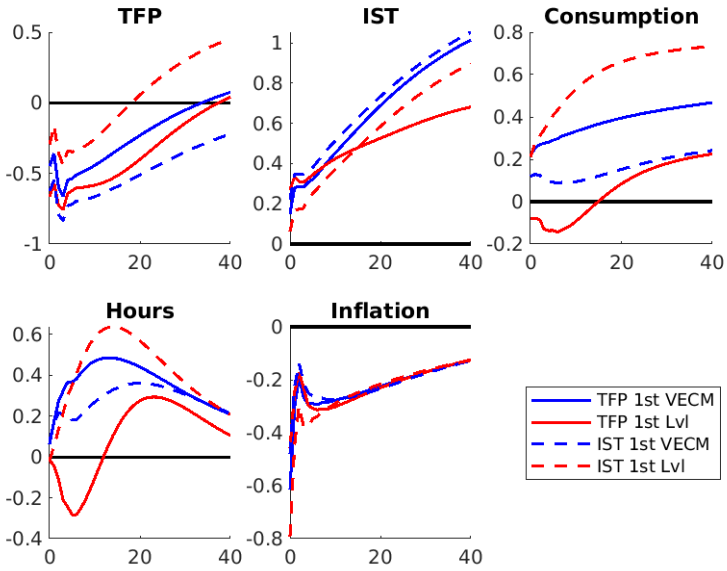
(a) Structural Impulse Responses



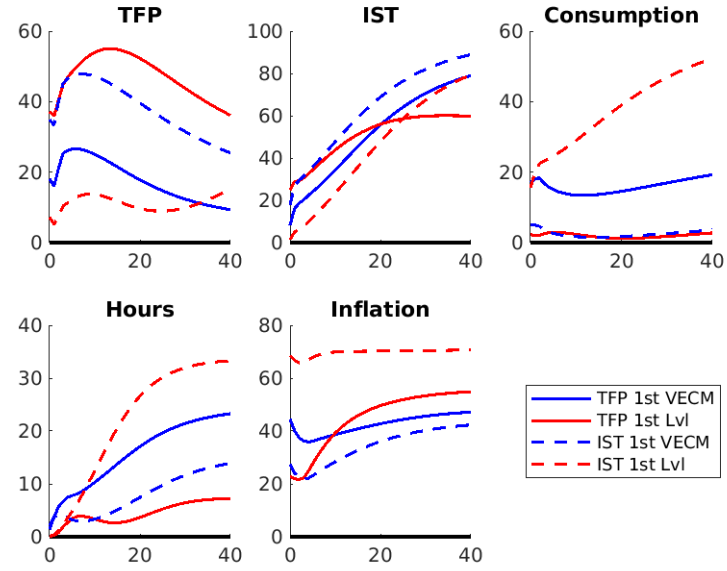
(b) Forecast Error Variance Decomposition

Notes: (1) Red color corresponds to level-based estimates. Blue color corresponds to "first-difference" estimates. (2) A solid line indicates the TFP shock is identified before the IST shock. A dashed line indicates the IST shock is identified before the TFP shock.

Figure 12: IST shock



(a) Structural Impulse Responses



(b) Variance Decomposition

Notes: (1) Red color corresponds to level-based estimates. Blue color corresponds to VECM- based estimates. (2) A solid line indicates the TFP shock is identified before the IST shock. A dashed line indicates the IST shock is identified before the TFP shock.