Non-Gaussian Structural Dynamic Factor Models: Identification and Estimation*

Andrzej Kocięcki† Christian Matthes‡ Michele Piffer§

Abstract

When can a dynamic factor model recover structural shocks without exclusion restrictions, sign restrictions, or external instruments? Macroeconomic data depart systematically from Gaussianity, and these departures carry identifying information. We show how to exploit non-Gaussianity to identify structural shocks and impulse responses in a structural dynamic factor model by establishing global identification under conditions that can hold even in modest cross sections. For Bayesian inference, we develop a Gibbs sampler based on the Student-t scale-mixture representation and a likelihood-preserving normalization.

JEL: C10, C51, C65, E32

Keywords: Dynamic factor model, Non-Gaussianity, Identification, Bayesian inference

*We thank seminar and conference participants at the ESEM in Dublin, the RCEA Bayesian Econometrics Conference (Madrid), the UEA Time Series Workshop (Norwich), 7th Vienna Workshop on High-Dimensional Times Series in Macroeconomics and Finance, BSE Summer Forum Workshop on Advances in Structural Shocks Identification, and the Milan Time Series Seminar for helpful comments.

†University of Warsaw, Faculty of Economic Sciences, Poland; email: akociecki@wne.uw.edu.pl

‡University of Notre Dame, Notre Dame, Indiana, USA; email: cmatthes@nd.edu

§Bank of England and King's Business School, King's College London, UK; email: m.b.piffer@gmail.com

> "This view of economic time series as unreliable or noisy witnesses to a common underlying phenomenon readily suggests a factor or hidden index model."
>
> — Litterman et al. (1984)

1 Introduction

Factor models are powerful dimension-reduction tools and are widely used in the time-series literature. They provide a low-dimensional summary that helps organize the large volume of economic data produced every day.

Three limitations currently constrain the use of factor models. First, consistent estimation of the factor structure typically relies on asymptotics in the cross-sectional dimension, a condition that many datasets do not satisfy. Second, factor models often impose strong restrictions on the autocorrelation and cross-correlation structure of idiosyncratic shocks, allowing only weak correlations in population. Third, although these models are useful descriptive summaries, they do not, on their own, deliver a causal interpretation of the factors. To obtain such interpretations, researchers introduce additional assumptions—such as exclusion restrictions, sign restrictions, or instruments—thereby turning factor models into *structural* factor models.

The first contribution of this paper is to provide a theory of global identification for structural factor models in a non-Gaussian setting. Our main identification result does not require a particular parametric shock distribution. We assume mutually independent non-Gaussian structural shocks whose distributions contain no nondegenerate independent Gaussian summand, a Gaussian component independent of the structural-shock process, and the stated rank and moment conditions. Under these restrictions, the structural impulse responses are identified up to signed permutation and the factor coordinates are identified up to an invertible transformation. Identification is global rather than local. Importantly, identification can be achieved even with a modest cross section, making the model usable in applications with limited cross-sectional coverage. The result allows for a wide class of time-series dynamics in the Gaussian component, including, for example, a vector autoregressive moving-average model—a flexibility that standard factor-model estimators cannot typically accommodate. In addition, as in the vector autoregression literature, non-Gaussianity identifies the structural shocks to the factors without sign restrictions, exclusion restrictions, or external instruments. Statistical identification distinguishes the shock distributions and their responses; assigning economic names to the shocks requires additional information.

The key insight is that higher-order features of the observable distribution distinguish structural dynamics from Gaussian noise, while the restriction on Gaussian summands identifies the scale assigned to each structural shock. Approximate-factor methods typically recover the factor space from second moments as the cross section grows. Our result instead identifies structural impulse responses in a fixed cross section under the stated independence and distributional restrictions.

The second contribution of the paper is to provide a Bayesian inference procedure that makes our identification result usable in applied work. Although the identification result applies to a wide class of non-Gaussian processes, we take a parametric approach and assume Student-t structural shocks to the factors. We then develop a Gibbs sampler based on the Student-t scale-mixture representation. We show that the generalized likelihood-preserving (LP) normalization discussed by Kocięcki et al. (2025) for non-Gaussian structural vector autoregressive (SVAR) models extends to factor models. This further generalizes the likelihood-preserving normalization of Waggoner and Zha (2003), on which Kocięcki et al. (2025) build for non-Gaussian SVAR models. Departures from Gaussianity provide evidence relevant to identification, and shock-independence tests can assess implications of the model. A rejection of Gaussianity, however, does not establish the absence of an independent Gaussian summand, and population identification does not guarantee precise inference when shocks are close to Gaussian (Hoesch et al., 2024).

The third contribution is finite-sample evidence that the model can recover impulse responses and structural shocks well with modest cross sections and Gaussian noise. We assess recovery against the true responses and shock sequences, including specifications with contaminated proxies.

To illustrate the relevance of our identification approach, we provide evidence that non-Gaussianity is pervasive in macroeconomic data. We apply the standard Jarque–Bera test to the residuals from a separate AR(4) regression with an intercept for each of 125 transformed series in the St. Louis Fed's Federal Reserve Economic Data–Quarterly (*FRED-QD*) database. Figure 1 shows that the test rejects Gaussianity for 119 series at the 5% significance level, even after applying the outlier-removal procedure used in the MATLAB code that accompanies *FRED-QD*.[1]

Our approach complements the approximate factor framework of Forni et al. (2000), Stock and Watson (2002a), Stock and Watson (2002b), and Bai (2003). Factor models have a long tradition in economics, dating back to the early contributions to principal component analysis by Pearson (1901); Hotelling (1933). Modern macroeconomic applications of factor models trace back to Sargent and Sims (1977); Geweke (1977); Litterman et al. (1984);

[1]The results are very similar when the residuals come from separate AR(8) regressions with intercepts.

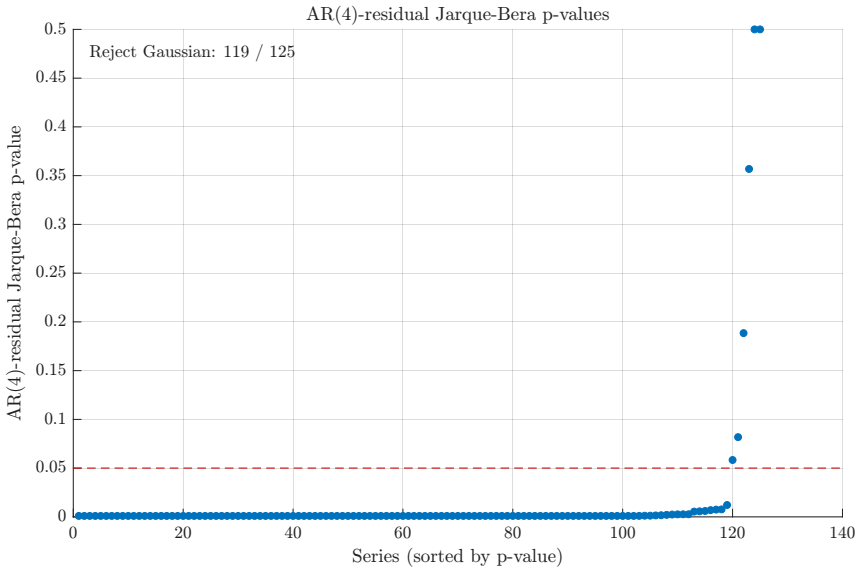


Figure 1: p-values of the Jarque and Bera (1980) test applied to residuals from separate AR(4) regressions with intercepts for 125 FRED-QD series (McCracken and Ng, 2020). The series are flagged for factor analysis in the database and use its recommended transformations and outlier-removal procedure. Series are ordered by p-value; the dashed line marks the 5% significance level.

Stock and Watson (2002b) and have since expanded in many directions; for example, to Factor-Augmented Vector Autoregressions (FAVARs) introduced by Bernanke et al. (2005). Dynamic factor models have also been used to draw causal conclusions. Early descriptions of structural dynamic factor models appear in Stock and Watson (2005) and Giannone et al. (2005). Since then, these models have been used in many applications, including Forni et al. (2009), Forni and Gambetti (2010), Stock and Watson (2012), Forni et al. (2014), Luciani (2015), Bjørnland and Thorsrud (2016), Forni et al. (2025a), and Forni et al. (2025c). A review appears in the *Handbook of Macroeconomics* (Stock and Watson, 2016).

As emphasized by Stock and Watson (2016), structural dynamic factor models (SDFMs) can in principle overcome shortcomings of SVARs. First, by including more observables than structural shocks, SDFMs make fundamentalness less fragile. Forni et al. (2009) show that, in large-dimensional factor models, heterogeneous impulse responses make nonfundamentalness nongeneric; in their setting, structural shocks can therefore be recovered from current and past common components. SDFMs are consequently less prone than small SVARs to invertibility problems or deformations (Canova and Ferroni, 2022; Hansen and Sargent, 1991; Lippi and Reichlin, 1994; Forni et al., 2019). Second, because SDFMs include idiosyncratic components that can accommodate measurement error, they naturally address error-in-variables problems. Measurement error can lead, for example, to contaminated or biased impulse responses; see, e.g., Lippi (2023) and Forni et al. (2025b). Traditionally, SVAR models were

used to assess fit and provide reference impulse responses for dynamic stochastic general equilibrium (DSGE) models. Forni et al. (2025c) warn against using SVARs for validating DSGE models and argue convincingly that SDFMs are better suited to this task. Postulating a relatively small number of common factors amounts to assuming a small number of structural shocks; thus, the methodology is consistent with a common critique of macroeconomic models, namely that they include too many (atheoretical) shocks.

In a comment on Stock and Watson (2012), Sims (2012) highlighted that neglecting non-Gaussianity is a serious drawback of the structural factor model in Stock and Watson (2012). Building on this insight, our identification argument is closely related to independent component analysis (ICA). Classical ICA uses mutual independence and non-Gaussianity to identify a static mixing matrix up to scale, sign, and permutation (Comon, 1994). Probabilistic independent component analysis extends this logic to noisy latent-variable models by combining non-Gaussian components with Gaussian measurement noise (Beckmann and Smith, 2004). The approach in this paper has the same basic structure, but the structure is dynamic rather than static. After eliminating the latent factors, the observables can be viewed as the sum of a dynamic mixture of non-Gaussian structural shocks and a Gaussian component. Thus, the paper is not merely an application of ICA after factor extraction. Instead, it identifies the dynamic structural transfer function, and therefore impulse responses, in the noisy factor model itself.

Most empirical work that uses non-Gaussianity for structural identification in macroeconomics has focused on vector autoregressions (VARs). For example, Brunnermeier et al. (2021) investigate the interaction of financial and monetary policy shocks and their impact on the economy; Braun (2023) studies identification of supply and demand shocks in oil-price fluctuations; Jarociński (2024) disentangles the structural shocks underlying financial-market reactions to Federal Open Market Committee announcements; Anttonen et al. (2023) tracks the effects of monetary policy shocks, see also Anttonen et al. (2024); and Keweloh et al. (2025) estimate the effects of exogenous changes in tax revenues and government spending. The literature exploiting non-Gaussianity in empirical macroeconomics and finance is growing rapidly; see Lewis (2025), which appears to be the first review article on the topic. Non-Gaussian SVARs use independence and distributional restrictions to identify structural shocks and their impacts up to sign and ordering.

Sentana and Fiorentini (2001) provide an important complementary route to factor identification. They study static factor models in which uncorrelated common factors are conditionally heteroskedastic, and they show that time variation in the diagonal factor covariance matrix supplies second-moment restrictions that can identify the factor loadings. Bonhomme and Robin (2009) are closer to the ICA logic: in a cross-sectionally i.i.d. setting, they use

higher-order cumulants implied by mutual independence and non-Gaussianity of the factors to identify the loading matrix, while allowing structured dependence in measurement errors. Our contribution is different in both object and scope. We study a structural dynamic factor model, allow dynamic loadings and a Gaussian vector autoregressive moving-average component, and target observable structural impulse responses and shock distributions, with the contemporaneous impact matrix identified after fixing the factor coordinates. The innovation is therefore not merely another way to estimate static loadings; it is a finite-cross-section identification result for structural dynamics.

The rest of the paper is organized as follows. Section 2 outlines the model. Section 3 presents and discusses our key identification result. Section 5 develops our Bayesian estimation approach. Section 6 reports Monte Carlo experiments. Section 7 decomposes Gaussian and non-Gaussian variation in a small US macroeconomic panel. Section 8 concludes. The Appendix gives the proof of the main identification result. The Online Appendix contains supporting proofs, posterior simulation details, and additional results.

2 The model setup

We study the following structural dynamic factor model (SDFM):

$$y_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \cdots + \Lambda_s f_{t-s} + u_t, \tag{1}$$

$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_p f_{t-p} + B\epsilon_t, \tag{2}$$

where y_t is an $n \times 1$ vector of observables; f_t is a $q \times 1$ vector of unobserved factors ("dynamic" because factors can enter the observation equation with lags); u_t is an $n \times 1$ vector representing the Gaussian component; and ϵ_t is a $q \times 1$ vector of structural shocks. For each $i = 0, \ldots, s$, Λ_i is $n \times q$ (factor loadings). For each $i = 1, \ldots, p$, Φ_i is $q \times q$. The matrix B is nonsingular of dimension $q \times q$. We assume $n \geq q$ (and allow $n = q$), in contrast to the usual factor-analysis setting $n \gg q$.[2]

Several familiar settings lead to (1). First, if all variables in a structural vector autoregression (SVAR) for y_t are measured with error, then (1) is the natural measurement-error representation. Set $\Lambda_0 = I_q$, $\Lambda_i = 0$ for $i > 0$, and $u_t \sim$ i.i.d. $N(0, \Omega)$. Then f_t follows an SVAR(p), and $y_t = f_t + u_t$.[3] More generally, an SVAR with measurement error can be embedded in a larger SDFM; see Bai and Wang (2016). Second, when at least one factor is

[2]The model (1) implicitly assumes that the common component $\chi_t := \Lambda_0 f_t + \Lambda_1 f_{t-1} + \cdots + \Lambda_s f_{t-s}$ spans a finite-dimensional space.

[3]The model (1) accommodates the case in which only a subset of SVAR variables is measured with error. In our setup, set $n = q$ and set the corresponding entries of the Gaussian component to zero.

observed, the SDFM nests the factor-augmented VAR (FAVAR) of Bernanke et al. (2005).

Let us define $\Phi(L) := I_q - \Phi_1 L - \cdots - \Phi_p L^p$, where L is the backward shift (lag) operator (other polynomials in L will be defined analogously).

Assumption 1. *(Errors), a) u_t follows a causal, stationary and invertible Gaussian VARMA process, $\Pi(L)u_t = \Theta(L)e_t$, where the Gaussian innovations satisfy $e_t \sim$ i.i.d. $N(0, \Omega)$ and Ω is positive definite; b) ϵ_t is i.i.d., mean zero, and non-Gaussian, with mutually independent components*[4] *that need not be identically distributed; c) the processes $\{\epsilon_t\}_{t\in\mathbb{Z}}$ and $\{e_t\}_{t\in\mathbb{Z}}$ are independent; d) each structural shock has no normal component: if $\epsilon_{j,t} \stackrel{d}{=} X + G$, where X and G are independent and G is normally distributed, then $\mathrm{Var}(G) = 0$. This restriction applies to every admissible representation.*

Independence and non-Gaussianity provide the information used to distinguish the structural shocks, while part (d) prevents Gaussian variation from being reassigned between those shocks and the Gaussian component.

Non-Gaussianity alone does not exclude a normal component. Let $\{W_t\}$, $\{G_t\}$, and $\{H_t\}$ be mutually independent i.i.d. sequences, with W_t taking the values -1 and 1 with equal probability and $G_t, H_t \sim N(0,1)$. Then

$$y_t = W_t + (G_t + H_t) = \sqrt{2}\,\widetilde{\epsilon}_t + H_t, \qquad \widetilde{\epsilon}_t = \frac{W_t + G_t}{\sqrt{2}}.$$

Both W_t and $\widetilde{\epsilon}_t$ are non-Gaussian with variance one, but the two representations imply different responses to a unit-standard-deviation structural shock. Part (d) excludes the second representation because $\widetilde{\epsilon}_t$ contains an independent normal component.

This restriction also appears in the static factor model $X = \beta U + \varepsilon$ discussed by Kagan et al. (1973, p. 315). With a full-column-rank loading matrix, mutually independent non-Gaussian factors having no normal components, and independent multivariate normal errors, the loading matrix and factor distributions are unique up to permutation and scale, and the error distribution is unique, after fixing means. Part (d) imposes the corresponding restriction on the structural innovations of our dynamic model. The Student-t distributions used below satisfy this restriction.

By Assumption 1a, we can write $u_t = [\Pi(L)]^{-1}\Theta(L)e_t := \Upsilon(L)e_t = \sum_{i=0}^{r} \Upsilon_i e_{t-i}$, with normalization $\Upsilon_0 = I_n$ and $r = \infty$ (unless u_t is a pure MA process, in which case r may be finite). Because the structural shocks are non-Gaussian, we refer to (1) as the non-Gaussian

[4]The mutual-independence restriction can also be checked using the recovered shocks. In non-Gaussian SVAR settings, Montiel Olea et al. (2022) discuss bootstrap tests of shock independence based on cross-correlations of squared estimated shocks, and Amengual et al. (2024) develop specification tests that compare the joint distribution of estimated shocks with the product of their marginal distributions.

SDFM (NG-SDFM). Gaussian innovations and independence of the two innovation processes are central to identification.[5]

An additional finite-dimensional non-Gaussian component can be absorbed into an expanded factor vector when its innovations satisfy the same distributional restrictions and the expanded system satisfies the rank and minimality conditions below. In particular, the total number of non-Gaussian innovations cannot exceed n; Online Appendix O.3.2 gives the algebra.

We allow each component of ϵ_t to have its own distribution. We require that sufficiently high moments exist; in particular, we assume at least a finite fourth moment:[6]

Assumption 2. *(Moments), There exist common integers $r \geq 3$ and even $a \geq r$ such that every structural-shock component has a finite absolute a-th moment and a nonzero r-th cumulant. The same orders r, a apply to both admissible representations compared in the identification result.*

The proof uses finite, nonzero higher-order cumulants—quantities capturing departures from Gaussianity—to distinguish structural shocks, because independent Gaussian noise contributes nothing to these cumulants. For estimation, we assume independent Student-t structural shocks with finite degrees of freedom $\nu_i > 4$ for each $i = 1, \ldots, q$. Their fourth cumulants are nonzero, so $r = a = 4$ satisfies Assumption 2. They also have no independent Gaussian summand, as shown in Lemma 2; thus part (d) of Assumption 1 leaves the Student-t specification unchanged. More generally, Assumption 2 allows skewed shock distributions with a common nonzero third cumulant and finite fourth moments; see Gouriéroux et al. (2020). The moment condition requires each component to be non-Gaussian, since all Gaussian cumulants of order at least three vanish. This is stronger than in the non-Gaussian SVAR literature, where (at most) one shock may be Gaussian; see, e.g., Lanne et al. (2017).

We normalize the variances of the structural shocks:

Assumption 3. *(Scale Normalization), Individual components of ϵ_t obey $var(\epsilon_{i,t}) = 1$, for $i = 1, ..., q$.*

[5]The Gaussian innovations e_t could also be given a structural interpretation by imposing standard identifying restrictions. For example, one could write $e_t = A_0^{-1}\eta_t$, with $\eta_t \sim N(0, I_n)$, and identify A_0^{-1} using recursive, exclusion, sign, or external-instrument restrictions. This additional step is separate from our identification result, which uses non-Gaussianity to identify the structural shocks ϵ_t and their impulse responses.

[6]Assuming higher-order moments for shocks is not unique to our non-Gaussian setup; it is common in the DFM literature; see, e.g., Doz et al. (2011), Forni et al. (2015), and Barigozzi et al. (2024).

This normalization fixes the scale of each shock, so that its impulse responses measure the effects of a one-standard-deviation innovation.

Assumption 4. *(Stationarity), $det(\Phi(z)) \neq 0$ for all $z \in \mathbb{C}$ (a set of complex numbers) such that $|z| \leq 1$.*

This condition makes the effects of past shocks decay over time, allowing us to express the factors and observables as convergent sums of current and past innovations.
We find it useful to write (1) in state-space form (i.e. static form of the DFM, see e.g. Stock and Watson (2005)). To this end we need to distinguish two cases. Assuming $s+1 \leq p$, our NG-SDFM can be written as:

$$\underbrace{\begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{bmatrix}}_{s_t} = \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I_q & 0 & \cdots & 0 \\ \vdots & \ddots & \dots & \vdots \\ 0 & \dots & I_q & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-p} \end{bmatrix}}_{s_{t-1}} + \underbrace{\begin{bmatrix} B \\ 0_{q(p-1)\times q} \end{bmatrix}}_{\mathbf{B}} \epsilon_t, \tag{3}$$

$$y_t = \underbrace{\begin{bmatrix} \Lambda_0 & \Lambda_1 & \dots & \Lambda_s & 0 & \dots & 0 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{bmatrix}}_{s_t} + u_t. \tag{4}$$

On the other hand, when $s+1 > p$, the state-space form is given by:

$$\underbrace{\begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-s} \end{bmatrix}}_{s_t} = \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p & 0 & \dots & 0 & 0 \\ I_q & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & I_q & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_q & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & I_q & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-s-1} \end{bmatrix}}_{s_{t-1}} + \underbrace{\begin{bmatrix} B \\ 0_{qs\times q} \end{bmatrix}}_{\mathbf{B}} \epsilon_t, \tag{5}$$

$$y_t = \underbrace{\begin{bmatrix} \Lambda_0 & \Lambda_1 & \dots & \Lambda_s \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-s} \end{bmatrix}}_{s_t} + u_t. \tag{6}$$

The state vector s_t stacks the current factor and its lags and thus contains the "static" factors.

Regardless of whether A is defined as in (3) or (5), our Assumption 4 is equivalent to the assumption that A is stable (i.e. the absolute value of the largest eigenvalue of A is less than 1).[7] In line with the above development, let us define:

$$\tau = max\{p, s+1\}, \tag{7}$$

which is an index needed to mark the appropriate dimension of A matrix from either (3) or (5).

Since A is stable, iterating the state equation and substituting into the observation equation yields the moving-average (MA) representation of the NG-SDFM:

$$\begin{aligned} y_t = \sum_{i=0}^{\infty} CA^i \mathbf{B}\epsilon_{t-i} + u_t &:= \sum_{i=0}^{\infty} \Psi_i \epsilon_{t-i} + \sum_{i=0}^{r} \Upsilon_i e_{t-i}, \\ &:= \Psi(L)\epsilon_t + \Upsilon(L)e_t, \end{aligned} \tag{8}$$

where possibly $r = \infty$. Stability of the rational filters implies that $\{\Psi_i\}$ and $\{\Upsilon_i\}$ are absolutely summable and hence square-summable. Together with Assumption 2, this ensures finite second moments of y_t. Taking into account (8), our model may be thought as a dynamic, parametric version of the semi-parametric setup of Bonhomme and Robin (2009).

We define the $n \times q$ rational transfer function

$$\Psi(z) := C(I_{\tau q} - zA)^{-1}\mathbf{B}, \tag{9}$$

where $z \in \mathbb{C}$ is a complex variable used in place of the backward-shift operator L.[8] We say that $\Psi(z)$ is left-invertible if there exists a (rational) $q \times n$ matrix $\Psi_{\text{left}}(z)$ such

[7] The case covered by (3) is the standard result, see e.g. Brockwell and Davis (2006) p. 424. The case in (5) may be obtained by slight modification of the proof of Proposition 10.1 in Hamilton (1994), pp. 285-286.

[8] A transfer function is rational if each element is the ratio of polynomials in z with real coefficients. Let ρ denote the maximum eigenvalue (in modulus) of A. Since A is stable, $\rho < 1$. For all $|z| < \frac{1}{\rho}$, (i) $\det(I_{\tau q} - zA) \neq 0$ and (ii) $(I_{\tau q} - zA)^{-1} = \sum_{j=0}^{\infty} A^j z^j$. Thus there is a one-to-one mapping between the transfer function (9) and $CA^i\mathbf{B}$ for $i = 0, 1, \dots$; see, e.g., Kocięcki and Kolasa (2023), p. 21. This motivates working with (9) rather than the MA coefficients.

that $\Psi_{\text{left}}(z)\Psi(z) = I_q$. We call $\Psi_{\text{left}}(z)$ a left inverse. Define the normal rank as $\text{normrank}(\Psi(z)) := \max\{\text{rank}(\Psi(z)) \mid z \in \mathbb{C}\}$. Except for finitely many points, $\text{rank}(\Psi(z))$ equals $\text{normrank}(\Psi(z))$. An $n \times q$ $(n \geq q)$ rational matrix $\Psi(z)$ is left-invertible if and only if $\text{normrank}(\Psi(z)) = q$; see Trentelman et al. (2001), p. 31. Hence we assume:

Assumption 5. $\text{rank}(\Psi(z)) = q$ *for all* $z \in \mathbb{C}$ *such that* $|z| = 1$.

At every frequency, the structural shocks have linearly independent effects on the common component, so no nonzero combination of them cancels across all observables. Assumption 5 ensures that a rational left inverse can be chosen with no poles on the unit circle. It therefore admits a Laurent expansion—a series in integer powers of z, including negative powers—that converges in a neighborhood of the unit circle and has absolutely summable coefficients. In lag notation,

$$\Psi_{\text{left}}(L) = \sum_{j=-\infty}^{\infty} \Theta_j L^j, \qquad \Psi_{\text{left}}(L)\Psi(L) = I_q.$$

Negative powers of L correspond to leads, so this inverse may use both past and future observations. The left inverse need not be unique.[9] One can construct such an inverse via the Smith–McMillan factorization; see, e.g., Kailath (1980) and Anderson and Deistler (2008).

The full-rank condition on the unit circle supplies the stable two-sided left inverse used in the identification proof. A related rank condition appears in Chan et al. (2006). It generalizes assumptions used in the square case $(n = q)$; see, e.g., Chen et al. (2017), Plagborg-Møller (2019), Gouriéroux et al. (2020), and Funovits (2024).

We impose two further rank conditions:

Assumption 6. Λ_0 *has full column rank.*

Every structural shock must affect at least one observable on impact, which, together with causality, rules out equivalent representations that shift its entire response to earlier or later dates. In the MA representation (8),

$$\Psi_0 = C\mathbf{B} = \Lambda_0 B. \tag{10}$$

Since B is nonsingular, Assumption 6 holds if and only if Ψ_0 has full column rank. The

[9] If we strengthen Assumption 5 to $\text{rank}(\Psi(z)) = q$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$, then the identification results below continue to hold and one can take a one-sided left inverse, $\Psi_{\text{left}}(L) = \sum_{j=0}^{\infty} \Theta_j L^j$. See Anderson and Deistler (2008), Forni et al. (2009), and Forni et al. (2015). A one-sided left inverse need not be unique either; see, e.g., Anderson and Deistler (2008).

second is a standard observability condition. Define

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{\tau q-1} \end{bmatrix}. \tag{11}$$

Assumption 7. *$\mathcal{O}$ has full column rank.*

This condition rules out combinations of current and lagged factors that never affect any observable, even at later dates, thereby excluding state directions that are irrelevant for observable dynamics.
Assumption 7 is part of the classic minimality assumption for linear systems; see, e.g., Kailath (1980).

2.1 Further discussion of Assumptions 5, 6, and 7

Define $\Lambda = [\Lambda_0 \vdots \Lambda_1 \vdots \ldots \vdots \Lambda_s]$. A simple sufficient condition for Assumptions 5 and 6 is full column rank of Λ:

Proposition 1. *Assume that Λ has full column rank. Then Assumptions 5 and 6 hold.*

We also use the system matrix (see, e.g., Anderson and Deistler, 2008):

$$M(z) := \begin{bmatrix} I_{\tau q} - zA & -\mathbf{B} \\ C & 0_{n\times q} \end{bmatrix}. \tag{12}$$

Studying (12) is convenient for left-invertibility of the transfer function (9). By Lemma O.1 in Online Appendix O.2.1, the transfer function is left-invertible if and only if $\text{normrank}(M(z)) = q(\tau+1)$, i.e., $M(z)$ has full column rank for almost all $z \in \mathbb{C}$.

Online Appendix O.2.1 shows that if Λ has full column rank, then $M(z)$ has full column rank for all $z \in \mathbb{C}$. Beyond implying Assumption 5, this condition yields a particularly simple left inverse of $\Psi(z)$: a finite polynomial in z. Equivalently, $\Psi_{\text{left}}(L) = \sum_{j=0}^{k} \Theta_j L^j$ for some finite k; see Theorems 8.28 and 7.17 in Trentelman et al. (2001). Premultiplying (8) by $\Psi_{\text{left}}(L)$ gives $\epsilon_t = \sum_{j=0}^{k} \Theta_j y_{t-j} - \sum_{i=0}^{r+k} D_i e_{t-i}$. For example, in case (5)–(6) one may take $\Psi_{\text{left}}(L) = [B^{-1} \vdots 0](I_{\tau q} - AL)(\Lambda'\Lambda)^{-1}\Lambda'$, which uses only one lag.

As noted by Anderson and Deistler (2008) (see also Forni et al., 2009), if the transfer function (9) is "tall" i.e. $n > q$, and since our underlying state-space is minimal (see

Assumption 7 and Lemma 1 in Appendix A.1), for generic values of (C, A, B), the system matrix (12) has full column rank for all $z \in \mathbb{C}$. By the results in Online Appendix O.2.1, it "generically" implies our Assumption 5 and existence of a finite polynomial left inverse of $\Psi(z)$ (for the latter see Trentelman et al., 2001, Anderson and Deistler, 2008). It suggests that when working with a generic SDFM one should not worry about Assumption 5. This insight is increasingly used in modern DFM literature, see e.g. Forni et al. (2015), Barigozzi et al. (2024), Forni et al. (2025b). However we prefer to keep Assumption 5 since our setup includes the square case (i.e. $n = q$), in which the nonsingularity of the system matrix (12) for all $z \in \mathbb{C}$ is nongeneric (unusual) as demonstrated below, see also Anderson and Deistler (2008), Forni et al. (2009). The next proposition sheds some light on "genericity" of the full column rank assumption for (12) for all $z \in \mathbb{C}$ in the model (1):

Proposition 2. *If B is nonsingular, the system matrix (12) has full column rank if and only if $\Lambda_0 + z\Lambda_1 + z^2\Lambda_2 + \cdots + z^s\Lambda_s$ has full column rank.*

The above proposition reveals the surprising feature that the full rank condition for the system matrix does not depend on the structure of Φ_i's. Next, by Assumption 6, one obvious consequence is that if $s = 0$, then a) the system matrix has full column rank for all $z \in \mathbb{C}$, b) Assumption 5 holds and c) there exists a left inverse of $\Psi(z)$, which is a finite polynomial in z.

To illustrate the above points consider the SDFM with $s = 1$. Then the system matrix (12) has full column rank if and only if $\Lambda_0 + z\Lambda_1$ has full column rank. In the square case, Λ_0, Λ_1 are $n \times n$ matrices and it amounts to checking $det(\Lambda_0 + z\Lambda_1) \neq 0$. By our Assumption 6 we are interested in finding $z \in \mathbb{C}$ such that $det(I_n + z\Lambda_0^{-1}\Lambda_1) = 0$. The solutions are the negative reciprocals of nonzero eigenvalues of $\Lambda_0^{-1}\Lambda_1$. Hence for generic Λ_0, Λ_1, $det(\Lambda_0 + z\Lambda_1) \neq 0$ for almost all $z \in \mathbb{C}$. It follows the system matrix (12) has full column rank for almost all $z \in \mathbb{C}$, but not all $z \in \mathbb{C}$. In such a case a left inverse of $\Psi(z)$ will be two-sided, in general. In the "tall" case ($n > q$), Λ_0, Λ_1 are $n \times q$ matrices. To make the point simpler suppose that $n = 2$ and $q = 1$. Then $rank(\Lambda_0 + z\Lambda_1) \leq 1$. So the question is whether $\Lambda_0 + z\Lambda_1$ can drop the rank, which means it contains only zeros. For generic Λ_0, Λ_1 this is impossible, since it would amount to finding a single $z \in \mathbb{C}$ that solves two linear equations. Hence $\Lambda_0 + z\Lambda_1$ has full column rank for all $z \in \mathbb{C}$ and the same is true for the system matrix (12). In this case there exists a left inverse of $\Psi(z)$, which is a finite polynomial in z.

In the case of Assumption 7, we do need to distinguish between two state-space models:

Proposition 3. *Assume that $\tau = max\{p, s + 1\} = s + 1 \geq p$. Provided that Λ has full column rank, Assumption 7 holds.*

On the other hand:

Proposition 4. *Assume that* $\tau = \max\{p, s+1\} = p > s+1$, Λ *has full column rank, and* Φ_p *is nonsingular. Then Assumption 7 holds.*

Full column rank of Λ appears explicitly in Bai and Wang (2014). It is sufficient for Assumptions 5 and 6, but its implication for observability depends on the lag orders. When $p \leq s+1$, Proposition 3 applies; when $p > s+1$, Proposition 4 additionally requires nonsingular Φ_p. It should be emphasized that Assumptions 5 and 7 can hold even without full column rank of Λ.

The proofs of these propositions are in Online Appendix O.2.

3 Identification

3.1 Main result

Let $\theta = (\Lambda_0, \ldots, \Lambda_s, \Phi_1, \ldots, \Phi_p, B, \theta_u)$ collect the parameters in (1), where $\theta_u = (\Upsilon, \Omega)$ specifies the normalized causal, invertible moving-average representation of the Gaussian component u_t. The pair (Υ, Ω) determines the joint distribution of u_t across dates. For an unrestricted VARMA representation, identifying the autoregressive and moving-average coefficient matrices separately requires additional restrictions. The structural-shock distributions are not included in θ because we allow them to differ across observationally equivalent models.

Let $\bar{\theta}$ be an alternative parameter vector. Let ϵ_t and $\bar{\epsilon}_t$ be i.i.d. vectors of mutually independent, mean-zero, non-Gaussian shocks that satisfy Assumption 1b–d and the common-order condition in Assumption 2, but possibly have different distributions.[10]

We say that (θ, ϵ_t) and $(\bar{\theta}, \bar{\epsilon}_t)$ are *observationally equivalent*, written $(\theta, \epsilon_t) \sim (\bar{\theta}, \bar{\epsilon}_t)$, if they induce the same distribution for the observables $\{y_t\}$.

We can now state our main result:

Theorem 1. *Fix* n, q, p, s. *Let both representations satisfy Assumptions 1–7. Then* $(\theta, \epsilon_t) \sim (\bar{\theta}, \bar{\epsilon}_t)$ *if and only if:*

1) $\bar{\Lambda}_i = \Lambda_i T^{-1}$ *for each* $i = 0, \ldots, s$;

2) $\bar{\Phi}_i = T\Phi_i T^{-1}$ *for each* $i = 1, \ldots, p$;

3) $\bar{B} = TBP_s$;

[10] Since $\bar{\epsilon}_t$ and ϵ_t have mutually independent components, their probability distribution functions are the same, written $\bar{\epsilon}_t \overset{d}{=} \epsilon_t$, if the corresponding marginal distribution functions are the same. It follows that $\bar{\epsilon}_t$ and ϵ_t have different distributions if at least one corresponding marginal distribution differs.

4) $\bar{\theta}_u = \theta_u$*;*

5) $\bar{\epsilon}_t \overset{d}{=} P_s' \epsilon_t$*,*

where T *is a nonsingular* $q \times q$ *matrix, uniquely determined by the pair* $(\theta, \bar{\theta})$*, and* P_s *is a signed permutation matrix.*

The theorem identifies the observable impulse responses and structural-shock distributions up to signed permutation. The remaining freedom to transform the factors changes their coordinates but leaves these responses unchanged.

For a given parameterization, any nonsingular T generates an observationally equivalent model by transforming the factors and adjusting the loadings, VAR coefficients, and impact matrix as in the theorem. Once both parameterizations are fixed, the relation $\bar{\Lambda}_0 T = \Lambda_0$ determines T because $\bar{\Lambda}_0$ has full column rank.

Proof sketch. Independence and non-Gaussianity imply that two observationally equivalent structural transfers have the same range. Applying their left inverses and comparing joint cumulants of order at least three eliminates the Gaussian noise and restricts the relation between the transfers to a permutation, component-specific scales, and time shifts. Once the transfers are aligned, each filtered coordinate is a structural shock plus independent Gaussian noise. Any difference between the two Gaussian variances would give one of the matched shocks a normal component. Assumption 1d rules this out, identifying the aligned shock distributions. Causality and nonzero impact columns remove the time shifts, and unit variances reduce the scales to signs. Subtracting the identified structural covariance from the observable covariance identifies the autocovariances of u_t at every lag and hence its joint distribution across dates. Its normalized causal, invertible representation identifies (Υ, Ω). The remaining argument uses minimality and the factor companion structure to obtain the common factor-coordinate transformation T. Appendix A.1 gives the proof.

The expression (5) in the theorem means that the joint distribution function of the (independent) structural shocks can only be identified up to signed permutation of individual shocks. The theorem also implies that, to pin down the NG-SDFM up to signed permutations of the columns of B, we must impose q^2 normalizing restrictions—the number of free elements in T; see Bai and Wang (2014) and Bai and Wang (2015). There are many ways to complete identification, i.e., to normalize the model. Following Bai and Wang (2015), one convenient choice is the "named factor normalization" (see Stock and Watson, 2016). In a dynamic factor model, it sets the upper $q \times q$ block of one loading matrix Λ_i (for some $i \in \{0, 1, \ldots, s\}$) equal to I_q:

$$\Lambda_i = \begin{bmatrix} I_q \\ \times \end{bmatrix} \tag{13}$$

This identifies the NG-SDFM up to signed permutation. This normalization is admissible if the selected $q \times q$ block of Λ_i is nonsingular. Assumption 6 guarantees that such a selection exists for Λ_0. In the leading case $y_t = \Lambda_0 f_t + u_t$, this normalization equates the factors to the common components of the first q series. The observed series also contain Gaussian variation. Thus the ordering of observables matters: the first q series must be sufficiently informative about the q factors; see Bai and Wang (2015) and Stock and Watson (2016).

3.2 A state-space interpretation of Theorem 1

Theorem 1 is stated in structural dynamic factor notation, but the identification argument is fundamentally a state-space result. In the representation $x_t = Ax_{t-1} + R\epsilon_t$, $y_t = Cx_t + u_t$, the Gaussian component $u_t = \Upsilon(L)e_t$ can itself be included in the state vector. For example, if $u_t = \Pi_1 u_{t-1} + e_t$, the joint state $\xi_t = [x_t' \; u_t']'$ satisfies

$$\xi_t = \begin{bmatrix} A & 0 \\ 0 & \Pi_1 \end{bmatrix} \xi_{t-1} + \begin{bmatrix} R & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} \epsilon_t \\ e_t \end{bmatrix}, \tag{14}$$

$$y_t = \begin{bmatrix} C & I_n \end{bmatrix} \xi_t. \tag{15}$$

The first state block is driven by the non-Gaussian structural shocks, and the second is Gaussian. The two blocks are independent under Assumption 1. For a general finite-order VARMA specification of u_t, the Gaussian state is augmented with the required lags of u_t and e_t. The same block structure applies, with a selection matrix replacing I_n in the observation equation.

In this joint representation, the transfer function from the non-Gaussian structural shocks to the observables is

$$\Psi(z) = C(I - zA)^{-1}R = \sum_{h=0}^{\infty} CA^h R z^h. \tag{16}$$

The full transfer function from $[\epsilon_t' \; e_t']'$ to y_t is $[\Psi(z) \; \Upsilon(z)]$, with $\Upsilon(z) = (I_n - z\Pi_1)^{-1}$ in the VAR(1) example. The rank and left-invertibility conditions concern the non-Gaussian block $\Psi(z)$; they are not imposed on the full transfer matrix. Under the maintained distributional and rank conditions, the joint distribution of the observables across dates identifies the Markov parameters CA^hR, $h = 0, 1, \ldots$, up to signed permutation. Thus observable impulse responses are identified before choosing a particular coordinate system for the latent state.

Minimality concerns the realization (A, R, C) of $\Psi(z)$, rather than the entire augmented state. If this realization is minimal, then, after aligning the signs and order of the structural shocks, it is unique up to a nonsingular linear transformation of the latent state. For $\tilde{x}_t =$

Hx_t, where H is nonsingular, the transformed matrices are

$$\tilde{A} = HAH^{-1}, \qquad \tilde{R} = HR, \qquad \tilde{C} = CH^{-1}. \tag{17}$$

They satisfy $\tilde{C}\tilde{A}^h\tilde{R} = CA^hR$ at every horizon. Minimality rules out redundant states; a normalization selects their coordinates. In factor models, this is the familiar need to normalize the factors and loadings; see Stock and Watson (2016, Section 2.1.3). Online Appendix O.3.1 states the result formally.[11]

For example, suppose $p = 1$, $s = 0$, and there are two factors, so $x_t = (f_{1t}, f_{2t})'$. The transformation

$$H = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad \tilde{x}_t = \begin{bmatrix} f_{1t} + f_{2t} \\ f_{2t} \end{bmatrix}, \tag{18}$$

replaces the first factor by the sum of the two factors. If $C = [c_1\ c_2]$, the transformed loading matrix is $\tilde{C} = [c_1\ \ c_2 - c_1]$. If the original factors are normalized as indices of economic activity and financial conditions, the transformed first factor combines these indices. Both parameterizations generate the same observable process and the same responses to each structural shock. A factor normalization selects one parameterization; it does not alter the structural shocks or give them economic labels.

3.3 Significance of Theorem 1

Although stated in dynamic-factor notation, the structural identification argument also applies to SVARs with measurement error, FAVARs, and the proxy/instrument systems discussed below. In our model, $y_t = \chi_t + u_t$, where χ_t is the common component driven by the non-Gaussian structural shocks and u_t is the Gaussian component. In a conventional factor application, u_t can be restricted to have weak cross-sectional dependence, as in standard approximate factor models. Our model allows general, potentially strong, cross-sectional dependence in u_t, so both components can generate comovement across observables. Much of the DFM literature assumes that χ_t and u_t are orthogonal at all leads and lags. We impose the stronger assumption that the two processes are independent, as in, e.g., Bai and Ng (2008), Doz et al. (2011), Forni et al. (2015), Han (2018), and Barigozzi et al. (2024). We model u_t as a stationary Gaussian VARMA process.

The key implication of Theorem 1 is identification in finite cross sections. This result ap-

[11] The distribution-free result requires the non-Gaussian structural transfer function to be left-invertible, which rules out more non-Gaussian structural innovations than observables. Assuming a specific parametric family for the non-Gaussian distribution (such as a Student-t specification) could in principle deliver identification in such a setting, but such an argument would require additional moment and tensor-uniqueness conditions and is conceptually distinct from the theorem developed here.

plies in the case of a small number of observables, provided the stated distributional and rank conditions hold. In Gaussian DFMs, the common and idiosyncratic components are not separately identified without additional restrictions; modern "approximate" factor arguments therefore rely on $n \to \infty$. Our identification instead uses independence and non-Gaussianity of the common shocks, exclusion of independent Gaussian summands from those shocks, and Gaussian innovations independent of the structural-shock process.

It is also useful to contrast Theorem 1 with principal components analysis (PCA). PCA is a second-moment method. In standard approximate-factor settings, consistency of the principal-component space relies on an eigenvalue separation: the common component must generate large leading eigenvalues, while the idiosyncratic component must not. Our identification argument does not require this eigenvalue structure. The eigenvalues of $\mathrm{Var}(y_t)$ may be small, tied, or dominated by Gaussian variation. Instead, identification comes from Assumptions 1–7: the structural shocks are mutually independent, non-Gaussian, and have no independent Gaussian summand; the Gaussian innovation process is independent of the structural-shock process, and the structural transfer function $\Psi(z)$ satisfies the required rank and minimality conditions. The Gaussian component affects second moments and can therefore affect the principal components, but it contributes no higher-order cumulants. Thus PCA can select high-variance directions driven by the Gaussian component rather than by the non-Gaussian factor component. Conversely, even when PCA recovers the common factor space, it does not identify the structural rotation within that space. Theorem 1 identifies the observable structural transfer function, and hence impulse responses, up to signed permutation. PCA uses only a second-moment summary under our maintained non-Gaussian model, and it can be misleading when large covariance eigenvalues are driven by the Gaussian component rather than by the non-Gaussian factor component.

These differences have two practical consequences. First, the model can be used when n is modest. Second, we can accommodate substantial cross-sectional and temporal dependence in u_t, rather than restricting u_t to have only "mild" cross-sectional correlation as in many approximate-factor setups.

A further distinctive feature is that we allow $n = q$. In contrast to classical static factor identification based on second moments (see, e.g., Anderson and Rubin, 1956 and Bekker and ten Berge, 1997), our distributional restrictions let us nest non-Gaussian SVARs with measurement error within the NG-SDFM framework.

4 Examples nested in the SDFM

Two special cases connect the SDFM to familiar VAR settings: macroeconomic variables measured with error and a macroeconomic system augmented by a noisy instrument.

4.1 A VAR with mismeasured data

Let $q = n$ and set $s = 0$ and $\Lambda_0 = I_n$, so that the observation equation becomes

$$y_t = f_t + u_t, \tag{19}$$

where u_t represents measurement error or other additive Gaussian noise. Assume the factor dynamics are a structural VAR(p):

$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_p f_{t-p} + B\epsilon_t, \tag{20}$$

where ϵ_t is a vector of non-Gaussian shocks and $u_t \sim \mathcal{N}(0, \Sigma_u)$, with u_t independent of ϵ_t at all leads and lags.

4.2 An instrument-augmented VAR

We next augment the macroeconomic system with a noisy instrument that may respond to several structural shocks. In this example, the q macroeconomic outcomes are observed without measurement error, i.e.

$$y_t = f_t. \tag{21}$$

The inclusion of measurement error for the factors f_t is straightforward and nested in our setting. In addition, there is a scalar instrument (the generalization to more than one instrument is, again, straightforward). We call the instrument cleaned of measurement error g_t and it is determined by

$$g_t = a' x_{t-1} + \gamma' \epsilon_t,$$

where the control variables are given by lags of both the instrument and the macroeconomic variables:

$$x_{t-1} := \begin{bmatrix} 1 & g_{t-1} & f'_{t-1} & g_{t-2} & f'_{t-2} & \cdots & g_{t-p} & f'_{t-p} \end{bmatrix}'. \tag{22}$$

We assume that the $q+1$ dimensional vector of factors $\tilde{f}_t = [g_t \; f'_t]'$ follows a structural

VAR of order p with non-Gaussian shocks:

$$\tilde{f}_t = \Phi_1 \tilde{f}_{t-1} + \cdots + \Phi_p \tilde{f}_{t-p} + B\epsilon_t, \tag{23}$$

where $\epsilon_t \in \mathbb{R}^{q+1}$ collects the $q+1$ non-Gaussian structural shocks in the VAR, so we have more structural shocks than macroeconomic data y_t and thus structural shocks cannot be recovered without error from the forecast errors based on y_t alone - we have non-invertibility. The measurement equations are given by Equation (21) and

$$z_t = g_t + \eta_t.$$

η_t is a Gaussian shock that is independent of structural shocks, could be persistent and potentially cross-correlated in the case of more than one instrument, and that we interpret as a measurement error.

For the augmented observables $(z_t, y_t')'$, the loading matrix is the identity. Stability of the augmented VAR and nonsingularity of B therefore give full impact and unit-circle transfer rank. Under the structural-shock restrictions in Assumptions 1b–d and 2, our transfer identification argument allows contamination of the instrument by several structural shocks through γ. It identifies responses to unit-standard-deviation shocks, up to signed permutation, rather than only relative responses as in approaches that do not exploit non-Gaussianity (Plagborg-Moller and Wolf, 2021; Stock and Watson, 2018).[12] If the macroeconomic outcomes are measured exactly, the augmented Gaussian noise is singular because only the instrument is measured with error. The structural impulse responses and shock distributions remain identified up to signed permutation, and the Gaussian measurement-error process is identified, including its serial dependence. An unrestricted Gaussian representation for all augmented observables can nevertheless contain unidentified dynamics in directions with zero innovation variance.

5 Bayesian estimation

To estimate the model, we supplement the identification assumptions with a distributional specification for the structural shocks. We adopt Bayesian estimation because shrinkage priors regularize the loading and dynamic coefficients, helping to limit overfitting when many

[12] In related work, Keweloh et al. (2025) show how non-Gaussianity in structural VARs can allow for contaminated instruments that are correlated with multiple shocks. In contrast to the example discussed here, their approach uses a pseudo-likelihood that combines the non-Gaussian likelihood with moments conditions for the instruments. Their setting assumes invertibility.

parameters must be estimated. We assume independent, standardized Student-t shocks with shock-specific degrees of freedom. This distributional assumption is used for estimation; our identification results do not require Student-t shocks. The Student-t distribution can be represented as a Gaussian distribution with a random variance. Conditional on these latent variances, the model is Gaussian, allowing us to alternate between drawing factor paths and updating parameters, following Geweke (1993) and Kocięcki et al. (2025). For estimation, we specify a VAR for the Gaussian component:

$$y_t = \sum_{a=0}^{s} \Lambda_a f_{t-a} + u_t, \tag{24}$$

$$f_t = \sum_{h=1}^{p} \Phi_h f_{t-h} + B D_t^{1/2} \varepsilon_t, \tag{25}$$

$$u_t = \sum_{h=1}^{l} \Pi_h u_{t-h} + e_t. \tag{26}$$

Here $\{\varepsilon_t\}$, $\{e_t\}$, and $\{D_t\}$ are mutually independent processes. The Gaussian innovations are i.i.d. $N(0, I_q)$ and $N(0, \Omega)$, respectively.

The diagonal entries of $D_t = \mathrm{diag}(d_{1t}, \ldots, d_{qt})$ are independent across components and dates, with

$$d_{jt} \sim IG\left(\frac{\nu_j}{2}, \frac{\nu_j - 2}{2}\right).$$

We parameterize the inverse-gamma density as $p(d) \propto d^{-\alpha-1} \exp(-\beta/d)$. Thus $E[d_{jt}] = 1$, and the structural shocks $\epsilon_t = D_t^{1/2} \varepsilon_t$ have variance one. Large draws of d_{jt} generate the heavy tails. We restrict $\nu_j > 4$, so the shock-distribution and moment assumptions hold; Appendix A.2 verifies the Gaussian-convolution restriction.

Write $\Lambda = (\Lambda_0, \ldots, \Lambda_s)$, $\Phi = (\Phi_1, \ldots, \Phi_p)$, and $\Pi = (\Pi_1, \ldots, \Pi_l)$. We augment the parameters with the factor paths and scales and update the blocks $f, \Lambda, \Pi, \Omega, \Phi, B, D, \nu$.

5.1 Prior specification and factor normalization

In this section, we describe the priors used in all Monte Carlo experiments and empirical applications. We use the named factor normalization: the first q rows of Λ_0 equal I_q. This fixes factor coordinates while leaving the factors latent. The free loading coefficients have independent $N(0, 1)$ priors. We assign independent, zero-centered Minnesota-type normal priors to the factor VAR coefficients. Before restricting the dynamics to be stable, their standard deviations are

$$\mathrm{sd}\left((\Phi_h)_{ij}\right) = \frac{0.5}{h^2} \frac{\widehat{\sigma}_i}{\widehat{\sigma}_j} \begin{cases} 1, & i = j, \\ 0.5, & i \neq j. \end{cases} \tag{27}$$

The scales $\widehat{\sigma}_i$ are residual standard deviations from univariate autoregressions for the q named observables and are fixed before posterior simulation. The prior shrinks higher lags more strongly and adjusts cross-lag coefficients for equation scales.

For unrestricted dynamics of the Gaussian component, let $k = nl$ and use the joint prior

$$p(\Pi, \Omega) \propto IW(\Omega; \bar{Q}, \bar{\nu})\, MN(\Pi; M_0, \Omega, V_0)\, \mathbf{1}_{\mathcal{S}_\Pi}(\Pi), \tag{28}$$

where MN denotes the matrix-normal density with $\mathrm{Var}(\mathrm{vec}\,\Pi \mid \Omega) = V_0 \otimes \Omega$, and $\mathcal{S}_\Pi$ is the stable region. We set $M_0 = 0$, $V_0 = 0.1 I_k$, and $\bar{\nu} = n + 2$; $\bar{Q}$ is diagonal, with the squared univariate innovation scales of the observables on its diagonal. The stability restriction applies to the joint density, so its marginal prior for Ω need not be inverse Wishart. Diagonal-noise specifications instead use separate normal–inverse-gamma priors for the univariate autoregressions. We use a flat prior on B and identical independent priors for ν_j, with weights proportional to $\exp(-0.1\nu_j)$ on the grid $4.01, 4.02, \ldots, 100$.

We fix the number of factors and the lag orders in each reported specification. For a given number of factors, Bayes factors could instead be used to choose lag orders while keeping the factor normalization and the prior on B unchanged.[13] Posterior model probabilities could then be used to average inference across these specifications for quantities with a common interpretation, such as forecasts of the observables. Under the flat prior on B used here, these methods cannot compare specifications with different numbers of factors.

Predictive distributions can instead be combined across specifications, including different numbers of factors, using the optimal prediction pools of Geweke and Amisano (2011). These pools use nonnegative weights summing to one, chosen to maximize the pooled density's log predictive score over a forecast evaluation sample. This approach requires well-defined predictive densities but does not require marginal likelihoods; the pooling weights are not posterior model probabilities.

[13] This qualification concerns our flat prior on the impact matrix B. A marginal likelihood integrates the likelihood against the prior, so an improper prior's arbitrary multiplicative constant matters. When models use the same factor dimension, factor normalization, and prior on B, this constant cancels from their Bayes factor, provided the remaining priors are proper and the integrals are finite. Changing the number of factors changes the dimension of B, so this cancellation no longer applies. See Koop (2003, Section 3.6).

5.2 Conditional-regression updates

Factor paths and loadings. Substituting the observation equation into the VAR for the Gaussian component yields

$$\widetilde{y}_t = \sum_{a=0}^{s+l} K_a f_{t-a} + e_t, \tag{29}$$
$$\widetilde{y}_t = y_t - \sum_{h=1}^{l} \Pi_h y_{t-h}, \qquad K_a = \Lambda_a - \sum_{h=1}^{l} \Pi_h \Lambda_{a-h},$$

with $\Lambda_a = 0$ outside $0 \le a \le s$. Stacking $\max\{p, s+l+1\}$ factor blocks gives a linear Gaussian state-space model with factor-innovation covariance BD_tB' and measurement covariance Ω. We draw the factor path jointly by simulation smoothing (Carter and Kohn, 1994), using a square-root Kalman filter as in Bai and Wang (2015).[14]

Conditional on the factors, the same equation is linear in the loading coefficients. Its design block for $\mathrm{vec}(\Lambda_a)$ is

$$W_t^a = f'_{t-a} \otimes I_n - \sum_{h=1}^{l} f'_{t-a-h} \otimes \Pi_h.$$

After subtracting the contribution of the fixed loading entries, stack the equations as $Y_\lambda = W\lambda + e$, where λ contains the free entries and $V = \mathrm{Var}(e)$ is block diagonal with blocks Ω. The loading update is Gaussian with covariance $(W'V^{-1}W + I)^{-1}$ and mean $(W'V^{-1}W + I)^{-1}W'V^{-1}Y_\lambda$.

Dynamics of the Gaussian component. Recover $u_t = y_t - \sum_a \Lambda_a f_{t-a}$ and write its VAR as $U = \Pi X + E$, where the columns of X contain the l lagged residual vectors. If N_e is the number of usable dates, the update for Π is matrix normal with row covariance Ω, column covariance C, and mean M:

$$C = (XX' + V_0^{-1})^{-1}, \qquad M = (UX' + M_0V_0^{-1})C.$$

We reject draws outside $\mathcal{S}_\Pi$. Conditional on the accepted coefficients, put $E = U - \Pi X$. The covariance update is

$$\Omega \mid \bullet \sim IW\Big(\bar{Q} + EE' + (\Pi - M_0)V_0^{-1}(\Pi - M_0)',\ \bar{\nu} + N_e + k\Big), \tag{30}$$

[14] In all reported estimations, the factor simulation smoother uses $s_1 = (f_1', f_0', f_{-1}', f_{-2}')' \sim N(0, 100I_{4q})$ before the first filtered observation at $t = 2$. This distribution is specified in the factor coordinates associated with the standardized observables, with its mean and covariance fixed independently of the model parameters and scale-mixture variables.

where $\bullet$ denotes the other blocks. The coefficient prior contributes both its quadratic term and k degrees of freedom: its determinant factor is $|\Omega|^{-k/2}$. Conditioning on Π does not remove that dependence on Ω. In diagonal-noise specifications, the corresponding variance and coefficient updates are univariate; their variance updates retain these prior contributions.

Factor dynamics. Let $\bar{f}_{t-1} = (f'_{t-1}, \ldots, f'_{t-p})'$, $X_t = \bar{f}'_{t-1} \otimes I_q$, and $\phi = \text{vec}(\Phi)$. Conditional on the factors, B, and D, equation (25) is a Gaussian regression with time-varying covariance $Q_t = BD_tB'$. If V_ϕ is the diagonal prior covariance defined by (27), the Gaussian update has moments

$$\begin{aligned} V_\phi^* &= \left(V_\phi^{-1} + \sum_t X_t' Q_t^{-1} X_t \right)^{-1}, \\ m_\phi^* &= V_\phi^* \sum_t X_t' Q_t^{-1} f_t. \end{aligned} \tag{31}$$

We draw from $N(m_\phi^*, V_\phi^*)$ and reject draws with unstable companion roots.

Impact matrix, scales, and tail parameters. Given the factors and Φ, the residuals $v_t = f_t - \sum_h \Phi_h f_{t-h}$ satisfy $v_t = B\epsilon_t$. We update B using the factorization $B^{-1} = \Gamma LU$ of Kocięcki et al. (2025), where Γ is diagonal and L, U are unit-diagonal triangular matrices. Updating these blocks separately avoids sampling the nonstandard joint conditional of B directly. Given $\epsilon_t = B^{-1}v_t$, the scale updates are independent:

$$d_{jt} \mid \bullet \sim IG\left(\frac{\nu_j + 1}{2}, \frac{\nu_j - 2 + \epsilon_{jt}^2}{2} \right).$$

Finally, we evaluate $p(\nu_j) \prod_t p(d_{jt} \mid \nu_j)$ over the finite grid, normalize these weights, and draw ν_j using Griddy-Gibbs (Ritter and Tanner, 1992). Online Appendix O.1 gives the conditional formulas in full.

5.3 Shock normalization

The loading normalization fixes factor coordinates but leaves structural-shock signs and ordering undetermined. To normalize the shocks, we use a fixed nonsingular reference impact matrix $\widehat{B}$. The generalized likelihood-preserving normalization of Kocięcki et al. (2025) chooses

$$P_s = \underset{P \in \mathcal{P}_\pm}{\text{argmin}} \left\| \widehat{B}^{-1} BP - I_q \right\|_F^2, \tag{32}$$

for each draw B, where $\mathcal{P}_{\pm}$ is the set of signed permutation matrices and $\|\cdot\|_F$ is the Frobenius norm. We transform $B \mapsto BP_s$, $\epsilon_t \mapsto P_s'\epsilon_t$, and $D_t \mapsto P_s'D_tP_s$, and permute the degrees of freedom accordingly. These transformations leave factor innovations and the likelihood unchanged.

$\widehat{B}$ can be estimated by Student-t quasi-maximum likelihood using factor-VAR residuals from a preliminary sampler run. In our Monte Carlo exercises we set it equal to the true impact matrix to ease interpretation of the resulting IRFs. Online Appendix O.1.7 gives the matching algorithm and reference construction; Online Appendix O.4 gives the Monte Carlo calibrations and simulation lengths.

6 Monte Carlo evidence

We use Monte Carlo experiments to assess how accurately the Bayesian estimator recovers impulse responses and structural shocks in finite samples. For each specification, we estimate the model on 100 independently simulated datasets and compare the estimates with the true responses and shock sequences. The panel experiments use ten observables and vary the non-Gaussianity of the structural shocks and the size, persistence, and cross-sectional correlation of the Gaussian component. The proxy experiments examine monetary policy responses when a noisy proxy may also load on another structural shock. We show estimated impulse responses alongside the true responses and summarize shock recovery using correlations between estimated and true shock sequences. Online Appendix O.4 reports additional measures of response accuracy and estimation diagnostics.

6.1 Response recovery in modest cross sections

The three data-generating processes (DGPs) in Figure 2 have $n = 10$ observables, $q = 2$ factors, static loadings, and a factor VAR(4). We simulate $T = 250$ observations from each process. In all three cases, the estimated model is correctly specified in its factor dimension, lag orders, loading structure, and shock distributions. Structural shocks are mutually independent Student-t variables, scaled to unit variance. The parameter ν is their common degrees of freedom; larger values imply thinner tails and distributions closer to Gaussian. The Gaussian component follows

$$u_t = \rho u_{t-1} + e_t, \qquad e_t \sim N\Big(0, \sigma^2[(1-r)I_n + r\mathbf{1}\mathbf{1}']\Big). \tag{33}$$

Here ρ is the common autoregressive coefficient, σ is the marginal standard deviation of each innovation e_{it}, and r is the correlation between any two distinct innovations. The

moderate-noise specifications in the first two columns hold $(\rho, \sigma, r) = (0.40, 0.60, 0.30)$ fixed and change ν from 5 to 20. The strong-noise specification in the third column uses $\nu = 8$ and $(\rho, \sigma, r) = (0.80, 1.00, 0.60)$. Thus its Gaussian noise is larger, more persistent, and more correlated across observables. Its unconditional marginal variance, $\sigma^2/(1-\rho^2)$, is 2.78, compared with 0.43 in the first two columns. The third column changes both the noise process and ν; the first two isolate the change in shock tails.

The factor dynamics, impact matrix, and loadings are identical across these three specifications. The first two loading rows form I_2, so $y_{1t} = f_{1t} + u_{1t}$ and $y_{2t} = f_{2t} + u_{2t}$. The structural impact matrix is

$$B = \begin{pmatrix} 1 & 0.35 \\ -0.20 & 1 \end{pmatrix}.$$

Online Appendix O.4 gives all four factor-VAR coefficient matrices and the remaining loading rows. Its Table O.1 reports response errors and interval coverage for these three specifications.

Each specification has 100 independent replications. We set the factor dimension and lag orders to their true values and estimate the factor paths and shock sequences. In every design, signs and permutations are normalized against the true impact matrix for the first q standardized observables, using (32). We use the same zero-centered Minnesota prior in every specification, with own-lag standard deviation $0.5/k^2$ at lag k and cross-lag scale 0.5, adjusted for equation scales. Each chain discards 2,000 iterations and retains every fourth draw from a further 2,000. The experiments hold the factor dimension and lag orders fixed and do not study their selection.

Figure 2 compares the true responses with the posterior medians from all 100 replications. Each pale curve is one estimated response path. The panels show the same two responses in each specification, after reversing sample standardization to express them in common DGP units.

In the first column, most estimated own responses of variable 1 follow the true decay toward zero. Estimates of the smaller response of variable 2 are more dispersed, including estimates with the wrong impact sign. The second column shows greater dispersion when ν rises to 20. In the third column, many estimated own responses understate the true impact, and some decay too slowly. Online Appendix O.4 reports additional statistics.

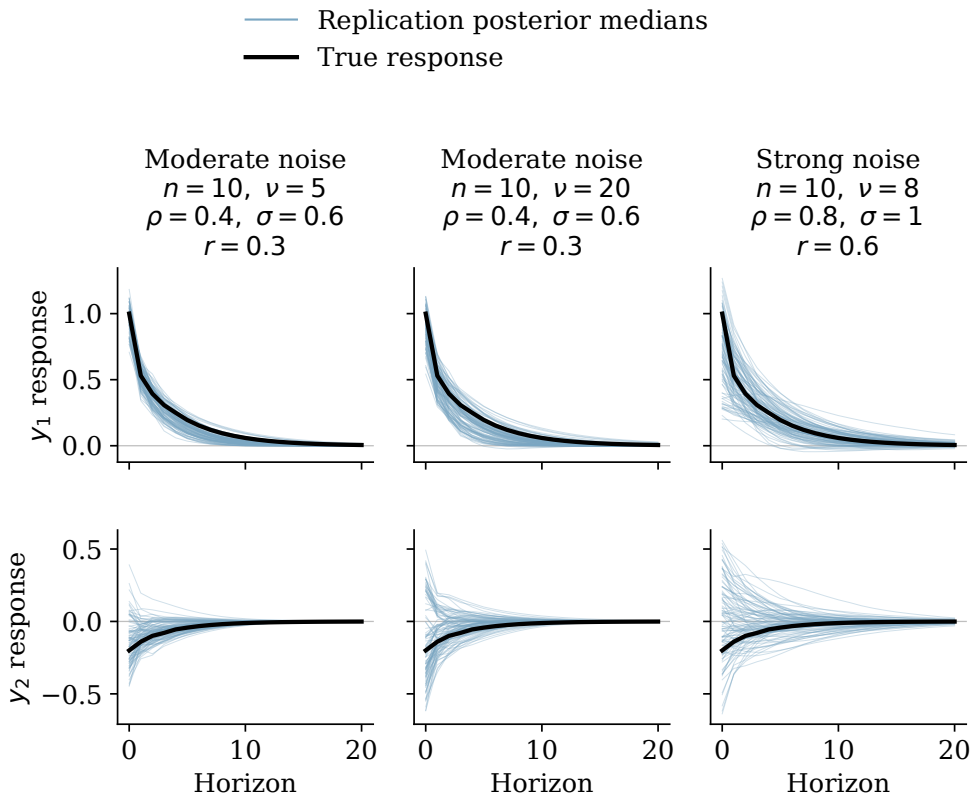


Figure 2: Estimated and true responses with ten observables and two factors. All columns use $T = 250$, a factor VAR(4), and static loadings. Rows show variables 1 and 2 responding to a one-standard-deviation shock 1; their true impact responses are 1 and -0.2. Structural shocks are independent, unit-variance Student-t variables with degrees of freedom $\nu = 5, 20, 8$ across columns. Gaussian noise satisfies (33): ρ is its AR coefficient, σ its innovation standard deviation, and r its innovation correlation. The first two columns use $(0.4, 0.6, 0.3)$; the third uses $(0.8, 1, 0.6)$. Black curves are the fixed true responses. Each pale blue curve is the posterior median from one of the 100 independently simulated datasets; all replications are shown. Responses are in common raw DGP units, with each dataset's standardization reversed. Curve dispersion represents sampling variation in point estimates, not posterior credible intervals.

6.2 Recovery of structural shocks

Impulse responses describe the propagation of structural shocks. We now examine how well the estimator recovers the timing and magnitude of the shocks themselves.

Figure 3 compares the true shocks with their posterior medians in replication 1 of the first specification ($n = 10$, $\nu = 5$). The replication and the first 100 available dates are chosen by index, without inspecting fit. Estimated amplitudes are neither smoothed nor rescaled. Shading shows the central 90% posterior interval for each shock at each date, measuring uncertainty about the latent shock realization.

The estimated sequences track much of the true variation, although individual magnitudes remain uncertain. Correlations over all 249 available dates are 0.93 and 0.92 for shocks 1 and 2.

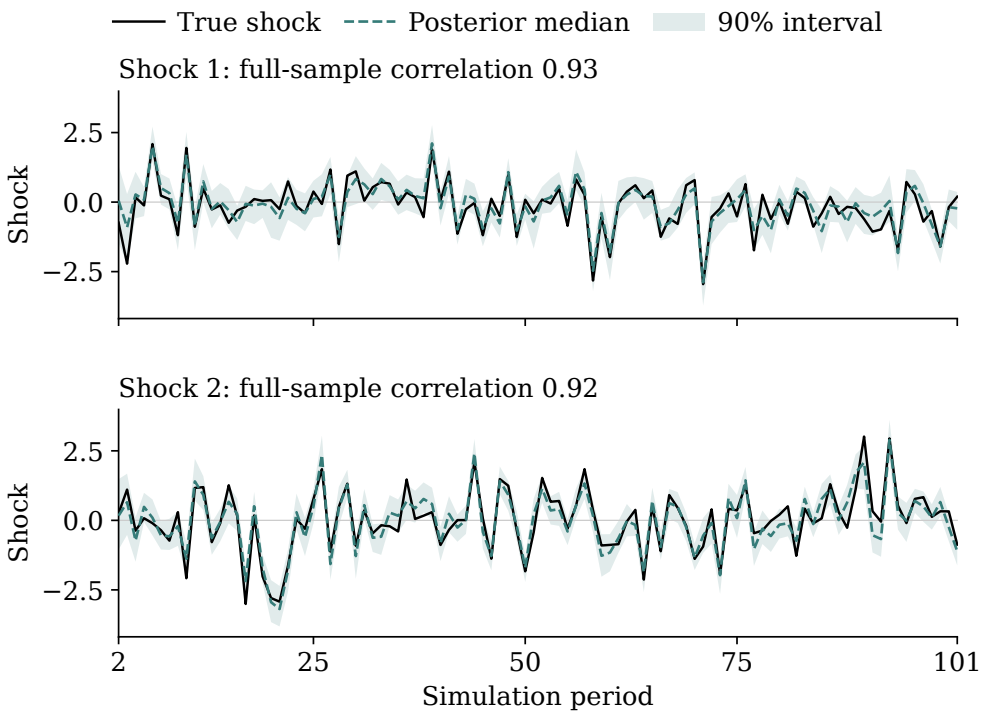


Figure 3: Recovery of structural shocks in a fixed simulation. The first specification in Figure 2 has $n = 10$, $T = 250$, two factors, a factor VAR(4), and independent unit-variance Student-t shocks with $\nu = 5$. Gaussian noise has AR coefficient 0.4, innovation standard deviation 0.6, and innovation correlation 0.3. Black lines show the true shocks; teal dashed lines show their pointwise posterior medians across 500 retained draws. Shaded bands span the 5th and 95th posterior percentiles at each date; they are pointwise intervals, not simultaneous bands. Both panels use replication 1 and sample dates 2–101, the first 100 dates with saved shock draws. Shocks retain their unit-unconditional-variance scale. The correlations reported in the panels use all 249 available dates. The replication and display window are fixed by index rather than selected for fit.

Table 1 summarizes shock recovery across all 100 datasets in each of the seven specifications shown in the main figures. For the panel designs, moderate Gaussian noise has AR coefficient $\rho = 0.40$, innovation standard deviation $\sigma = 0.60$, and pairwise innovation correlation $r = 0.30$ in (33). Strong Gaussian noise sets these parameters to 0.80, 1.00, and 0.60, respectively. The corresponding unconditional noise variances are 0.43 and 2.78. For each dataset, we correlate the true shock with its estimated sequence over all available dates. These table statistics use posterior-mean shock estimates, whereas Figure 3 uses posterior medians. We average the correlations across all structural shocks within each dataset, including all four shocks in the proxy design. Estimated columns are matched to the true shocks. The table describes variation in recovery across datasets; it does not report posterior uncertainty about the correlations.

With moderate Gaussian noise, mean shock correlation falls from 0.91 to 0.86 as ν rises from 5 to 20. With strong Gaussian noise and ten observables, mean correlation is 0.48,

Table 1: Correlation between true and estimated structural shocks

Specification	Mean	SD	10th pct.	90th pct.
Moderate Gaussian noise ($n = 10, q = 2$)				
$\nu = 5$	0.91	0.03	0.88	0.94
$\nu = 20$	0.86	0.06	0.78	0.92
Strong Gaussian noise ($n = 10, q = 2, \nu = 8$)				
$n = 10$	0.48	0.20	0.19	0.76
Noisy monetary proxy ($n = q = 4$)				
$\nu = 5,\ c = 0.0$	0.82	0.04	0.76	0.87
$\nu = 5,\ c = 0.2$	0.86	0.04	0.81	0.90
$\nu = 5,\ c = 0.4$	0.88	0.03	0.83	0.91
$\nu = 8,\ c = 0.4$	0.84	0.03	0.79	0.88

Each row uses 100 independently simulated datasets. Within a dataset, we correlate each true shock sequence with its posterior-mean estimate over all available dates, then average across all q shocks, including all four in the proxy specifications. Estimated columns are matched to the true shocks. Columns report the mean, standard deviation, and 10th and 90th percentiles across datasets. These percentiles describe sampling variation, not credible intervals. The table contains the seven specifications shown in the main figures. The panel designs use $T = 250$ and the proxy designs use $T = 500$; one initial observation is lost to the lag in the Gaussian component. For Gaussian panel noise in (33), the AR coefficient, innovation standard deviation, and pairwise innovation correlation are $(\rho, \sigma, r) = (0.40, 0.60, 0.30)$ under moderate noise and $(0.80, 1.00, 0.60)$ under strong noise. Here c is proxy contamination as defined in (34). The factor dimension and lag orders are fixed at their true values. High correlation alone does not establish correct shock amplitudes.

and the 10th and 90th percentiles across datasets are 0.19 and 0.76. Across the four proxy specifications, mean correlations range from 0.82 to 0.88.

6.3 An extension with a contaminated proxy

This experiment implements the instrument-augmented VAR in Section 4.2. We ask whether jointly modeling a monetary policy proxy and macroeconomic variables allows us to recover monetary impulse responses when the proxy contains measurement error and also responds to an information shock. The latter contamination violates the usual exclusion restriction for an instrument. Our model allows the proxy to load on several independent non-Gaussian shocks and separates their effects from Gaussian measurement error. The simulations vary both the proxy's contamination and the non-Gaussianity of the structural shocks.

Four mutually independent, unit-variance Student-t shocks drive three macroeconomic variables, interpreted as output, prices, and the policy rate, and a latent proxy. The macroeconomic variables are measured exactly. A one-standard-deviation monetary tightening lowers output and prices on impact by 0.4 and 0.3 and raises the policy rate by 0.8. An information shock raises the three variables by 1.6, 1.1, and 1.2, respectively. We choose these impact coefficients to distinguish a monetary tightening from an information shock

that raises all three variables.

Let x_{1t}, x_{2t}, and x_{3t} denote the three macroeconomic variables and z_t^* the latent proxy. The latter follows

$$\begin{aligned} z_t^* &= 0.10x_{1,t-1} - 0.05x_{2,t-1} + 0.08x_{3,t-1} + 0.25z_{t-1}^* \\ &\quad + \sqrt{1-c}\,\epsilon_{m,t} + \sqrt{c}\,\epsilon_{I,t}, \\ z_t &= z_t^* + u_{z,t}, \qquad u_{z,t} = 0.3u_{z,t-1} + \eta_t, \quad \eta_t \sim N(0, 0.25(1-0.3^2)). \end{aligned} \tag{34}$$

Here m denotes the monetary shock and I the information shock. The parameter c is the information shock's share of the latent proxy's innovation variance; it is not a share of the observed proxy's total variance. Gaussian measurement error is independent of all economic shocks and has unconditional variance 0.25. At $c = 0$ the proxy is uncontaminated by other economic shocks but still contains measurement error. The augmented latent system has a full-rank impact matrix. The three-variable macro block alone cannot recover the monetary shock. Online Appendix O.4 gives the macro impact matrix and VAR coefficients.

Each specification uses $T = 500$, a VAR(4), and 100 independent datasets. Figure 4 reports all four specifications: $(\nu, c) = (5, 0), (5, 0.2), (5, 0.4), (8, 0.4)$. The first three increase contamination at fixed shock tails; the fourth holds contamination at 40% and raises the degrees of freedom from 5 to 8. The true macroeconomic responses are identical across columns. As in Figure 2, the pale curves are individual posterior-median response paths in common raw units. We match the estimated columns to the true shocks before constructing these responses. The figure therefore measures recovery of the monetary response conditional on this matching, rather than success in assigning its economic label.

The first three columns show similar response shapes as contamination increases at $\nu = 5$. Most estimated output responses are negative and most policy-rate responses are positive on impact, but the estimated impact magnitudes tend to be smaller than the truth. The fourth column shows weaker recovery with thinner shock tails. At $c = 0.4$, the estimated output impact has the wrong sign in 3 of 100 datasets when $\nu = 5$ and 12 of 100 when $\nu = 8$, even after matching columns to the true shocks. Thus increasing contamination and reducing non-Gaussianity have distinct effects in these experiments.

For comparison, we estimate a *proxy-first recursive VAR*: an OLS VAR(4) with the proxy ordered first. Its first recursive shock is interpreted as monetary, so other orthogonal shocks cannot affect the proxy contemporaneously. This restriction fails when the proxy also loads on the information shock. NG-SDFM instead estimates the proxy's loadings on all four independent non-Gaussian shocks and separates its Gaussian measurement error.

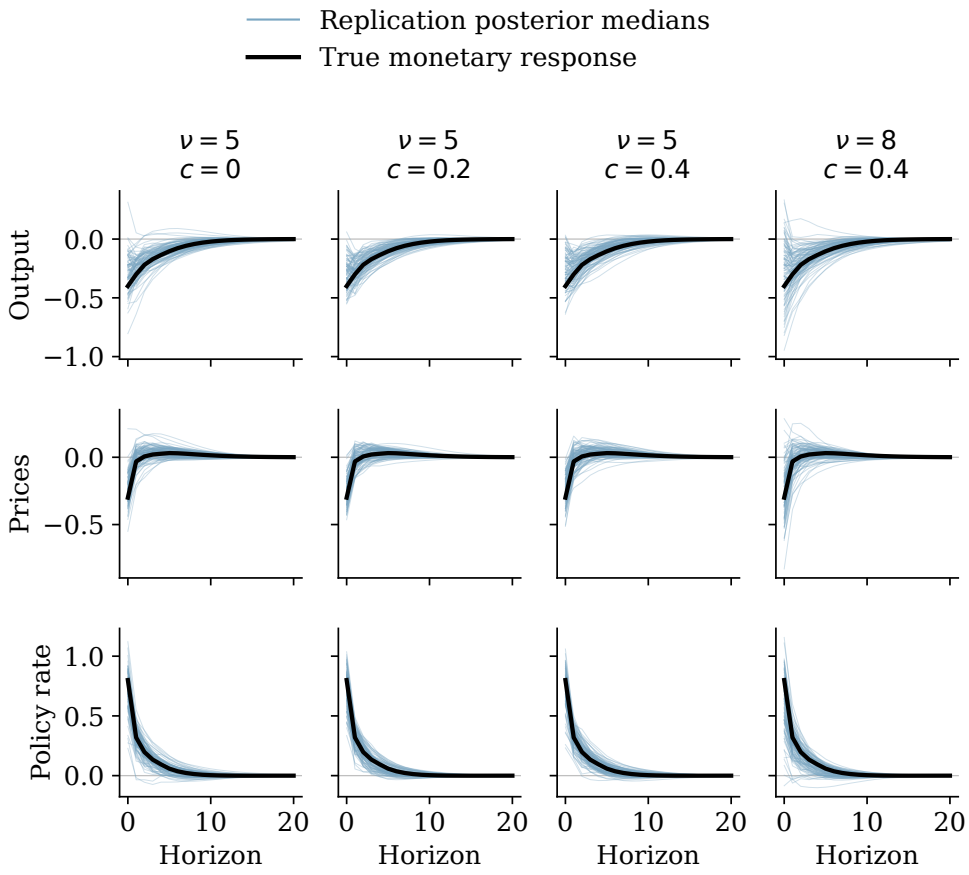


Figure 4: Recovery of monetary responses with a noisy, contaminated proxy. Rows show output, prices, and the policy rate responding to a one-standard-deviation monetary tightening. Columns give the Student-t degrees of freedom ν and the information shock's share c of latent proxy innovation variance in (34). Every specification has four structural shocks, a VAR(4), $T = 500$, and Gaussian proxy measurement error with AR coefficient 0.3 and unconditional variance 0.25. The three macro variables are measured exactly. Black curves are the true responses; pale blue curves are posterior medians from all 100 datasets, with sample standardization reversed. Estimated shock columns are matched to the truth. The figure does not incorporate errors in assigning the monetary label. Curve dispersion represents sampling variation, not posterior credible intervals.

Online Appendix Table O.3 reports root mean squared errors (RMSEs) for responses to a one-standard-deviation monetary shock. Both estimators have an RMSE of 0.03 in the clean-proxy specification after rounding. NG-SDFM has lower RMSE in the three contaminated-proxy specifications. At $\nu = 5$, its RMSE remains between 0.02 and 0.03 as contamination increases, whereas the recursive VAR's RMSE rises from 0.03 to 0.11. With 40% contamination and thinner shock tails ($\nu = 8$), NG-SDFM's RMSE increases to 0.04, compared with 0.11 for the recursive VAR.

7 Gaussian and non-Gaussian variation in US data

We use the model to assess the quantitative importance of non-Gaussian shocks for US macroeconomic fluctuations. The variance decomposition measures how much of each variable's volatility they explain, while the covariance decomposition measures their contribution to comovement. Substantial Gaussian contributions are consistent with Figure 1: rejecting Gaussianity does not establish that non-Gaussian shocks account for most of the variance. A non-Gaussian component can generate detectable departures from normality even when its variance contribution is small. Figure 1 tests residuals from univariate AR(4) regressions, whereas this section decomposes the model-implied unconditional covariance of the transformed series.

We apply the model to 22 quarterly US macroeconomic series from FRED-QD (McCracken and Ng, 2020). We use a modest cross section rather than the full database to illustrate factor extraction without relying on a large number of observables. The selected series cover real activity, labor markets, housing, prices, interest rates, equities, and money. The sample contains 242 observations from 1959Q3 to 2019Q4. Prices enter in first log differences, so CPI measures quarterly inflation; real GDP and other quantity series generally enter in log growth. Interest rates, the credit spread, unemployment, and average weekly hours enter in levels. We demean and standardize each input. Online Appendix O.5 lists all variables, transformations, and estimation details.

We specify four factors with contemporaneous loadings and a VAR(4). Their structural innovations are independent standardized Student-t variables. The Gaussian component follows an unrestricted VAR(1), with unrestricted innovation covariance. It can therefore generate persistent comovement across observables. The decomposition separates Gaussian from non-Gaussian variation; it does not separate common variation from variable-specific noise. We use the Bayesian procedure in Section 5, with Minnesota shrinkage for the factor dynamics.

Table 2 decomposes unconditional variance into the aggregate non-Gaussian component and the Gaussian component. These are population variances implied by the stationary model for the transformed inputs, not variances of cumulated GDP or price levels. Gaussian variation accounts for 79.46% of GDP-growth variance, 66.40% of productivity-growth variance, and 92.87% of unemployment variance at the posterior mean. The non-Gaussian component accounts for 78.51% and 90.31% of the variance of the federal funds rate and ten-year yield, respectively. The allocation is less precise for inflation and the credit spread: their posterior-mean Gaussian shares are 39.13% and 66.42%, respectively, and both 90% intervals include one-half. Thus the non-Gaussian factors need not account for most of each

Table 2: Unconditional variance shares in selected US variables

Observable	Non-Gaussian (%)	Gaussian (%)
Real GDP growth	20.54 [8.11, 42.22]	79.46 [57.78, 91.89]
CPI inflation	60.87 [39.32, 84.14]	39.13 [15.86, 60.68]
Federal funds rate	78.51 [58.69, 94.65]	21.49 [5.35, 41.31]
Baa–Treasury spread	33.58 [14.73, 63.89]	66.42 [36.11, 85.27]
Unemployment	7.13 [1.33, 19.33]	92.87 [80.67, 98.67]
Payroll employment growth	12.72 [3.89, 31.08]	87.28 [68.92, 96.11]
Ten-year Treasury yield	90.31 [78.49, 98.24]	9.69 [1.76, 21.51]
Labor productivity growth	33.60 [15.53, 60.91]	66.40 [39.09, 84.47]

Posterior mean shares (%), with 5th and 95th posterior percentiles in brackets. Non-Gaussian combines all four structural shocks; Gaussian denotes the joint correlated component. Both components can generate comovement across observables. The first four variables define the identity loading block. All variables retain their estimation transformations. Shares concern model-implied unconditional variance, including all frequencies, and are calculated within each posterior draw.

observable's variation.

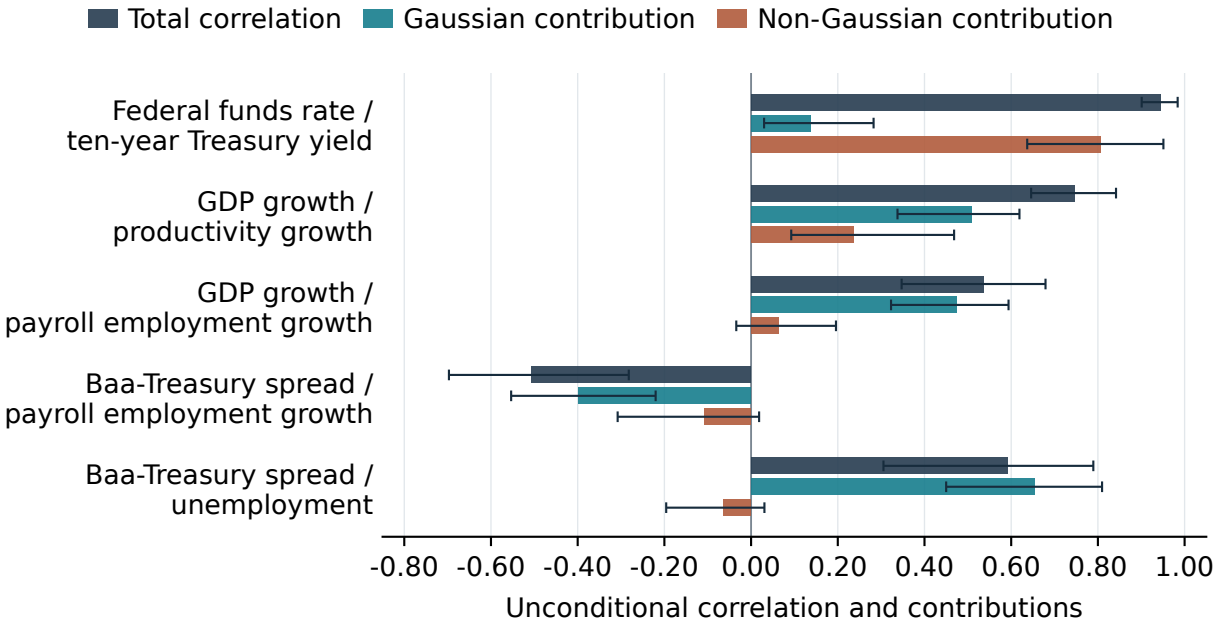


Figure 5: Unconditional correlations and component contributions. Bars report posterior means; whiskers mark the 5th and 95th posterior percentiles. Each contribution divides the component covariance by the two variables' total marginal standard deviations. Gaussian and non-Gaussian contributions add to total correlation within each draw. The correlations are model-implied and use the estimation transformations, including variation at all frequencies.

Figure 5 asks how the two components contribute to comovement. Their covariances add by independence. Dividing each by the same total marginal standard deviations gives addi-

tive contributions to unconditional correlation. These contributions differ from correlations computed using each component's own standard deviations. The non-Gaussian component contributes 0.81 to the 0.94 correlation between the two interest rates. The Gaussian component contributes 0.51 to the 0.75 correlation between GDP growth and productivity growth.

The table and figure describe different entries of the same covariance decomposition. Variance shares bound the component contributions to correlation, but their signs and magnitudes also depend on correlation within each component. A variable's predominantly Gaussian variance does not by itself imply predominantly Gaussian comovement with another variable. The two contributions can also offset each other. At the posterior mean, the spread's correlation with unemployment is 0.59, comprising a Gaussian contribution of 0.65 and a non-Gaussian contribution of -0.06. The latter's 90% interval includes zero. Online Appendix O.5 derives this relationship and reports unconditional variance shares for all 22 variables.

8 Conclusions

This paper develops a practical framework for structural analysis in dynamic factor models when the common shocks are non-Gaussian. Our main theoretical contribution is a global identification result for non-Gaussian SDFMs: under the stated distributional, moment, and rank conditions, structural-shock distributions and impulse responses are identified up to signed permutation without exclusion restrictions, sign restrictions, or external instruments.

A key practical implication is that identification does not rely on large cross sections. The model can be applied when the number of observables is modest, and it remains valid even when the Gaussian component exhibits serial dependence and general cross-sectional correlation. This substantially broadens the set of empirical environments in which factor-based structural analysis is feasible.

For inference, we provide a Gibbs sampler based on the Student-t scale-mixture representation and introduce a likelihood-preserving normalization that resolves sign and ordering ambiguities. The Monte Carlo experiments show that the model can recover impulse responses and structural shocks well with modest cross sections and Gaussian noise.

References

Dante Amengual, Gabriele Fiorentini, and Enrique Sentana. Specification tests for non-Gaussian structural vector autoregressions. *Journal of Econometrics*, 244(2):105803, 2024. doi: 10.1016/j.jeconom.2024.105803.

Brian D.O. Anderson and Manfred Deistler. Generalized Linear Dynamic Factor Models - a Structure Theory. In *2008 47th IEEE Conference on Decision and Control*, pages 1980–1985, 2008. doi: 10.1109/CDC.2008.4739367.

Theodore W. Anderson and Herman Rubin. Statistical inference in factor analysis. In Jerzy S. Neyman, editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 5, pages 111–150. University of California Press, 1956.

Jetro Anttonen, Markku Lanne, and Jani Luoto. Bayesian inference on fully and partially identified structural vector autoregressions. Technical report, SSRN, 2023.

Jetro Anttonen, Markku Lanne, and Jani Luoto. Statistically identified structural VAR model with potentially skewed and fat-tailed errors. *Journal of Applied Econometrics*, 39 (3):422–437, 2024.

Jushan Bai. Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71 (1):135–171, 2003.

Jushan Bai and Serena Ng. Large Dimensional Factor Analysis. *Foundations and Trends in Econometrics*, 3(2):89–163, 2008.

Jushan Bai and Peng Wang. Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics*, 178(2):794–804, 2014. doi: https://doi.org/10.1016/j.jeconom.2013.11.001. URL https://www.sciencedirect.com/science/article/pii/S0304407613002315.

Jushan Bai and Peng Wang. Identification and Bayesian Estimation of Dynamic Factor Models. *Journal of Business & Economic Statistics*, 33(2):221–240, 2015. doi: 10.1080/07350015.2014.941467. URL https://doi.org/10.1080/07350015.2014.941467.

Jushan Bai and Peng Wang. Econometric Analysis of Large Factor Models. *Annual Review of Economics*, 8:53–80, 2016.

Matteo Barigozzi, Marc Hallin, Matteo Luciani, and Paolo Zaffaroni. Inferential theory for generalized dynamic factor models. *Journal of Econometrics*, 239(2):105422, 2024.

Christian F. Beckmann and Stephen M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23(2):137–152, 2004. doi: 10.1109/TMI.2003.822821.

Paul A. Bekker and Jos M.F. ten Berge. Generic global identification in factor analysis. *Linear Algebra and its Applications*, 264:255–263, 1997.

Ben S. Bernanke, Jean Boivin, and Piotr Eliasz. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1):387–422, 2005.

Hilde C. Bjørnland and Leif A. Thorsrud. Boom or Gloom? Examining the Dutch Disease in Two-speed Economies. *The Economic Journal*, 126(598):2219–2256, 2016.

Stéphane Bonhomme and Jean-Marc Robin. Consistent noisy independent component analysis. *Journal of Econometrics*, 149(1):12–25, 2009. doi: 10.1016/j.jeconom.2008.12.019.

Robin Braun. The importance of supply and demand for oil prices: Evidence from non-Gaussianity. *Quantitative Economics*, 14(4):1163–1198, 2023.

Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer, second edition, 2006.

Markus Brunnermeier, Darius Palia, Karthik A. Sastry, and Christopher A. Sims. Feedbacks: Financial markets and economic activity. *American Economic Review*, 111(6):1845–1879, 2021.

Fabio Canova and Filippo Ferroni. Mind the gap! stylized dynamic facts and structural models. *American Economic Journal: Macroeconomics*, 14(4):104–35, October 2022. doi: 10.1257/mac.20200054. URL https://www.aeaweb.org/articles?id=10.1257/mac.20200054.

C. K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3): 541–553, 09 1994.

Kung-Sik Chan and Lop-Hing Ho. On the unique representation of non-Gaussian multivariate linear processes. Technical report, Department of Statistics and Actuarial Science, The University of Iowa, 2004.

Kung-Sik Chan, Lop-Hing Ho, and Howell Tong. A note on time-reversibility of multivariate linear processes. *Biometrika*, 93(1):221–227, 2006.

Bin Chen, Jinho Choi, and Juan Carlos Escanciano. Testing for fundamental vector moving average representations. *Quantitative Economics*, 8(1):149–180, 2017. doi: https://doi.org/10.3982/QE393. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/QE393.

Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36: 287–314, 1994.

Catherine Doz, Domenico Giannone, and Lucrezia Reichlin. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205, 2011.

Mario Forni and Luca Gambetti. The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics*, 57(2):203–216, 2010.

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics*, 82 (4):540–554, 2000.

Mario Forni, Domenico Giannone, Marco Lippi, and Lucrezia Reichlin. Opening the Black Box: Structural Factor Models with Large Cross Sections. *Econometric Theory*, 25(5): 1319–1347, 2009.

Mario Forni, Luca Gambetti, and Luca Sala. No News in Business Cycles. *The Economic Journal*, 124(581):1168–1191, 2014.

Mario Forni, Marc Hallin, Marco Lippi, and Paolo Zaffaroni. Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185(2):359–371, 2015.

Mario Forni, Luca Gambetti, and Luca Sala. Structural VARs and noninvertible macroeconomic models. *Journal of Applied Econometrics*, 34(2):221–246, 2019.

Mario Forni, Luca Gambetti, Antonio Granese, Luca Sala, and Stefano Soccorsi. An Ameri-

can Macroeconomic Picture: Supply and Demand Shocks in the Frequency Domain. *American Economic Journal: Macroeconomics*, 17(3):311–341, 2025a.

Mario Forni, Luca Gambetti, Marco Lippi, and Luca Sala. Common Components Structural VARs. *Journal of Business & Economic Statistics*, pages 1–14, 2025b.

Mario Forni, Luca Gambetti, Marco Lippi, and Luca Sala. Informing DSGE Models Through Dynamic Factor Models. *Journal of Applied Econometrics*, 40(5):487–507, 2025c.

Bernd Funovits. Identifiability and estimation of possibly non-invertible SVARMA Models: The normalised canonical WHF parametrisation. *Journal of Econometrics*, 241(2):105766, 2024. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2024.105766. URL https://www.sciencedirect.com/science/article/pii/S030440762400112X.

John Geweke. The dynamic factor analysis of economic time series models. In Dennis J. Aigner and Arthur S. Goldberger, editors, *Latent Variables in Socio-Economic Models*, pages 365–383. North-Holland, Amsterdam, 1977.

John Geweke. Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*, 8(S1):S19–S40, 1993.

John Geweke and Gianni Amisano. Optimal prediction pools. *Journal of Econometrics*, 164 (1):130–141, 2011. doi: 10.1016/j.jeconom.2011.02.017.

Domenico Giannone, Lucrezia Reichlin, and Luca Sala. Monetary policy in real time. In Mark Gertler and Kenneth Rogoff, editors, *NBER Macroeconomics Annual 2004*, volume 19, pages 161–200. MIT Press, 2005.

Christian Gouriéroux, Alain Monfort, and Jean-Paul Renne. Identification and Estimation in Non-Fundamental Structural VARMA Models. *The Review of Economic Studies*, 87 (4):1915–1953, 2020.

James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

Xu Han. Estimation and inference of dynamic structural factor models with over-identifying restrictions. *Journal of Econometrics*, 202(2):125–147, 2018.

Edward J. Hannan and Manfred Deistler. *The Statistical Theory of Linear Systems*. SIAM, 2012.

Lars P. Hansen and Thomas J. Sargent. Two difficulties in interpreting vector autoregressions. In Lars P. Hansen and Thomas J. Sargent, editors, *Rational Expectations Econometrics*, pages 77–119. Westview Press, 1991.

Lukas Hoesch, Adam Lee, and Geert Mesters. Locally robust inference for non-Gaussian SVAR models. *Quantitative Economics*, 15(2):523–570, 2024. doi: 10.3982/QE2274.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933. doi: 10.1037/h0071325.

Marek Jarociński. Estimating the Fed's unconventional policy shocks. *Journal of Monetary Economics*, 144:103548, 2024.

Carlos M. Jarque and Anil K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259, 1980. doi: 10.1016/0165-1765(80)90024-5. URL https://doi.org/10.1016/0165-1765(80)90024-5.

A.M. Kagan, Yu. V. Linnik, and C. R. Rao. *Characterization Problems in Mathematical Statistics*. John Wiley & Sons Ltd., 1973.

Thomas Kailath. *Linear Systems*. Prentice-Hall, Inc., 1980.

Sascha A. Keweloh, Mathias Klein, and Jan Prüser. Estimating fiscal multipliers by combining statistical identification with potentially endogenous proxies. *The Econometrics Journal*, 2025. doi: 10.1093/ectj/utaf027.

Andrzej Kocięcki and Marcin Kolasa. A solution to the global identification problem in DSGE models. *Journal of Econometrics*, 236(2):105477, 2023.

Andrzej Kocięcki, Christian Matthes, and Michele Piffer. A Scalable Framework for Statistical Identification in Structural VARs. Working paper, 2025.

Gary Koop. *Bayesian Econometrics*. John Wiley & Sons, Chichester, 2003.

Markku Lanne, Mika Meitz, and Pentti Saikkonen. Identification and Estimation of Non-Gaussian Structural Vector Autoregressions. *Journal of Econometrics*, 196(2):288–304, 2017. doi: 10.1016/j.jeconom.2016.06.002.

Daniel J. Lewis. Identification Based on Higher Moments in Macroeconometrics. *Annual Review of Economics*, 17:665–693, 2025.

Marco Lippi. Validating DSGE models with SVARs and high-dimensional dynamic factor models. *Econometric Theory*, 39(6):1273–1291, 2023.

Marco Lippi and Lucrezia Reichlin. VAR analysis, nonfundamental representations, Blaschke matrices. *Journal of Econometrics*, 63(1):307–325, 1994.

Robert B. Litterman, Danny Quah, and Thomas J. Sargent. Business cycle analysis with unobservable index models and the methods of the NBER. Unpublished manuscript, June 1984.

Matteo Luciani. Monetary policy and the housing market: a structural factor analysis. *Journal of Applied Econometrics*, 30(2):199–218, 2015.

Michael W. McCracken and Serena Ng. Fred-qd: A quarterly database for macroeconomic research. Working Paper 2020-005B, Federal Reserve Bank of St. Louis, March 2020. URL https://doi.org/10.20955/wp.2020.005. Revision date: March 2020.

Jorge A. Montiel Olea, Martin Plagborg-Møller, and Emma Qian. Svar identification from higher moments: Has the simultaneous causality problem been solved? *AEA Papers and Proceedings*, 112:481–485, 2022.

Karl Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2:559–572, 1901. doi: 10.1080/14786440109462720.

Mikkel Plagborg-Møller. Bayesian inference on structural impulse response functions. *Quantitative Economics*, 10(1):145–184, 2019. URL https://onlinelibrary.wiley.com/doi/abs/.

Mikkel Plagborg-Moller and Christian K. Wolf. Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*, 89(2):955–980, March 2021. doi: 10.3982/ECTA17813.

Gregory C. Reinsel. *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New

York, Inc., second edition, 1997.

Christian Ritter and Martin A. Tanner. Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association*, 87(419): 861–868, 1992.

Thomas J. Sargent and Christopher A. Sims. Business cycle modeling without pretending to have too much a priori economic theory. Working Paper 55, Federal Reserve Bank of Minneapolis, 1977. URL https://www.minneapolisfed.org/research/working-papers/business-cycle-modeling-without-pretending-to-have-too-much-a-priori-economic-theory.

Enrique Sentana and Gabriele Fiorentini. Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics*, 102(2):143–164, 2001. doi: 10.1016/S0304-4076(01)00051-3.

Christopher A. Sims. Comment on: Disentangling the channels of the 2007–09 recession, by James H. Stock and Mark W. Watson. *Brookings Papers on Economic Activity*, 1: 141–148, 2012.

James H. Stock and Mark W. Watson. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002a.

James H. Stock and Mark W. Watson. Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460): 1167–1179, 2002b. URL https://doi.org/10.1198/016214502388618960.

James H. Stock and Mark W. Watson. Implications of dynamic factor models for VAR analysis. NBER Working Papers No.11467, 2005.

James H. Stock and Mark W. Watson. Disentangling the Channels of the 2007-09 Recession. *Brookings Papers on Economic Activity*, 1:81–135, 2012.

James H. Stock and Mark W. Watson. Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In John B. Taylor and Harald Uhlig, editors, *Handbook of Macroeconomics*, volume 2, pages 415–525. Elsevier, 2016.

James H. Stock and Mark W. Watson. Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *Economic Journal*, 128(610):917–948, May 2018. doi: 10.1111/ecoj.12593.

Harry L. Trentelman, Anton A. Stoorvogel, and Malo Hautus. *Control Theory for Linear Systems*. Springer London, 2001.

Daniel F. Waggoner and Tao Zha. Likelihood preserving normalization in multiple equation models. *Journal of Econometrics*, 114(2):329–347, 2003.

Appendix

A.1 The proof of Theorem 1

We start with the following definition:

Definition 1. Two processes y_t and y_t^* are said to be distributionally equivalent if, for any positive integer n and any integers $t_1, t_2, ..., t_n$, $(y_{t_1}, ..., y_{t_n})$ and $(y_{t_1}^*, ..., y_{t_n}^*)$ have the same joint probability distribution function.

We first establish identification of the (two-sided) moving-average representation. The higher-order cumulant argument follows the approach of Chan and Ho (2004); Chan et al. (2006). With additive Gaussian noise, equality of the shock distributions requires a separate argument.

Proposition 5. *Under Assumptions 1, 2, 4, and 5, suppose that stochastic processes*

$$y_t = \sum_{i \in \mathbb{Z}} \Psi_i \epsilon_{t-i} + u_t, \qquad y_t^* = \sum_{i \in \mathbb{Z}} \Psi_i^* \epsilon_{t-i}^* + u_t^*,$$

are distributionally equivalent. Both structural transfers are real rational matrices with absolutely summable Laurent coefficients and column rank q on the unit circle. Then there exist real constants $c_j \neq 0$, a permutation π of $\{1, \ldots, q\}$, and integers $m(j)$ such that

$$\epsilon_{j,t-m(j)}^* \overset{d}{=} c_j \epsilon_{\pi(j),t}, \qquad \Psi_{i,j}^* = c_j^{-1} \Psi_{i-m(j),\pi(j)}, \tag{A.1}$$

for every $i \in \mathbb{Z}$, $t \in \mathbb{Z}$ and $j = 1, \ldots, q$, where $\Psi_{i,j}$ denotes column j of Ψ_i. Furthermore, the shock processes satisfy this equivalence jointly across all coordinates and dates.

Proof. Let $\kappa_j = \mathrm{cum}_r(\epsilon_{j,t}) \neq 0$, where r is the common order in Assumption 2. Mutual independence of the structural and Gaussian processes, combined with the vanishing of Gaussian cumulants of order $r \geq 3$, implies that the structural components share identical order-r joint cumulants across arbitrary coordinates and dates. Absolute summability of the filter coefficients together with finite absolute r-th moments from Assumption 2 guarantees L^r-convergence of the truncated filter sums, justifying the term-by-term interchange of the infinite summation and the r-th order cumulant. Applying a common stable filter to the observationally equivalent processes preserves equality of their joint distributions across dates.

Let $H(z)$ and $H^*(z)$ be rational left inverses of $\Psi(z)$ and $\Psi^*(z)$ with no poles on the unit circle and absolutely summable two-sided coefficients, as described after Assumption 5. Put $Q(z) = I_n - \Psi^*(z)H^*(z)$. Under the starred representation, $Q(L)y_t^* = Q(L)u_t^*$ is purely Gaussian. Under the unstarred representation,

$$Q(L)y_t = Q(L)\Psi(L)\epsilon_t + Q(L)u_t.$$

If any entry of $Q(z)\Psi(z)$ were non-zero, the corresponding non-Gaussian structural innova-

tion would enter as an independent summand of a Gaussian random variable. By Cramér's decomposition theorem, that innovation would be forced to be Gaussian, contradicting Assumption 1b. Thus, $Q(z)\Psi(z) = 0$. Reversing the roles of the two representations yields:

$$\begin{aligned} \Psi(z) &= \Psi^*(z)D(z), \qquad \Psi^*(z) = \Psi(z)E(z), \\ D(z) &= H^*(z)\Psi(z), \qquad E(z) = H(z)\Psi^*(z), \\ E(z)D(z) &= D(z)E(z) = I_q. \end{aligned} \tag{A.2}$$

All filters corresponding to (A.2) have absolutely summable coefficients. The identity $E(z)D(z) = I_q$ guarantees that $D(z)$ is invertible ($\det D(z) \neq 0$). Write $D(L) = \sum_{a\in\mathbb{Z}} D_a L^a$ and $E(L) = \sum_{a\in\mathbb{Z}} E_a L^a$, where D_a and E_a are $q \times q$ coefficient matrices.

Put $\zeta_t = D(L)\epsilon_t$. Applying $H^*(L)$ to the two observationally equivalent processes gives

$$\zeta_t + g_t \overset{d}{=} \epsilon_t^* + g_t^*, \qquad g_t = H^*(L)u_t, \quad g_t^* = H^*(L)u_t^*, \tag{A.3}$$

as an equality of joint distributions across coordinates and dates, with Gaussian noises independent of the respective structural processes. Thus, ζ_t and ϵ_t^* share identical order-r joint cumulants across all coordinates and dates. Fix $(i,h) \neq (k,l)$. Any order-r cumulant whose first two arguments are $\zeta_{i,h}$ and $\zeta_{k,l}$ and whose remaining $r-2$ arguments are coordinates of ζ at arbitrary dates equals the corresponding cumulant of ϵ^*, which is zero due to the mutual independence of the starred innovations. Because $\epsilon_t = E(L)\zeta_t$, rewriting $\epsilon_{j,0}$ in terms of ζ reduces $\operatorname{cum}(\zeta_{i,h}, \zeta_{k,l}, \epsilon_{j,0}, \ldots, \epsilon_{j,0})$ to a linear combination of such vanishing cross-cumulants, establishing that it equals zero. Evaluating the same cumulant by expanding its first two arguments via $\zeta_{i,h} = \sum_{a\in\mathbb{Z}} \sum_{b=1}^{q} (D_a)_{ib}\epsilon_{b,h-a}$ leaves only the terms where every argument is the identical innovation $\epsilon_{j,0}$. The coefficients of $\epsilon_{j,0}$ in the first two arguments are $(D_h)_{ij}$ and $(D_l)_{kj}$. Hence

$$\begin{aligned} 0 &= \operatorname{cum}\left(\zeta_{i,h}, \zeta_{k,l}, \underbrace{\epsilon_{j,0}, \ldots, \epsilon_{j,0}}_{r-2}\right) \\ &= \kappa_j (D_h)_{ij}(D_l)_{kj}. \end{aligned} \tag{A.4}$$

Since $\kappa_j \neq 0$, it implies $(D_h)_{ij}(D_l)_{kj} = 0$ for all $(i,h) \neq (k,l)$. Consequently, the sequence of column coefficients $\{(D_h)_{\cdot,j}\}_{h\in\mathbb{Z}}$ contains at most one non-zero scalar entry across all rows i and lag/leads h. Invertibility ($\det D(z) \neq 0$) rules out identically zero columns, forcing each column of $D(L)$ to contain exactly one non-zero scalar coefficient. Furthermore, no two columns can share a row index without leaving an entire row of $D(L)$ identically zero—violating the non-singularity of $D(z)$. Defining $\pi(j)$ as the unique column index containing the non-zero entry of row j, row j takes the form of a single lag monomial $c_j L^{-m(j)}$ in

column $\pi(j)$, with $c_j \neq 0$ and $m(j) \in \mathbb{Z}$. Hence, $D(L) = P \cdot \text{diag}\left(c_1 L^{-m(1)}, \ldots, c_q L^{-m(q)}\right)$ for a permutation matrix P. Componentwise, row j of $\zeta_t = D(L)\epsilon_t$ reduces to:

$$\zeta_{j,t} = D_{j,\pi(j)}(L)\epsilon_{\pi(j),t} = c_j \epsilon_{\pi(j),t+m(j)}$$

Post-multiplying $\Psi^*(z)$ by $D(z)$ recombines its columns. Since column $\pi(j)$ of $D(z)$ contains a single non-zero entry $c_j z^{-m(j)}$ in row j, column $\pi(j)$ of $\Psi(z)$ is simply column j of $\Psi^*(z)$ scaled and shifted:

$$\Psi_{\cdot,\pi(j)}(z) = c_j z^{-m(j)} \Psi^*_{\cdot,j}(z) \iff \Psi^*_{\cdot,j}(z) = c_j^{-1} z^{m(j)} \Psi_{\cdot,\pi(j)}(z)$$

Expanding both transfer function columns as power series in z:

$$\sum_{i\in\mathbb{Z}} \Psi^*_{i,j} z^i = c_j^{-1} z^{m(j)} \sum_{k\in\mathbb{Z}} \Psi_{k,\pi(j)} z^k = \sum_{k\in\mathbb{Z}} c_j^{-1} \Psi_{k,\pi(j)} z^{k+m(j)} = \sum_{i\in\mathbb{Z}} c_j^{-1} \Psi_{i-m(j),\pi(j)} z^i$$

Equating coefficients of z^i establishes $\Psi^*_{i,j} = c_j^{-1} \Psi_{i-m(j),\pi(j)}$ for all $i \in \mathbb{Z}$. This proves the second relation in (A.1).

It remains to identify the shock distributions. Fix coordinate j and date t in (A.3), and let $v_j = \text{Var}(g_{j,t})$ and $v_j^* = \text{Var}(g^*_{j,t})$. Write $\varphi_X(b) = E[\exp(ibX)]$ for the characteristic function of a scalar random variable X. Independence gives

$$\varphi_{\zeta_{j,t}}(b) e^{-v_j b^2/2} = \varphi_{\epsilon^*_{j,t}}(b) e^{-v_j^* b^2/2}, \qquad b \in \mathbb{R}.$$

If $v_j > v_j^*$, the distribution of $\epsilon^*_{j,t}$ is the convolution of the distribution of $\zeta_{j,t}$ with a normal distribution of variance $v_j - v_j^* > 0$. If $v_j < v_j^*$, the reverse relation holds. Both possibilities contradict Assumption 1d, since $\zeta_{j,t}$ is a nonzero multiple of a structural shock having no normal component. Therefore $v_j = v_j^*$ and the two characteristic functions agree everywhere. The permutation, rescaling, and component-specific time shifts preserve independence across coordinates and dates. Equality of the marginal distributions therefore gives equality of the joint distributions of $\{\zeta_t\}$ and $\{\epsilon_t^*\}$ across coordinates and dates. It follows $\epsilon^*_{j,t-m(j)} \overset{d}{=} \zeta_{j,t-m(j)} = c_j \epsilon_{\pi(j),t}$, establishing the relation between the shock distributions in (A.1). Q.E.D.

The proof uses only stationarity, finite variance, and Gaussianity of the noise process, together with independence from the structural shocks. Its conclusions about the transfer functions and shock distributions therefore also hold for singular Gaussian noise. Covariance subtraction below additionally identifies the joint distribution of the Gaussian component across dates. Uniqueness of an unrestricted $n \times n$ Gaussian transfer matrix additionally requires positive definiteness of the full innovation covariance; structural impulse-response

identification does not require this restriction.

Causality and the impact-rank condition remove the time shifts; unit variances then leave only component-specific signs.

Corollary 1. *Under Assumptions 1–6, suppose that the two representations in Proposition 5 are distributionally equivalent and that both $\Psi(L)$ and $\Psi^*(L)$ are causal. Then*

$$\epsilon^*_{j,t} \overset{d}{=} c_j \epsilon_{\pi(j),t}, \qquad \Psi^*_{i,j} = c_j \Psi_{i,\pi(j)}, \qquad c_j \in \{-1, 1\},$$

for every $i \geq 0$, t, and $j = 1, \ldots, q$.

Proof. Proposition 5 gives $\Psi^*_{i,j} = c_j^{-1} \Psi_{i-m(j),\pi(j)}$ for every integer i. Causality makes both coefficient sequences zero for $i < 0$, while the Assumption 6 makes every column of Ψ_0 and Ψ^*_0 nonzero. If $m(j) > 0$, setting $i = 0$ gives $\Psi^*_{0,j} = c_j^{-1} \Psi_{-m(j),\pi(j)} = 0$, contradicting the nonzero alternative impact column. If $m(j) < 0$, setting $i = m(j)$ gives $\Psi^*_{m(j),j} = c_j^{-1} \Psi_{0,\pi(j)} \neq 0$, contradicting causality because the horizon $m(j)$ is negative. Thus $m(j) = 0$. On the other hand, Proposition 5 states $\epsilon^*_{j,t-m(j)} \overset{d}{=} c_j \epsilon_{\pi(j),t}$. Setting $m(j) = 0$, the equality of shock distributions and the variance normalization give: $1 = \mathrm{Var}(\epsilon^*_{j,t}) = c_j^2 \, \mathrm{Var}(\epsilon_{\pi(j),t}) = c_j^2$ for each j. Hence $c_j \in \{-1, 1\}$ and $c_j = c_j^{-1}$. Q.E.D.

Now we are in position to state our first important result:

Proposition 6. *Under Assumptions 1-6, $y_t = \sum_{i=0}^{\infty} \Psi_i \epsilon_{t-i} + \sum_{i=0}^{r} \Upsilon_i e_{t-i}$ and $y^*_t = \sum_{i=0}^{\infty} \Psi^*_i \epsilon^*_{t-i} + \sum_{i=0}^{r} \Upsilon^*_i e^*_{t-i}$ are distributionally equivalent if and only if 1) $\Psi^*_i = \Psi_i PD$; 2) $\epsilon^*_t \overset{d}{=} DP' \epsilon_t$ (where P is the permutation matrix and D is diagonal matrix with ± 1 elements); and 3) $\Upsilon^*_i = \Upsilon_i$ and $e^*_t \overset{d}{=} e_t$.*

Proof. Corollary 1 identifies the structural transfer up to a signed permutation and establishes the corresponding equality of shock distributions. Write $v_t = \Psi(L)\epsilon_t$. The structural and Gaussian processes are independent, so

$$\Gamma_y(k) = \Gamma_v(k) + \Gamma_u(k).$$

where $\Gamma_x(k) = cov(x_t, x_{t-k})$. Observational equivalence gives $\Gamma_{y^*}(k) = \Gamma_y(k)$. Since both structural shock covariance matrices are I_q and the transfers differ only by a signed permutation, $\Gamma_{v^*}(k) = \Gamma_v(k)$. Subtraction identifies $\Gamma_u(k)$ at every lag and hence the joint distribution of the Gaussian component across dates and its spectral density. Under Assumption 1a, this spectral density has the normalized causal, stable, invertible factorization

$$S_u(\omega) = \frac{1}{2\pi} \Upsilon(e^{-i\omega}) \Omega \Upsilon(e^{i\omega})'.$$

Positive definite Ω and $\Upsilon_0 = I_n$, together with causality and invertibility, make this factorization unique; see Theorem 1.3.3 of Hannan and Deistler (2012). Thus $\Upsilon_i^* = \Upsilon_i$ and $\Omega^* = \Omega$. The converse follows by substitution and process independence. Q.E.D.

Let us define the following matrix:

$$\mathcal{C} = [\mathbf{B} \vdots A\mathbf{B} \vdots A^2\mathbf{B} \vdots \ldots \vdots A^{\tau q-1}\mathbf{B}]. \tag{A.5}$$

To proceed further we need to establish the following fact:

Lemma 1. *$\mathcal{C}$ has full row rank.*

Proof: The case $p = 1, s = 0$ is trivial since $\mathbf{B} = B$. Otherwise, w.l.o.g., we can consider A matrix from (3) setting the last $s + 1 - p$ Φ_i's to zeros to cover the case (5) if necessary. We have:

$$\mathcal{C} = [\begin{bmatrix} B \\ 0_{q(\tau-1)\times q} \end{bmatrix} \vdots \begin{bmatrix} \Phi_1 B \\ B \\ 0_{q(\tau-2)\times q} \end{bmatrix} \vdots \begin{bmatrix} \times \\ \Phi_1 B \\ B \\ 0_{q(\tau-3)\times q} \end{bmatrix} \vdots \ldots \vdots \begin{bmatrix} \times \\ \Phi_1 B \\ B \\ 0_{q\times q} \end{bmatrix} \vdots \begin{bmatrix} \times \\ \times \\ \Phi_1 B \\ B \end{bmatrix} \vdots \ldots \vdots \times].$$

Hence we factorized $\mathcal{C} = [C_1 \vdots C_2]$, where C_1 is the upper block triangular (quadratic) matrix, with the same diagonal block B. If B is nonsingular then C_1 is also nonsingular hence $\mathcal{C}$ has full row rank. Q.E.D.

We are in position to complete the proof of Theorem 1. From Proposition 6 we have $\bar{\Psi}_i = \Psi_i P_s$, hence $\bar{C}\bar{A}^i\bar{\mathbf{B}} = CA^i\mathbf{B}P_s$, for $i = 0, 1, 2, ...$, where $P_s := PD$ is the signed permutation matrix. It follows:

$$\bar{\mathcal{O}}\bar{\mathcal{C}} = \begin{bmatrix} \bar{C}\bar{\mathbf{B}} & \bar{C}\bar{A}\bar{\mathbf{B}} & \bar{C}\bar{A}^2\bar{\mathbf{B}} & \dots & \bar{C}\bar{A}^{\tau q-1}\bar{\mathbf{B}} \\ \bar{C}\bar{A}\bar{\mathbf{B}} & \bar{C}\bar{A}^2\bar{\mathbf{B}} & \bar{C}\bar{A}^3\bar{\mathbf{B}} & \dots & \bar{C}\bar{A}^{\tau q}\bar{\mathbf{B}} \\ \bar{C}\bar{A}^2\bar{\mathbf{B}} & \bar{C}\bar{A}^3\bar{\mathbf{B}} & \bar{C}\bar{A}^4\bar{\mathbf{B}} & \dots & \bar{C}\bar{A}^{\tau q+1}\bar{\mathbf{B}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{C}\bar{A}^{\tau q-1}\bar{\mathbf{B}} & \bar{C}\bar{A}^{\tau q}\bar{\mathbf{B}} & \bar{C}\bar{A}^{\tau q+1}\bar{\mathbf{B}} & \dots & \bar{C}\bar{A}^{2\tau q-2}\bar{\mathbf{B}} \end{bmatrix},$$

$$= \begin{bmatrix} C\mathbf{B} & CA\mathbf{B} & CA^2\mathbf{B} & \dots & CA^{\tau q-1}\mathbf{B} \\ CA\mathbf{B} & CA^2\mathbf{B} & CA^3\mathbf{B} & \cdots & CA^{\tau q}\mathbf{B} \\ CA^2\mathbf{B} & CA^3\mathbf{B} & CA^4\mathbf{B} & \dots & CA^{\tau q+1}\mathbf{B} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{\tau q-1}\mathbf{B} & CA^{\tau q}\mathbf{B} & CA^{\tau q+1}\mathbf{B} & \dots & CA^{2\tau q-2}\mathbf{B} \end{bmatrix} (I_{\tau q} \otimes P_s) = \mathcal{O}\mathcal{C}\tilde{P}, \tag{A.6}$$

where $\tilde{P} = I_{\tau q} \otimes P_s$. The next proof step is standard and follows Kailath (1980), pp. 364-365 with slight modifications. Since $\mathcal{O}$ has full column rank, it posseses the left inverse, hence $\bar{\mathcal{O}}\bar{\mathcal{C}} = \mathcal{O}\mathcal{C}\tilde{P} \Rightarrow \mathcal{C}\tilde{P} = \underbrace{(\mathcal{O}'\mathcal{O})^{-1}\mathcal{O}'\bar{\mathcal{O}}}_{T_1}\bar{\mathcal{C}} = T_1\bar{\mathcal{C}}$. Note that since $\mathcal{C}\tilde{P}$ and $\bar{\mathcal{C}}$ have full row rank i.e. τq, we have $\tau q = rank(\mathcal{C}\tilde{P}) = rank(T_1\bar{\mathcal{C}}) = rank(T_1)$, hence T_1 is nonsingular. On the other hand since $\mathcal{C}$ has full row rank $\bar{\mathcal{O}}\bar{\mathcal{C}}\tilde{P}' = \mathcal{O}\mathcal{C} \Rightarrow \mathcal{O} = \bar{\mathcal{O}}\underbrace{\bar{\mathcal{C}}\tilde{P}'\mathcal{C}'(\mathcal{C}\mathcal{C}')^{-1}}_{T} = \bar{\mathcal{O}}T$. Next, We have $T_1T = (\mathcal{O}'\mathcal{O})^{-1}\mathcal{O}'\underbrace{\bar{\mathcal{O}}\bar{\mathcal{C}}\tilde{P}'}_{=\mathcal{O}\mathcal{C}}\mathcal{C}'(\mathcal{C}\mathcal{C}')^{-1} = I_{\tau q}$, hence $T = T_1^{-1}$ (it follows T is nonsingular too). From $\bar{\mathcal{O}} = \mathcal{O}T^{-1}$ we get $\bar{C} = CT^{-1}$. Uniqueness of T follows from $\bar{\mathcal{O}} = \mathcal{O}T^{-1}$ under assumption that $\mathcal{O}$ has full column rank. Further, since $\mathcal{C}\tilde{P} = T_1\bar{\mathcal{C}} = T^{-1}\bar{\mathcal{C}}$ we get $\bar{\mathbf{B}} = T\mathbf{B}P_s$. Last since $\bar{\mathcal{O}}\bar{A}\bar{\mathcal{C}} = \mathcal{O}A\mathcal{C}\tilde{P} \Rightarrow (\mathcal{O}'\mathcal{O})^{-1}\mathcal{O}'\bar{\mathcal{O}}\bar{A}\bar{\mathcal{C}}\tilde{P}'\mathcal{C}'(\mathcal{C}\mathcal{C}')^{-1} = A$, it yields $T_1\bar{A}T = A \Leftrightarrow \bar{A} = TAT^{-1}$. Now $\begin{bmatrix} \bar{B} \\ 0_{q(\tau-1)\times q} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} BP_s \\ 0_{q(\tau-1)\times q} \end{bmatrix} = \begin{bmatrix} T_{11}BP_s \\ T_{21}BP_s \end{bmatrix}$. It follows $\bar{B} = T_{11}BP_s$. Since BP_s is nonsingular it also implies $T_{21} = 0_{q(\tau-1)\times q}$.

In the next step we will show that T is block diagonal with the same block. Since $\bar{A} = TAT^{-1} \Leftrightarrow \bar{A}T = TA$ we have:[15]

[15]The proof to follow assumes $\tau = max\{p, s+1\} = p \geq s+1$ just for visual convenience. Setting $\tau = max\{p, s+1\} = s+1 > p$ does not change further conclusions.

$$\begin{bmatrix} \bar{\Phi}_1 & \bar{\Phi}_2 & \dots & \bar{\Phi}_p \\ I_q & 0 & \cdots & 0 \\ \vdots & \ddots & \dots & \vdots \\ 0 & \dots & I_q & 0 \end{bmatrix} \begin{bmatrix} T_{11} & T_{12} & T_{13} & \dots & T_{1p} \\ 0_{q\times q} & T_{22} & T_{23} & \dots & T_{2p} \\ 0_{q\times q} & T_{32} & T_{33} & \dots & T_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{q\times q} & T_{p2} & T_{p3} & \dots & T_{pp} \end{bmatrix} = \\ = \begin{bmatrix} T_{11} & T_{12} & T_{13} & \dots & T_{1p} \\ 0_{q\times q} & T_{22} & T_{23} & \dots & T_{2p} \\ 0_{q\times q} & T_{32} & T_{33} & \dots & T_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{q\times q} & T_{p2} & T_{p3} & \dots & T_{pp} \end{bmatrix} \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I_q & 0 & \cdots & 0 \\ \vdots & \ddots & \dots & \vdots \\ 0 & \dots & I_q & 0 \end{bmatrix}, \tag{A.7}$$

where each T_{ij} is $q \times q$. Multiplying all terms and taking into account all rows except the first q we have:

$$\begin{bmatrix} T_{11} & T_{12} & T_{13} & \dots & T_{1(p-1)} & T_{1p} \\ 0_{q\times q} & T_{22} & T_{23} & \dots & T_{2(p-1)} & T_{2p} \\ 0_{q\times q} & T_{32} & T_{33} & \dots & T_{3(p-1)} & T_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{q\times q} & T_{(p-1)2} & T_{(p-1)3} & \dots & T_{(p-1)(p-1)} & T_{(p-1)p} \end{bmatrix}, \\ = \begin{bmatrix} T_{22} & T_{23} & T_{24} & \dots & T_{2p} & 0_{q\times q} \\ T_{32} & T_{33} & T_{34} & \dots & T_{3p} & 0_{q\times q} \\ T_{42} & T_{43} & T_{44} & \dots & T_{4p} & 0_{q\times q} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ T_{p2} & T_{p3} & T_{p4} & \dots & T_{pp} & 0_{q\times q} \end{bmatrix}. \tag{A.8}$$

Comparing two sides in (A.8) we have $T_{1p} = T_{2p} = ... = T_{(p-1)p} = 0$. Further, $T_{1(p-1)} = T_{2p} = 0$, $T_{2(p-1)} = T_{3p} = 0$, $T_{3(p-1)} = T_{4p} = 0, \dots, T_{(p-2)(p-1)} = T_{(p-1)p} = 0$. Then $T_{1(p-2)} = T_{2(p-1)} = 0$, $T_{2(p-2)} = T_{3(p-1)} = 0, \dots , T_{(p-3)(p-2)} = T_{(p-2)(p-1)} = 0$. In general it holds $T_{ij} = 0$, for $i < j$. In a similar way we have: $T_{32} = T_{42} = \dots = T_{p2} = 0$. Further $T_{43} = T_{32} = 0, T_{53} = T_{42} = 0, \dots , T_{p3} = T_{(p-1)2} = 0$ and $T_{54} = T_{43} = 0, \dots , T_{p4} = T_{(p-1)3} = 0$. In general it holds $T_{ij} = 0$, for $i > j$. As a result $T = diag(T_{11}, T_{22}, ..., T_{pp})$. Lastly we directly read off from (A.8): $T_{11} = T_{22}$, $T_{22} = T_{33}$, $T_{33} = T_{44}$, ..., $T_{(p-1)(p-1)} = T_{pp}$, hence diagonal blocks are the same.

Denote the common diagonal block by T. Then $\bar{\Phi}_i T = T\Phi_i$, and hence $\bar{\Phi}_i = T\Phi_i T^{-1}$ for each $i = 1, \dots, p$. From $\bar{C} = C(I_\tau \otimes T^{-1})$ and the block structure of C, we obtain $\bar{\Lambda}_i = \Lambda_i T^{-1}$ for each $i = 0, \dots, s$.

Lastly, Proposition 6 identifies $\theta_u = (\Upsilon, \Omega)$ as defined in the main text. If u_t follows a pure VMA process then we are done. If u_t follows a pure VAR process then under Assumption 1a, there is a 1-1 mapping between VAR parameters, say $\Pi_1, ..., \Pi_l$, and its IRFs i.e. $\{\Upsilon_i\}$. We conclude that $\Pi_1, ..., \Pi_l$ (and Ω) are globally identified. In the case of mixed VARMA process for u_t, some further (but well known) assumptions are needed in order to uniquely pin down their parameters from the infinite order MA representation $u_t = \Upsilon(L)e_t$, see e.g. Reinsel (1997), pp. 37-41 and 61-67.

The implication in the other direction holds trivially by inserting $\bar{\Lambda}_i = \Lambda_i T^{-1}$, $\bar{\Phi}_i = T\Phi_i T^{-1}$, $\bar{B} = TBP_s$ into the MA representation (8).

Q.E.D.

A.2 Verification for Student-t shocks

Lemma 2. *A standardized Student-t_ν variable with finite $\nu > 2$ has no nondegenerate independent Gaussian summand. For $\nu > 4$, its fourth cumulant is $6/(\nu - 4) > 0$.*

Proof. Write $X = \sqrt{V}Z$, where $V = (\nu-2)/Q$, $Q \sim \chi^2_\nu$, and $Z \sim N(0,1)$ is independent of Q. Then $E[V] = 1$ and $\mathrm{Var}(X) = 1$. For every $\delta > 0$, the unbounded support of Q gives $p_\delta := \Pr(V \leq \delta) > 0$. The characteristic function therefore satisfies

$$\varphi_X(b) = E[e^{-Vb^2/2}] \geq p_\delta e^{-\delta b^2/2} > 0.$$

If $X \overset{d}{=} W + G$ for independent W and $G \sim N(0, \sigma^2)$ with $\sigma^2 > 0$, then $|\varphi_X(b)| \leq e^{-\sigma^2 b^2/2}$. A nonzero Gaussian mean can be absorbed into W. Choosing $0 < \delta < \sigma^2$ would imply $p_\delta \leq e^{-(\sigma^2-\delta)b^2/2}$ for every b, which is impossible as $|b| \to \infty$. For $\nu > 4$, the chi-square inverse moments give $E[V^2] = (\nu-2)/(\nu-4)$, so

$$\mathrm{cum}_4(X) = E[X^4] - 3 = 3\frac{\nu-2}{\nu-4} - 3 = \frac{6}{\nu-4}.$$

Thus independent Student-t shocks with possibly different finite $\nu_j > 4$ satisfy the Gaussian-summand exclusion and the common choice $r = a = 4$. Their normal scale-mixture representation used for computation is consistent with both restrictions. Q.E.D.