Braun, Robin

Article

The importance of supply and demand for oil prices: Evidence from non-Gaussianity

Quantitative Economics

Provided in Cooperation with:
The Econometric Society

Suggested Citation: Braun, Robin (2023) : The importance of supply and demand for oil prices: Evidence from non-Gaussianity, Quantitative Economics, ISSN 1759-7331, The Econometric Society, New Haven, CT, Vol. 14, Iss. 4, pp. 1163-1198, https://doi.org/10.3982/QE2091

This Version is available at:
https://hdl.handle.net/10419/296352

Supplement to "The importance of supply and demand for oil prices: Evidence from non-Gaussianity"

(*Quantitative Economics*, Vol. 14, No. 4, November 2023, 1163–1198)

Robin Braun
Board of Governors of the Federal Reserve System

This document provides the Online Appendix to the paper "The importance of supply and demand for oil prices: Evidence from non-Gaussianity." It contains details on how to conduct Bayesian inference in the SVAR–DPMM via Markov Chain Monte Carlo methods (Appendix A), on convergence properties of the resulting MCMC (Appendix B), an illustration how to use the marginal likelihood estimator (Appendix C), supplementary figures to the oil market model (Appendix D), a discussion of the relationship between elasticities and variance decompositions in the oil market (Appendix E), and robustness analysis of the empirical findings with respect to alternative error specifications (Appendix F).

Appendix A: Bayesian inference

A.1 *General Markov chain Monte Carlo algorithm*

This part of the Appendix covers a generic MCMC algorithm to conduct inference for an A type of SVAR model where shocks follow Dirichlet Process mixture models (DPMM). Let $\alpha_+ = \text{vec}(\text{A}_+)$ and $A_{i\bullet}$ the ith row of A. Further, let $A_{i\bullet}' = w_i + W_i a_i$ where a_i is a vector of r_i free elements, W_i a $K \times r_i$ selection matrix of zeros and ones, and w_i an $K \times 1$ vector containing either zero or the constrained values. Then, following Section 2, the full hierarchical model (including prior distributions) reads for $i = 1, \ldots, K$ and $t = 1, \ldots, T$:

$$\text{A}(y_t - \text{A}_+ x_t) = \varepsilon_t, \tag{S.1}$$

$$\varepsilon_{it} | \theta_{it} \sim \mathcal{N}\big(\mu_{it}, \sigma_{it}^2\big), \tag{S.2}$$

$$a_i \sim p(a_i), \tag{S.3}$$

$$\alpha_+ \sim \mathcal{N}(m_{\alpha_+}, V_{\alpha_+}), \tag{S.4}$$

$$\theta_{it} \sim G_i, \tag{S.5}$$

$$G_i \sim \text{DP}(G_{i0}, \alpha_i), \tag{S.6}$$

$$G_{i0} \sim \mathcal{NIG}(s_i/2, S_i/2, m_i, \tau_i), \tag{S.7}$$

where $x_t = [y_{t-1}', \ldots, y_{t-p}', 1]'$ and $\text{A}_+ = [A_1, A_2, \ldots, A_p, c]$. Although optional, I outline the algorithm under the assumption that further hyperpriors are specified:

$$\alpha_i \sim \mathcal{G}(a_{\alpha_i}, b_{\alpha_i}), \tag{S.8}$$

Robin Braun: robin.a.braun@frb.gov

 https://doi.org/10.3982/QE2091

$$\tau_i \sim i\mathcal{G}(a_{\tau_i}, b_{\tau_i}), \tag{S.9}$$

$$m_i \sim \mathcal{N}(m_{m_i}, V_{m_i}). \tag{S.10}$$

In case that they are treated as fixed values, the corresponding steps in the MCMC algorithm can simply be skipped.

Define the set of parameters by $\varphi = \{\alpha_+, a_i, \alpha_i, \tau_i, m_i, i = 1, \dots, K\}$ and the collection of auxiliary mixing parameters by $\Theta = \{\theta_{it}, i = 1, \dots, K, t = 1, \dots, T\}$. Also, define the augmented set of parameters by $\xi = \{\varphi, \Theta\}$, and denote by ξ_{-x} all parameters in ξ but x. Based on arbitrary initial values, the following MCMC algorithm eventually generates draws $\xi^{(l)}, l = 1, 2, \dots$ from the posterior distribution of $p(\xi|Y)$, by cycling through blocks of conditional distributions of subsets in ξ. The algorithm involves the following steps:

1. For $i = 1, \dots, K$, draw from the mixture parameters $\theta_{it}, t = 1, \dots, T$. To achieve better mixing properties of the Markov chain, this step is performed using Algorithm 3 of Neal (2000). Neal further splits the mixing parameters into two components: $\theta_{it} = \theta^\star_{i,c_{it}}$, where c_{it} are latent discrete assignment variables and $\theta^\star_{ij}$ are unique cluster parameters. Given the conjugate base distribution, it is possible to integrate over the cluster parameters to increase efficiency. This yields the following two steps:

   (a) Draw from the conditional of the assignment variables $p(c_{it}|Y, \xi_{-\{c_{it},\theta^\star\}})$ for $t = 1, \dots, T$. These are discrete probability distributions given by

   $$\begin{aligned} &P(c_{it} = c_{ij}, j = 1, \dots, k_i | c_{i,-t}, \varepsilon_t) \\ &\quad = b \frac{n_{-t,c_{ij}}}{T - 1 + \alpha_i} \int F(\varepsilon_{it}|\theta)\, dH_{-t,c_{ij}}(\theta), \end{aligned} \tag{S.11}$$

   $$\begin{aligned} &P(c_{it} \neq c_{ij} \text{ for all } j \neq t | c_{i,-t}, \varepsilon_t) \\ &\quad = b \frac{\alpha}{T - 1 + \alpha_i} \int F(\varepsilon_{it}|\theta)\, dG_0(\theta), \end{aligned} \tag{S.12}$$

   where $c_{i,-t} = \{c_{ij}, j \neq t\}$, $c_{ij}, j = 1, \dots, k_i$ are the unique values in $c_{i,-t}$ each of count $n_{-t,c_{ij}}$. Furthermore, b is a normalizing constant and $H_{-t,c_{ij}}$ is the posterior distribution of θ based on prior G_0 and all shocks of $\varepsilon_{i,-t} = \{\varepsilon_{ij}, j \neq t\}$ assigned to cluster c_{ij}. Given the conjugate Base distribution G_0, both integrals are tractable and given in closed form. Hence, drawing from the distribution is straightforward.

   (b) Conditional on the assignment variables, the second step is to draw the (active) cluster parameters $p(\theta^\star_{ij}|Y, \xi_{-\theta^\star_{ij}}), j = 1, \dots, k_i$, which are given by

   $$\begin{aligned} \sigma^{\star 2}_{ij} &\sim i\mathcal{G}(\overline{a}_{ij}, \overline{b}_{ij}), \\ \mu^\star_{ij} &\sim \mathcal{N}\big(\overline{m}_{ij}, \sigma^{\star 2}_{ij}\overline{V}_{ik}\big) \end{aligned}$$

with moments defined as follows:

$$\overline{a}_{ij} = \frac{s_i + T_{ij}}{2}, \quad \text{with } T_{ij} = \sum_{t=1}^{T} \mathbb{1}\{c_{it} = j\},$$

$$\overline{b}_{ik} = 0.5\left(S_i + \frac{m_i^2}{\tau_i} + \sum_{t:c_{it}=j} {\varepsilon_{it}}^2 - \frac{\overline{m}_{ij}^2}{\overline{V}_{ij}}\right),$$

$$\overline{V}_{ij} = \left(\frac{1}{\tau_i} + T_{ij}\right)^{-1},$$

$$\overline{m}_{ij} = \overline{V}_{ij}\left(\frac{m_i}{\tau_i} + \sum_{t:c_{it}=j} \varepsilon_{it}\right).$$

2. The next step is to sample the hyperparameters $\{\alpha_i, m_i, \tau_i\}$ $(i = 1, \ldots, K)$ from their conditionals, which exactly follows Escobar and West (1995).

   (a) With respect to α_i, the procedure is given as follows. First, draw an auxiliary variable d_i and conditional on d_i, the concentration parameters α_i for $i = 1, \ldots, K$:

$$p(d_i|\alpha_i) \sim \text{Beta}(\alpha_i + 1, T),$$
$$p(\alpha_i|Y, \xi_{-\alpha}, d_i) \sim \pi_{d_i}\mathcal{G}\big(a_{\alpha_i} + k_i, b_{\alpha_i} - \log(d_i)\big) + (1 - \pi_{d_i})\mathcal{G}\big(a_{\alpha_i} + k_i - 1, b_{\alpha_i} - \log(d_i)\big),$$

   where π_{d_i} is defined as

$$\frac{\pi_{d_i}}{1 - \pi_{d_i}} = \frac{a_{\alpha_i} + k_i - 1}{T\big(b_{\alpha_i} - \log(d_i)\big)}.$$

   (b) Draw $p(m_i|Y, \xi_{-m_i}) \sim \mathcal{N}(\overline{m}_{m,i}, \overline{V}_{m,i})$ where $\overline{V}_{m,i} = \tau_i x_{\sigma_i^\star} V_{\sigma_i^\star}$, $\overline{m}_{m,i} = (1 - x_{\sigma_i^\star})m_{m_i} + x_{\sigma_i^\star} V_{\sigma_i^\star}(\sum_{j=1}^{k_i} \sigma_{ij}^{\star -2}\mu_{ij}^\star)$ for $V_{\sigma_i^\star}^{-1} = \sum_{j=1}^{k_i} \sigma_{ij}^{\star -2}$, and $x_{\sigma_i^\star} = V_{m_i}/(m_{m_i} + \tau_i V_{\sigma_i^\star})$.

   (c) Draw $p(\tau_i|Y, \xi_{-\tau_i}) \sim \mathcal{G}(\overline{a}_{\tau,i}, \overline{b}_{\tau,i})$ where $\overline{a}_{\tau,i} = a_{\tau_i} + \frac{k_i}{2}$ and $\overline{b}_{\tau,i} = b_{\tau_i} + \frac{\sum_{j=1}^{k_i}(\mu_{ij}^\star - m_i)/\sigma_{ij}^{\star 2}}{2}$.

3. The third step involves drawing from each row in A via an independent Metropolis–Hastings step, which is exact under a uniform prior. Recall that each row is given by $A_{i\bullet}' = w_i + W_i a_i$, where a_i is a vector of r_i free elements, W_i a $K \times r_i$ selection matrix, and w_i an $K \times 1$ vector containing constrained values. To develop a proposal distribution, I assume a uniform prior that is $p^\star(a_i) \propto c$. Let $U = [u_1 : \ldots : u_T]'$ for $u_t = y_t - \text{A}_+ x_t$, $\mu_i = [\mu_{i1}, \ldots, \mu_{iT}]'$ and $\Sigma_i = \text{diag}([\sigma_{i1}^2, \ldots, \sigma_{iT}^2])$. Then the conditional posterior is proportional to

$$p^\star(a_i|Y, \xi_{-a_i}) \propto |\text{A}|^T \exp\left(-\frac{T}{2}(a_i - \mu_{a_i})'\Omega_{a_i}^{-1}(a_i - \mu_{a_i})\right),$$

where $\Omega_{a_i}^{-1} = T^{-1}W_i'U\Sigma_i^{-1}UW_i$, $\mu_{a_i} = (W_i'U'\Sigma_i^{-1}UW_i)^{-1}W_i'Y'(\mu_i - Uw_i)$. Chan, Koop, and Yu (2023) derive an efficient way to sample from $p^\star(a_i|Y, \xi_{-a_i})$ for $w_i = 0$, which builds on previous work of Waggoner and Zha (2003) and Villani (2009). In the following, I generalize the sampling scheme for w_i containing nonzero elements. Hereby, I closely follow the exposition and notation of Villani (2009).

DEFINITION 1. A random variable X follows the generalized absolute normal distribution GAN(a, b, μ, ρ) if it has density function:

$$p_{\text{GAN}}(x; a, b, \mu, \rho) = c|a + bx|^{\frac{1}{\rho}} \exp\left(-\frac{1}{2\rho}(x-\mu)^2\right), \quad x \in R,$$

where c is a normalizing constant, $\rho \in R^+$, $a \in R$, $b \in R$, and $\mu \in R$.

Note that for $a = 0$, the absolute normal distribution is obtained as defined in Villani (2009).

In the following, denote B_{-i} the matrix B with the ith column deleted, $B_\perp$ the orthogonal complement of B, and chol(B) the Choleski decomposition of B such that chol(B) chol$(B)' = B$. Also, denote by $\|\cdot\|$ the Euclidean norm and $\overset{d}{=}$ equality in distribution.

PROPOSITION 1. *Under prior $p^\star(a_i)$, the conditional posterior $p^\star(a_i|Y, \xi_{a_i})$ is given by*

$$a_i \overset{d}{=} R_i \sum_{j=1}^{r_i} \gamma_j v_j, \tag{S.13}$$

where $R_i = \text{chol}(\Omega_{a_i})$, $\gamma_1 \sim \text{GAN}(\hat{a}, \hat{b}, \hat{\gamma}_1, T^{-1})$, $\gamma_j \sim \mathcal{N}(\hat{\gamma}_j, T^{-1})$ for $j = 2, \ldots, r_i$, $\hat{\gamma}_j = \mu_{a_i}'R_i'^{-1}v_j$, $v_1 = R_iW_i'(\text{A})_{-i\perp}/\|R_iW_i'(\text{A})_{-i\perp}\|$, $(v_2, \ldots, v_{r_i}) = v_{1\perp}$, $\hat{a} = \det([A_{1\bullet}', \ldots, w_i, \ldots, A_{K\bullet}'])$, and $\hat{b} = \det([A_{1\bullet}', \ldots, W_iR_iv_1, \ldots, A_{K\bullet}'])$.

PROOF. For the decomposition $a_i = R_i\sum_{j=1}^{r_i}\gamma_j v_j$, Waggoner and Zha (2003) shows that

$$p^\star(a_i|Y, \xi_{a_i}) \propto |A|^T \exp\left(-\frac{T}{2}\left[\sum_{j=1}^{r_i}(\gamma_j - \hat{\gamma}_j)^2\right]\right),$$

where $\hat{\gamma}_j = \mu_{a_i}'R_i'^{-1}v_j$. Next, note that the determinant A is given by

$$\begin{aligned}
|\text{A}| &= \det\left[A_{1\bullet}'|\cdots|w_i + W_iR_i\sum_{j=1}^{r_i}\gamma_j v_j|\cdots|A_{K\bullet}'\right] \\
&= \det\left[A_{1\bullet}'|\cdots|w_i|\cdots|A_{n\bullet}'\right] + \sum_{j=1}^{r_i}\gamma_j \det\left[A_{1\bullet}'|\cdots|W_iR_i\sum_{j=1}^{r_i}v_j|\cdots|A_{K\bullet}'\right]
\end{aligned}$$

$$= \underbrace{\det\left[A'_{1\bullet}|\cdots|w_i|\cdots|A'_{K\bullet}\right]}_{\hat{a}} + \underbrace{\det\left[A'_{1\bullet}|\cdots|W_i R_i \sum_{j=1}^{r_i} v_1|\cdots|A'_{K\bullet}\right]}_{\hat{b}} \gamma_1,$$

where the last line follows by construction of $(v_2, \ldots, v_{r_i})$ spanning the same space than $(\mathrm{A}')_{-i}$. The result follows that

$$p^\star(a_i|Y, \xi_{-a_i}) \propto |\hat{a} + \hat{b}\gamma_1|^T \exp\left(-\frac{T}{2}(\gamma_1 - \hat{\gamma}_1)^2\right) \prod_{j=2}^{r_i} \exp\left(-\frac{T}{2}(\gamma_j - \hat{\gamma}_j)^2\right) \qquad \square$$

In order to sample efficiently from $p^\star(a_i|Y, \xi_{-a_i})$, I follow Villani (2009) and use a mixture of two Gaussians to approximate $\gamma_1 \sim \mathrm{GAN}(\hat{a}, \hat{b}, \hat{\gamma}_1, T^{-1})$. The motivation for the approximation follows from the fact that $\mathrm{GAN}(a, b, \mu, \rho)$ is bimodal. Specifically, two roots are given at

$$\frac{b\mu - a \pm \sqrt{\left((a - b\mu)^2 + 4b(a\mu + b)\right)}}{2b},$$

Corresponding curvature is given by

$$-\left[\frac{d^2}{dx^2} \ln p_{\mathrm{GAN}}(x; a, b, \mu, \rho)\right]^{-1}\Big|_{x=x_0} = \rho \frac{(a + bx_0)^2}{a^2 + 2abx_0 + b^2x_0^2 + b^2}.$$

Hence, the following normal approximation:

$$p_{\mathrm{GAN}}(x; a, b, \mu, \rho) \approx w\mathcal{N}\left(x, \mu_1, \sigma_1^2\right) + (1 - w)\mathcal{N}\left(x, \mu_2, \sigma_2^2\right),$$

where $\mu_1 = \frac{b\mu - a + \sqrt{((a-b\mu)^2 + 4b(a\mu+b))}}{2b}$, $\mu_2 = \frac{b\mu - a - \sqrt{((a-b\mu)^2 + 4b(a\mu+b))}}{2b}$, $\sigma_i^2 = \rho \frac{(a+b\mu_i)^2}{a^2 + 2ab\mu_i + b^2\mu_i^2 + b^2}$, $i = 1, 2$, and $w = \frac{p_{\mathrm{GAN}}(\mu_1; a, b, \mu, \rho)}{\sum_{j=1}^2 p_{\mathrm{GAN}}(\mu_j; a, b, \mu, \rho)}$ is set to take into account different heights of the density at the modes. Similar to Villani (2009), I find that this approximation work extremely well in practice and can be taken as exact. If desired, however, one might obtain an exact sampler by correcting for the approximation error in the Metropolis– Hastings step.

Such a step is necessary when working with a more general prior for $p(a_i)$ than the uniform used to derive $p^\star(a_i|Y, \xi_{a_i})$. In most cases, it will suffice to use a Metropolis– Hastings step that corrects for the fact that $p^\star(a_i|Y, \xi_{a_i})$ is missing the information from a nonuniform prior $p(a_i)$. Denote by $a_i^{(l-1)}$ the current state of the Markov chain and by $a_i' \sim p^\star(a_i|Y, \xi_{a_i})$ the proposed value under a uniform prior. Then the MH algorithm proceeds setting $a_i^{(l)} = a_i'$ with probability $\alpha_{\mathrm{MH}} = \min\{1, \frac{p(a_i')}{p(a_i^{(l-1)})}\}$. If the proposed draw is not accepted, $a_i^{(l)} = a_i^{(l-1)}$.[1]

[1]The average acceptance probability varies with the strength of the prior. For priors of the type considered in the empirical application, the probability is between 0.88–0.98, depending on the row.

Finally, a researcher might prefer formulating a prior distribution for parameters that are nonlinear functions of a_i, say $z_i = H(a_i)$. In the empirical application of this paper, for example, the last row of A is parameterized by $A_{5\bullet} = [0, 0, 0, a_{54}, a_{55}]$ where $a_{54} = (\frac{\rho^\star}{1-\rho^\star}\chi^{-2}\hat{\omega}_3)^{-1/2}$ and $a_{55} = -\chi(\frac{\rho^\star}{1-\rho^\star}\chi^{-2}\hat{\omega}_3)^{-1/2}$ and prior distributions are spelled out for $\rho^\star$ and χ instead of a_{54} and a_{55}. In this case, the uniform prior underlying the proposal distribution $p^\star(a_i|Y, \xi_{a_i})$ implies a nonuniform prior for z_i. Hence, the MH step also needs to correct for the change of variables implicit in the proposal distribution. More formally, let $z_i = H(a_i)$. Let the Jacobian matrix evaluated at z_i be $J(z_i) = \frac{dH^{-1}(x)}{dx}|_{x=z_i}$. Then the density for z_i implied by the proposal distribution (equation (S.13)) is given by $p^\star(H^{-1}(z_i)|Y, \xi_{-z_i}) \times |\det(J(z_i))|$. Noting that the target posterior distribution is given by $p^\star(H^{-1}(z_i)|Y, \xi_{-z_i})p(z_i)$ the MH acceptance probability is then given by

$$\alpha_{\text{MH}} = \min\left\{1, \frac{p(z_i')|\det(J(z_i^{(l-1)})|}{p(z_i^{(l-1)})|\det(J(z_i'))|}\right\}$$

4. The fourth block draws from the conditional distribution of the VAR autoregressive parameters. Let $\mu_t = [\mu_{1t}, \ldots, \mu_{Kt}]'$ and $\Sigma_t = \text{diag}([\sigma_{1t}^2, \ldots, \sigma_{Kt}^2])$ The conditional posterior of α_+ is given by

$$p(\alpha_+|Y, \xi_{-\alpha_+}) \sim \mathcal{N}(\overline{\mu}_A, \overline{V}_A), \tag{S.14}$$

where

$$\overline{V}_{\alpha_+} = \left(V_{\alpha_+}^{-1} + \sum_{t=1}^{T}(x_t \otimes I_K)\big(A'\Sigma_t^{-1}A\big)\big(x_t' \otimes I_K\big)\right)^{-1}, \tag{S.15}$$

$$\overline{\mu}_{\alpha_+} = \overline{V}_{\alpha_+}\left(V_{\alpha_+}^{-1}m_{\alpha_+} + \sum_{t=1}^{T}(x_t \otimes I_K)\big(A'\Sigma_t^{-1}A\big)\tilde{y}_t\right), \tag{S.16}$$

for $\tilde{y}_t = y_t - A^{-1}\mu_t$.

A.2 *Adjustment for the oil market model*

The algorithm outlined in Appendix A.1 is not directly applicable to the oil market model outlined in Section 3. The reason is that $u_t^{i^\star}$, the forecast error of the scaled up oil inventories, is an unobserved latent variable. To get around this problem, I include $u_t^{i^\star}$ into the set of latent variables and infer it from the data within the MCMC algorithm. Specifically, the fourth block is altered as to draw from $p(\alpha_+, u^{i^\star}|Y, \xi_{\{-\alpha_+, -u^{i^\star}\}})$, where $u^{i^\star} = [u_1^{i^\star}, \ldots, u_T^{i^\star}]'$. Specifically, I make use of the possibility to marginalize over $u^{i^\star}$ when sampling α_+. The adjusted fourth block of the MCMC algorithm draws from

$$p\big(\alpha_+, u^{i^\star}|Y, \xi_{\{-\alpha_+, -u^{i^\star}\}}\big) = \underbrace{p(\alpha_+|Y, \xi_{\{-\alpha_+, -u^{i^\star}\}})}_{\text{normal}}\underbrace{p\big(u^{i_t^\star}|\alpha_+, Y, \xi_{\{-\alpha_+, -u^{i^\star}\}}\big)}_{\text{normal}}.$$

In words, first a draw of α_+ is generated from the conditional posterior marginal of $u^{i^\star}$. The second step draws $u^{i^\star}$ conditional on α_+. To derive both steps, note that one may readily marginalize out $u_t^{i^\star}$ to obtain the likelihood function of the observed forecast errors. Conditional on auxiliary mixture parameters in Θ, the model is given as

$$\mathrm{A}\begin{pmatrix} y_t - \mathrm{A}_+ x_t \\ u_t^{i^\star} \end{pmatrix} = \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\tilde{\mu}_t, \tilde{\Sigma}_t).$$

Since the measurement error $\varepsilon_{5t} \sim \mathcal{N}(0, \sigma_5^2)$ is Gaussian, we have that $\tilde{\mu}_t = [\mu_t', 0]'$ and $\tilde{\Sigma}_t = \mathrm{diag}([v_t', \sigma_5^2]')$. Manipulating the equation, the reduced form can be obtained:

$$\begin{pmatrix} y_t \\ u_t^{i^\star} \end{pmatrix} = \begin{pmatrix} \mathrm{A}_+ x_t \\ 0 \end{pmatrix} + \mathrm{A}^{-1}\tilde{\mu}_t + \tilde{\eta}_t, \quad \varepsilon_t \sim \mathcal{N}\big(0, \mathrm{A}^{-1}\tilde{\Sigma}_t \mathrm{A}^{-1'}\big), \tag{S.17}$$

which defines the joint likelihood of Y and $u_t^{i^\star}$. Define J s.t. $u_t = J\tilde{u}_t$. Then, using standard results of multivariate Gaussian densities, the marginal likelihood is simply given:

$$\begin{aligned} &p(Y|\alpha_+, \xi_{\{-\alpha_+, -u^{i^\star}\}}) \\ &\quad \propto |\Omega_t|^{-T/2} \exp\left(\sum_{t=1}^{T} (\tilde{y}_t - A_+ x_t)' \Omega_t^{-1} (\tilde{y}_t - A_+ x_t)\right) \end{aligned} \tag{S.18}$$

for $\tilde{y}_t = y_t - JA^{-1}\tilde{\mu}_t$ and $\Omega_t = J\mathrm{A}^{-1}\tilde{\Sigma}_t \mathrm{A}^{-1'} J'$. Given the likelihood, its straightforward to obtain the conditional posterior $p(\alpha_+|Y, \xi_{\{-\alpha_+, -u^{i^\star}\}}) \sim \mathcal{N}(\overline{\mu}_A, \overline{V}_A)$,

$$\overline{V}_A = \left(V_A^{-1} + \sum_{t=1}^{T} (x_t \otimes I_K)\Omega_t^{-1}\big(x_t' \otimes I_K\big)\right)^{-1}, \tag{S.19}$$

$$\overline{\mu}_A = \overline{V}_A\left(V_A^{-1} m_{\alpha_+} + \sum_{t=1}^{T} (x_t \otimes I_K)\Omega_t^{-1}\tilde{y}_t\right), \tag{S.20}$$

The second step involves drawing from $p(u^{i_t^\star}|\alpha_+, Y, \xi_{\{-\alpha_+, -u^{i^\star}\}})$, which can be obtained using standard results for multivariate normal distributions. Define

$$\mathrm{A}^{-1}\tilde{\Sigma}_t \mathrm{A}^{-1'} = \tilde{\Omega}_t = \begin{pmatrix} \tilde{\Omega}_{t,11} & \tilde{\Omega}_{t,12} \\ \tilde{\Omega}_{t,21} & \tilde{\Omega}_{t,22} \end{pmatrix},$$

and J_2 a $1 \times (K+1)$ vector s.t. $J_2\tilde{u}_t = u_t^{i^\star}$. Then, for $t = 1, \ldots, T$, this conditional is given as

$$\begin{aligned} p\big(u^{i_t^\star}|\alpha_+, Y, \xi_{\{-\alpha_+, -u^{i^\star}\}}\big) &\sim \mathcal{N}\big(\overline{u}_t^{i^\star}, \overline{V}_{u_t^{i^\star}}\big), \\ \overline{u}_t^{i^\star} &= J_2 A^{-1}\tilde{\mu}_t + \tilde{\Omega}_{t,21}\tilde{\Omega}_{t,11}^{-1}\big(\tilde{y}_t - \mathrm{A}_+ x_t - JA^{-1}\tilde{\mu}_t\big), \\ \overline{V}_{u_t^{i^\star}} &= \tilde{\Omega}_{t,22} - \tilde{\Omega}_{t,21}\tilde{\Omega}_{t,11}^{-1}\tilde{\Omega}_{t,12} \end{aligned}$$

A.3 *Adjustment for the oil market model—Gaussian model*

To keep results simple and comparable, the same MCMC algorithm for the Gaussian model is used, with one alteration. Specifically, instead of drawing from the DPMM auxiliary parameters, a simple Gibbs update is used for each shock variance. Given inverse Gamma priors $\sigma_i^2 \sim i\mathcal{G}(a_{\sigma_i}, b_{\sigma_i})$, the Gibbs sample steps are $p(\sigma_i^2|Y, \xi_{-\sigma_i^2}) \sim i\mathcal{G}(\overline{a}_{\sigma_i}, \overline{b}_{\sigma_i})$ where $\overline{a}_{\sigma_i} = a_{\sigma_i} + \frac{T}{2}$ and $\overline{b}_{\sigma_i} = b_{\sigma_i} + \frac{\sum_{j=1}^{T} \varepsilon_{ij}^2}{2}$. The result of the algorithm can proceed as in Sections A.1 and A.2, simply setting $\mu_{it} = 0$ and $\sigma_{it}^2 = \sigma_i^2$ for $i = 1, \ldots, K$ and $t = 1, \ldots, T$.

Appendix B: Convergence properties MCMC

To study the convergence properties of the MCMC, I simulate artificial data of size $T = 500$ from the following stylized bivariate model of supply and demand:

$$\begin{aligned} q_t &= \alpha_{qp} p_t + \sigma_1 \varepsilon_{1t}, \\ q_t &= \beta_{qp} p_t + \sigma_2 \varepsilon_{2t}, \end{aligned}$$

where $\varepsilon_t \sim (0, I_2)$. Regarding the error term, I set $\varepsilon_t^i = \sqrt{\frac{\nu}{\nu-2}}\tilde{\varepsilon}_t^i$, $i = 1, 2$ for $\tilde{\varepsilon}_t^i \sim t_\eta$ where t_η is the student-t distribution with η degrees-of-freedom. The values of the parameters are set to $\alpha_{qp} = 0.05$, $\beta_{qp} = -0.35$, $\sigma_1 = 1$, and $\sigma_2 = 0.5$. When estimating the model, the following prior is used for A: $p(\alpha_{qp}) \sim t_{0,\infty}(0.1, 0.2, 3)$ and $p(\beta_{qp}) \sim t_{0,\infty}(-0.1, 0.2, 3)$, that is, truncated t-distributions with modes at 0.1 and -0.1, scale of 0.2 and 3 degrees-of-freedom. In this scenario, generating 1000 random draws from the MCMC algorithm takes about 3 seconds using a standard $i5$ laptop processor.[2] To contrast the results to those of a Gaussian model, the model is also estimated using the methodology of Baumeister and Hamilton (2015).

B.1 *Strong identification via non-Gaussianity*

I start with simulating data using $\eta = 3$ degrees-of-freedom, which corresponds to strong identification from non-Gaussianity. First, Figure S.1 shows the simulated structural shocks (top panel) along with estimated 90% posterior credibility sets for the corresponding predictive density obtained in the non-Gaussian model. The latter, highlighted by red dashed lines, demonstrate that the DPMM–SVAR can well capture the strong non-Gaussian shape in the data. Particularly, the second shock has strong outliers leading to very heavy tails.

Second, Figure S.2 shows a Markov chain of length 100000 for α_{qp} and β_{qp} obtained by saving every 10th draw. For both models, Gaussian and Non-Gaussian, visual inspection indicates that the MCMC seems to have converged reasonably well. As a summary statistic of the underlying autocorrelation, Gewekes Relative Numerical Efficiency (RNE) statistics are printed into each subplots title. As described in Geweke (1992), the RNE carries the interpretation of the ratio of number of replications required to achieve the

[2] For the computations in this paper, a Intel(R) Core(TM) i5-6300U CPU with 2.40 GHz was used.

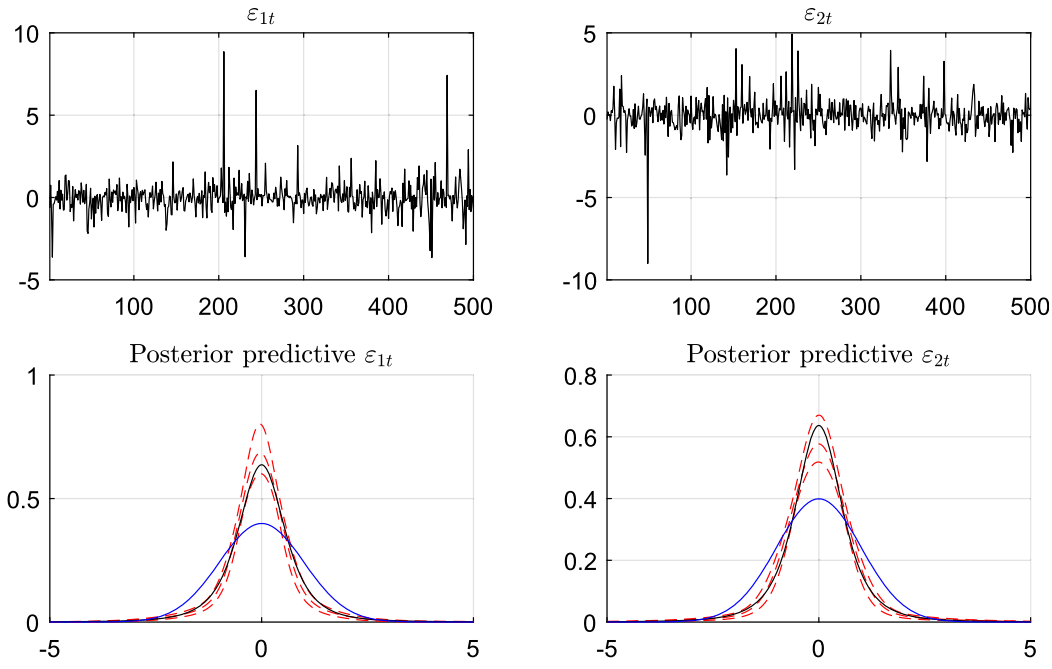

FIGURE S.1. Simulated structural shocks (top panel) and estimated posterior predictive densities under the non-Gaussian model. Red dashed lines indicate 90% posterior credibility sets, the black line that of a unit variance standardized t_3 distribution, and the blue line gives the standard normal density.

same efficiency than drawing i.i.d. from the posterior. The RNE values documented for the algorithm suggest a fairly high autocorrelation in the draws even after the thinning of the Markov chain by factor of 10. This suggests that similar to the algorithm of Baumeister and Hamilton (2015), one should consider a relatively large Markov chain of 100000 to obtain comparably precise results of at least 1000 i.i.d. draws.

Finally, Figure S.3 compares the priors used to the posterior distribution obtained in the Gaussian (top panel) and non-Gaussian model (bottom panel). In the Gaussian model, the data seems to be totally uninformative about the value of α_{qp}, while the value of β_{qp} is estimated fairly precisely. As expected, once non-Gaussianity is taken into account, posterior mass shifts toward the true value of α_{qp}, and further narrows down the value of β_{qp}.

B.2 *Weak identification via non-Gaussianity*

In the second case, I use $\eta = 10$ degrees-of-freedom, which should yield considerably less identifying information from non-Gaussianity. As evident in Figure S.4, the simulated shocks are closer to normality and estimated 90% posterior credibility sets of the posterior predictive distribution includes the Gaussian bell curve. Regarding MCMC efficiency, visual inspection of the Markov chains printed in Figure S.5 suggests no apparent problem with the MCMC. However, the RNE values deteriorates somewhat, which

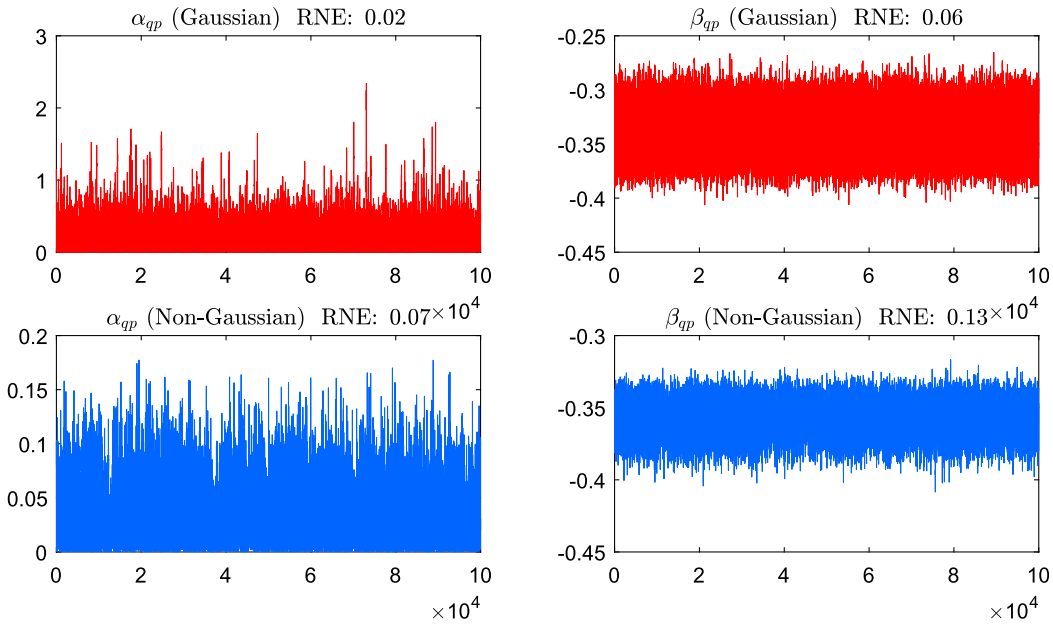

FIGURE S.2. Markov Chain Monte Carlo output of length 100'000. Top panel: Gaussian model with MCMC as in Baumeister and Hamilton (2015). Bottom panel: MCMC of non-Gaussian model as described in Appendix A.1.

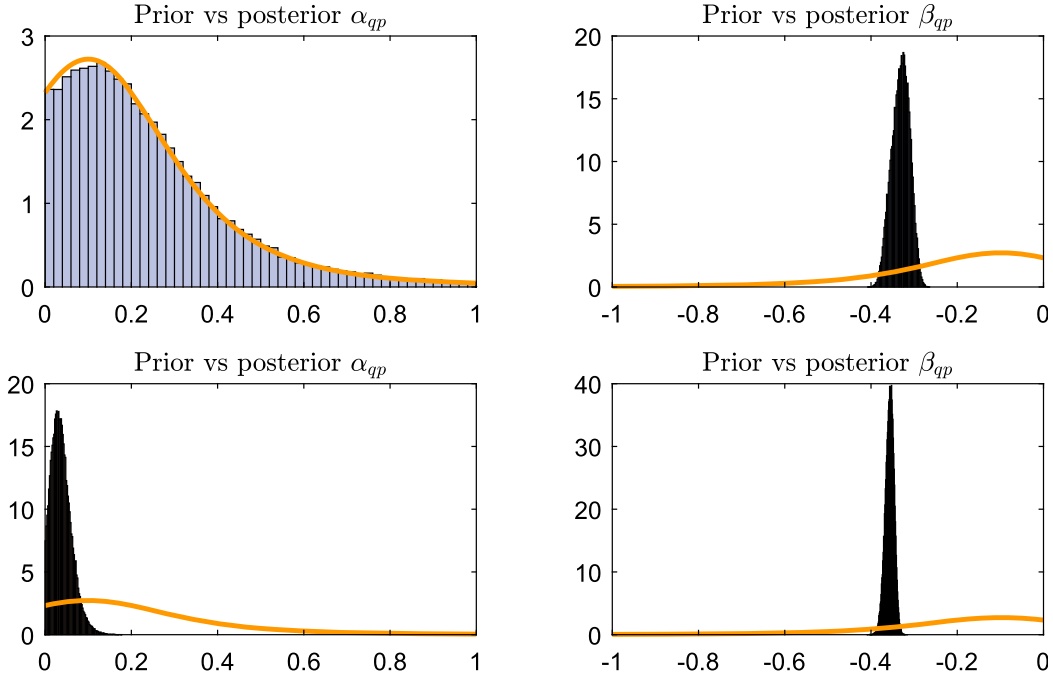

FIGURE S.3. Prior (orange line) and posterior density of the two structural parameters α_{qp} and β_{qp}. Top panel: Gaussian model. Bottom panel: non-Gaussian model.

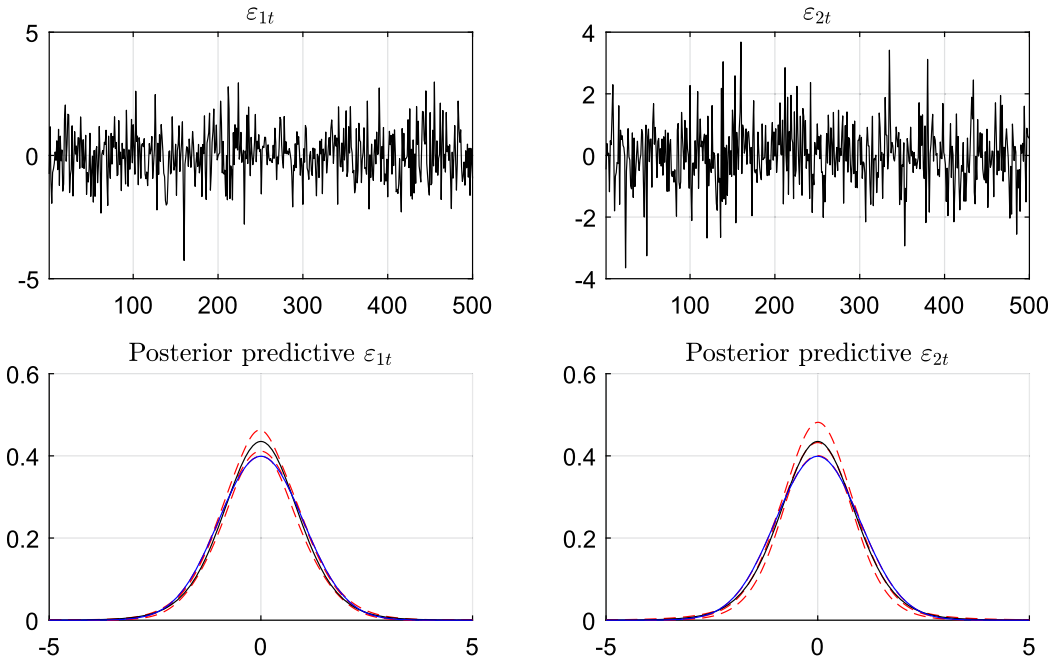

FIGURE S.4. Simulated structural shocks (top panel) and estimated posterior predictive densities under the non-Gaussian model. Red dashed lines indicate 90% posterior credibility sets, the black line that of a unit variance standardized t_{10} distribution, and the blue line gives the standard normal density.

is to be expected for Gibbs sampler type MCMC algorithms under weak identification. Finally, Figure S.6 shows that under weaker identification by non-Gaussianity, the posterior is naturally less informative about the structural parameters. However, given a more concentrated posterior of α_{qp} near zero, some additional information is contained in the likelihood if compared to the Gaussian model.

B.3 *Empirical application*

As a last exercise, Figure S.7 provides a plot of the Markov chains corresponding to each element of A in the empirical application (Section 3). It is fair to say that one might expect a slightly slower convergence given the additional complexity that comes with inferring the latent inventory series. Visual inspection suggest good convergence of the algorithm, however. Still, large RNE suggests a fairly high autocorrelation in the draws justifying the use of very long Markov chain.

APPENDIX C: ILLUSTRATION MARGINAL LIKELIHOOD

To illustrate the use and reliability of the marginal likelihood estimator, I use simulated data of size $T = 500$ from the bivariate static model outlined in the previous section of

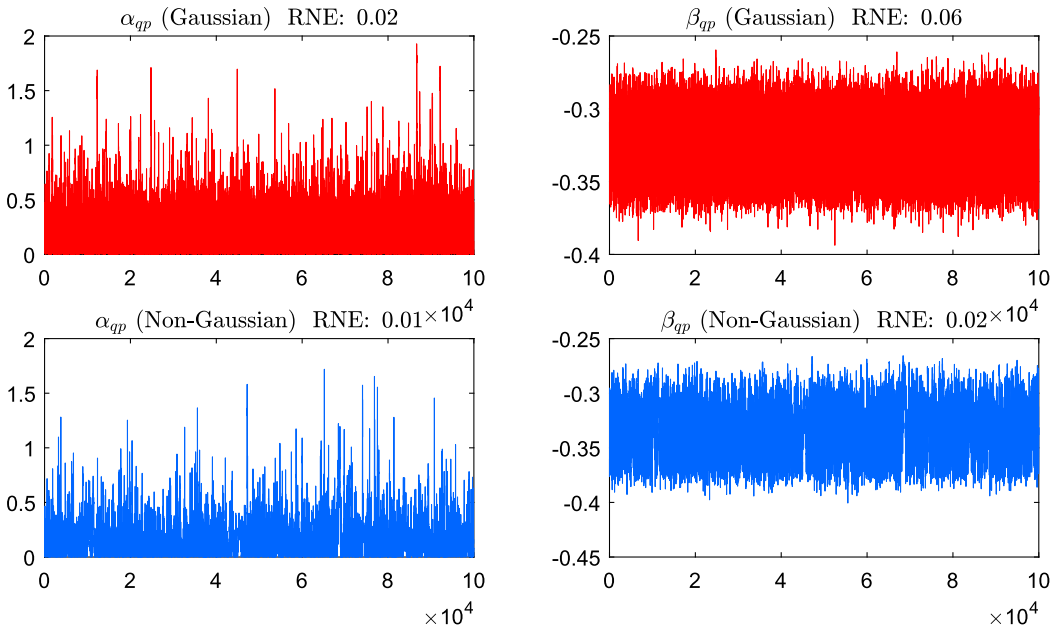

FIGURE S.5. Markov Chain Monte Carlo output of length 100'000. Top panel: Gaussian model with MCMC as in Baumeister and Hamilton (2015). Bottom panel: MCMC of non-Gaussian model as described in Appendix A.1.

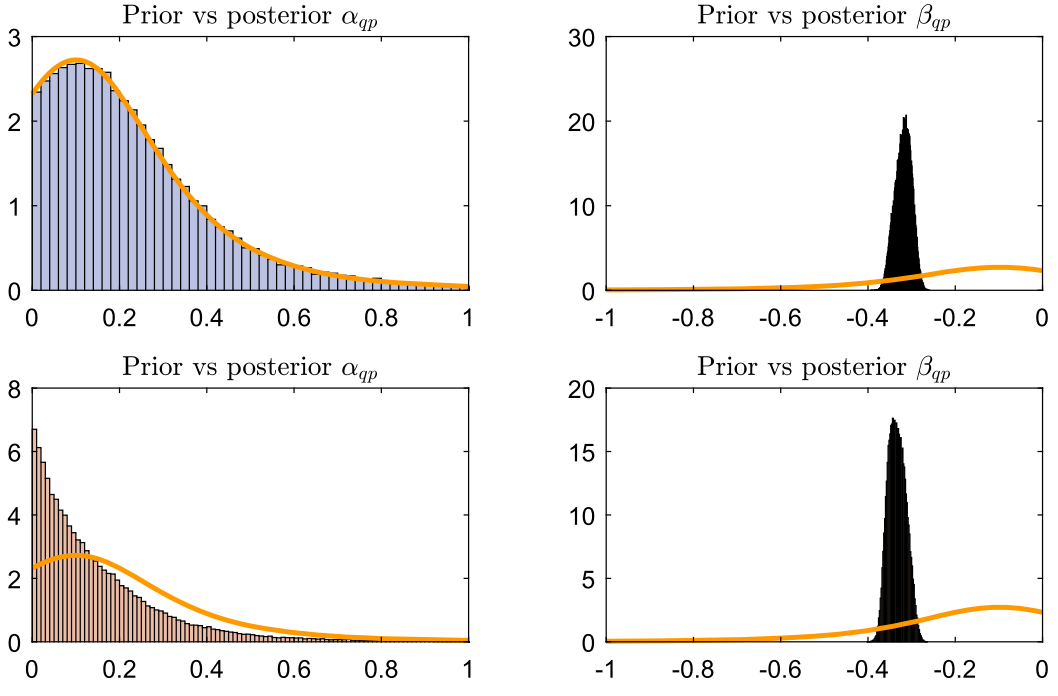

FIGURE S.6. Prior (orange line) and posterior density of the two structural parameters α_{qp} and β_{qp}. Top panel: Gaussian model. Bottom panel: non-Gaussian model.

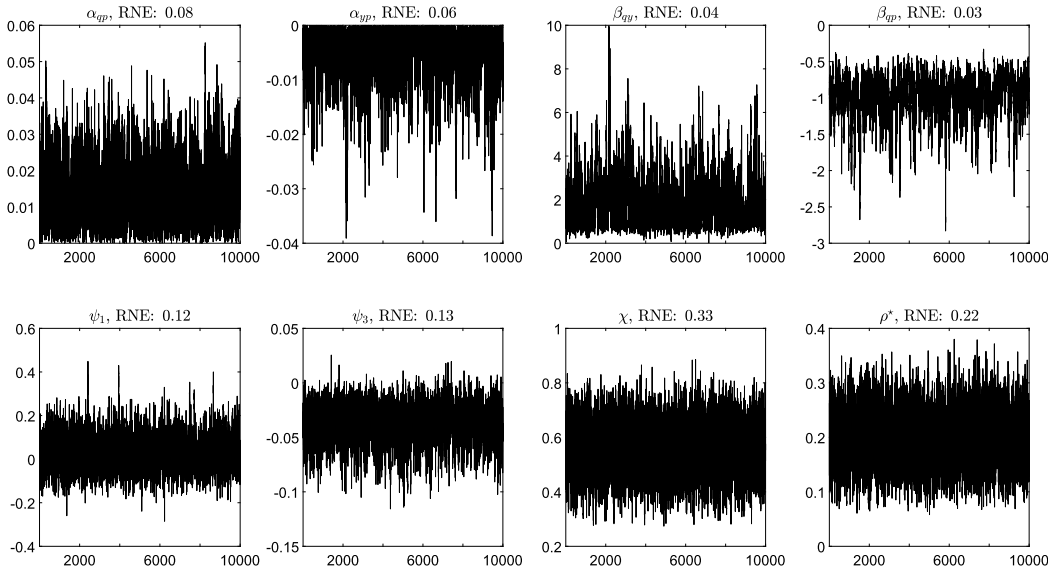

FIGURE S.7. Markov Chain Monte Carlo output of each element of A obtained under the non–Gaussian model of Section 3.

the Appendix:

$$q_t = \alpha_{qp} p_t + \sigma_1 \varepsilon_{1t},$$
$$q_t = \beta_{qp} p_t + \sigma_2 \varepsilon_{2t},$$

where $\varepsilon_t \sim (0, I_2)$, $\alpha_{qp} = 0.25$, $\beta_{qp} = -0.35$, $\sigma_1 = 1$, and $\sigma_2 = 0.5$. Regarding the error term, say that $\varepsilon_t^i = \sqrt{\frac{\nu}{\nu-2}}\tilde{\varepsilon}_t^i$, $i = 1, 2$ for $\tilde{\varepsilon}_t^i \sim t_\eta$ where t_η is the student-t distribution with $\nu = 3$ degrees-of-freedom.

Assume the goal is to test the wrong hypothesis that the supply is price inelastic, that is, $\alpha_{qp} = 0$. Given that the shocks are clearly non-Gaussian, it is possible to test the hypothesis using Bayes factors outlined in Section 2.5. To this end, let model M_1 be un unrestricted DPMM–SVAR with weakly informative priors $p(\alpha_{qp}) \sim t_{0,\infty}(0.1, 0.2, 3)$ and $p(\beta_{qp}) \sim t_{0,\infty}(-0.1, 0.2, 3)$. On the other hand, for the restricted model M_0 it holds that $\alpha_{qp} = 0$.[3]

For the cross-entropy method outlined in Section 2.5, I set $G = 50$ and $M = 5000$, which corresponds to the number of replications used to evaluate the likelihood (G) and the marginal likelihood (M), respectively. Estimated log Marginal Likelihoods are then given by $\ln \widehat{p(Y|M_1)} = -1406.09$ for the restricted model and $\ln \widehat{p(Y|M_0)} = -1412.55$ for the unrestricted model. Standard errors for these estimates can be readily obtained by the batch means method. Splitting the importance sampling simulation output into 10

[3] Furthermore, for both shocks $i = 1, 2$, set α_i be such that $E(k|T, \alpha_i) = 3$. With respect to the Base distribution, set uninformative values $m_i = 0$, $\tau_i = 5$, $s_i = 1/2$, and $S_i = 4$. Furthermore, although the true model is static, the number of lags is set to $p = 1$ with weakly informative prior $p(\alpha_+) \sim \mathcal{N}(0, 100 \times I_4)$.

TABLE S.1. Categories of interpretation according to Kass and Raftery (1995).

$2\ln(BF_{10})$	B_{10}	Evidence against M_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very strong

equally-sized buckets yields a standard error of 0.22 and 0.44, respectively, suggesting fairly accurate estimates.

The (log) marginal likelihood of the unrestricted model M_1 is clearly higher than that of the restricted model M_0. This should be no surprise given that the true supply curve is not inelastic. In order to interpret the magnitudes, it is common to look at twice the natural logarithm of the Bayes factor $B_{10} = p(Y|M_1)/p(Y|M_0)$, which operates on the same scale than the more familiar likelihood ratio test statistic. For the simulated data above, this yields a value of $2\ln(BF_{10}) = 12.93$. One then can make use of the popular reference point categories provided in Kass and Raftery (1995) to interpret the exact magnitude. As suggested by Table S.1, the evidence against the null hypothesis is very strong.

APPENDIX D: OIL MARKET MODEL: SUPPLEMENTARY FIGURES

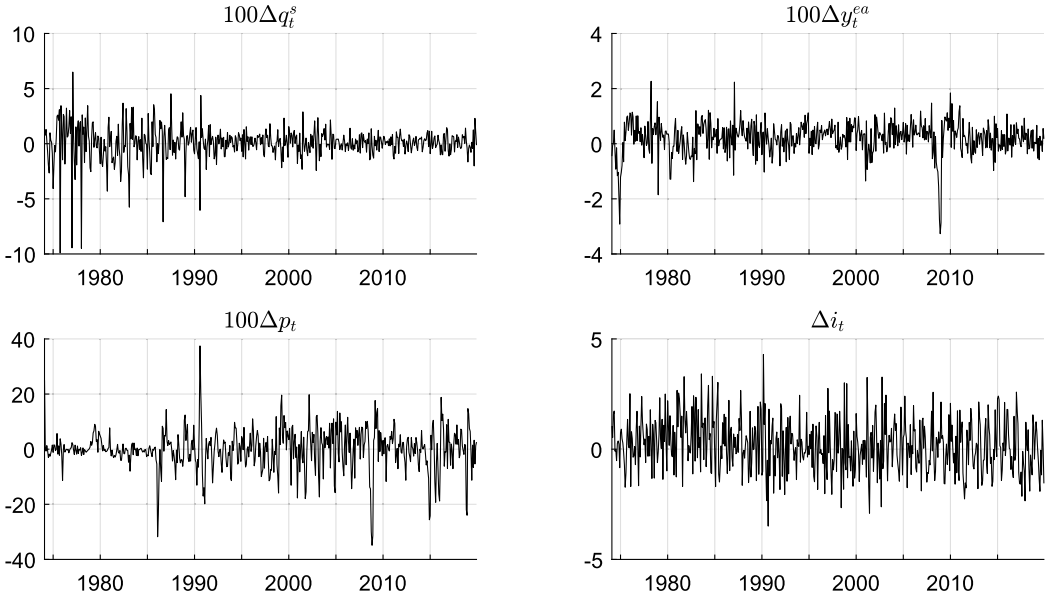

FIGURE S.8. Oil market data set.

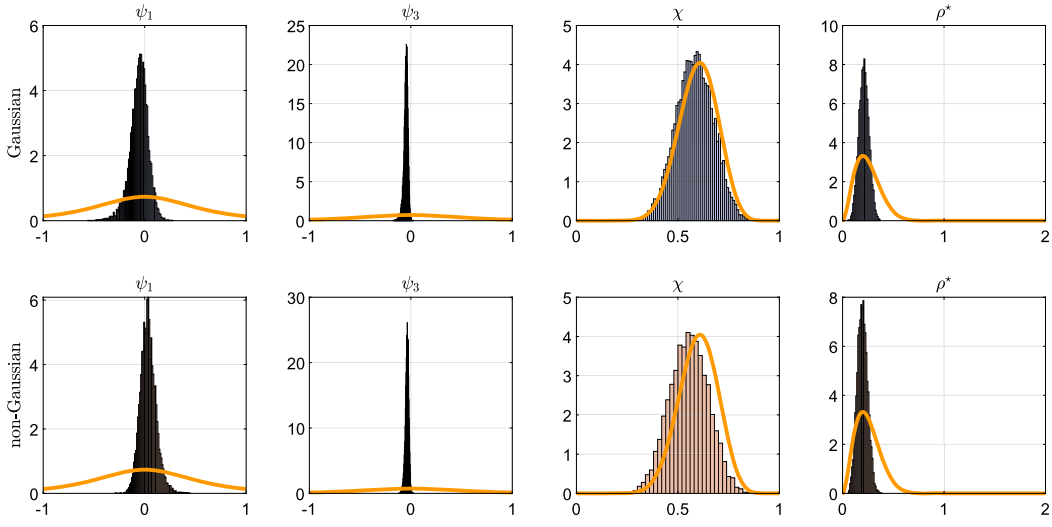

FIGURE S.9. Prior (orange line) and posterior density of the remaining structural parameters. Top panel: Gaussian model. Bottom panel: non-Gaussian model.

APPENDIX E: OIL MARKET MODEL: ON THE RELATIONSHIP BETWEEN ELASTICITIES AND VARIANCE DECOMPOSITION

In this part of the Appendix, I demonstrate the link between restrictions on α_{qp} and resulting estimates of the demand elasticity β_{qp}, and the variance decomposition of the real oil price. Using the (Gaussian) baseline model, I fix α_{qp} at values between 0 and 0.15, and estimate the remaining parameters by maximizing the posterior. Note that the resulting maximum-a-posteriori (MAP) estimates will reflect a combination of prior and covariance structure in the data, but do not further rely on independence and non-Gaussianity. For a similar exercise, see also Caldara, Cavallo, and Iacoviello (2019).

Figure S.10 displays MAP estimates of the demand elasticity (left) and the forecast error variance decomposition of the real oil price (right), both obtained after fixing α_{qp} (x-axis). First, note that the smaller the short-run elasticity of supply, the larger are estimates of the demand elasticity β_{qp} (in absolute terms). This is well in line with the empirical results of Section 3. Here, the posterior of the non-Gaussian model (red) concentrates at very small values of α_{qp}, and relatively high values for $|\beta_{qp}|$. On the other hand, the model identified as in Baumeister and Hamilton (2019) (BH19) suggests relatively high estimates of α_{qp} and a less steep demand curve.

The right panel shows the implication of varying α_{qp} for the variance decomposition of the real oil price, calculated at the $h = 16$ months forecast horizon. Low values for the supply elasticity come with a very small contribution of supply shocks ε_t^s to the variance, while demand shocks ε_t^{cd} are very important. On the other hand, larger values for the supply elasticity imply a substantial role for supply shocks in driving oil price, and hence, less importance of demand shocks.

Figure S.11 repeats the exercise excluding the earlier years of the sample, covering only data from 1985M1–2019M12 (robustness exercise R2). The results suggest that the

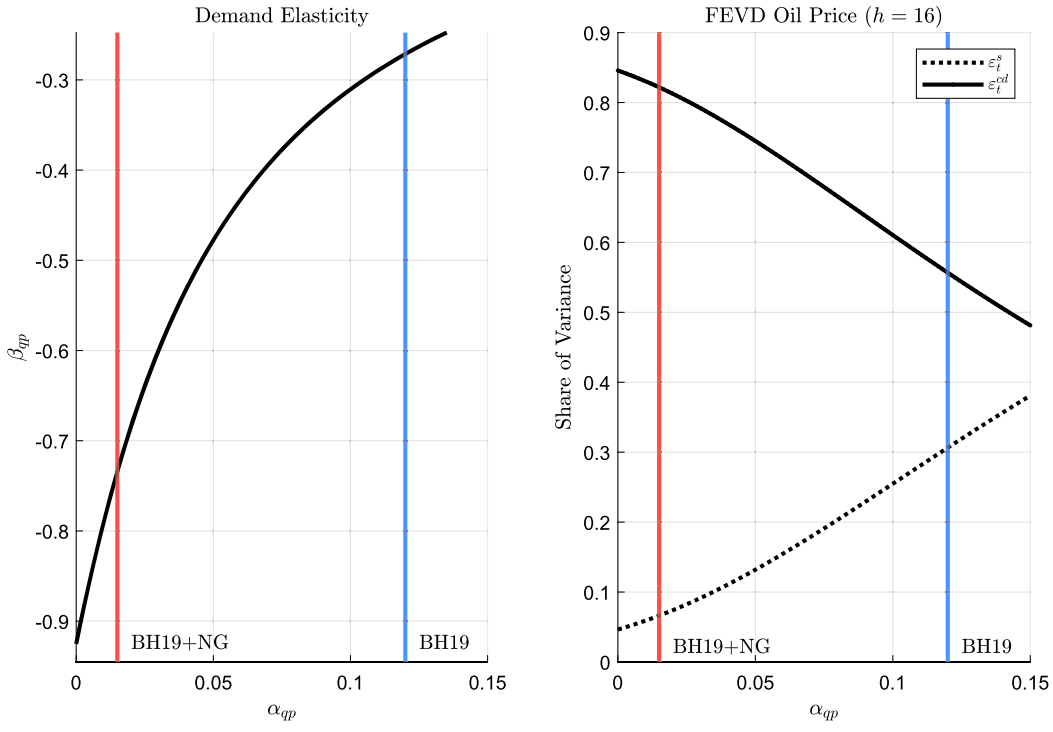

FIGURE S.10. Maximum a posteriori (MAP) estimates obtained under the Gaussian model when the short-run supply elasticity α_{qp} is fixed at the values shown on the x-axis. Left panel: MAP estimates obtained for β_{qp}. Right panel: MAP estimates for the contributions of supply and consumption demand shocks to the variance of the real oil price (at $h = 16$ months forecast horizon).

trade-off between a low supply and high demand elasticity is less pronounced. As reported in Table 4, the non-Gaussian model arrives at posterior median estimates of around -0.3 for β_{qp} while α_{qp} is still very low (0.03). On the other hand, the implications of varying α_{qp} for the variance decomposition of the real oil price remains the same. Models that estimate a inelastic supply curve, such as the non-Gaussian SVAR, estimate a minor role of supply shocks for fluctuations in the oil price.

APPENDIX F: OIL MARKET MODEL: ROBUSTNESS TO ERROR SPECIFICATION

In the following, two more robustness exercises are conducted to assess the sensitivity of the results to the error specifications used to exploit the combined (non-Gaussian) identification strategy. The first, labeled as R3, uses parametric student-t distributions for the shock marginals instead of nonparametric DPMMs.[4] Here, the goal is to to understand if there are any practical gains from using the more involved DPMM machin-

[4]Specifically, I assume $\varepsilon_{it} = \sigma_i \tilde{\varepsilon}_{it}$, $\tilde{\varepsilon}_{it} \sim t_{\nu_i}$ for $i = 1, \ldots, 4$. The scales are given the same priors than in the Gaussian model (see Table 1). The degrees-of-freedom parameters are given a uniform prior between 2 and 100. Posterior sampling of η is implemented via an independence-chain MH algorithm; see, for example, Chan (2020).

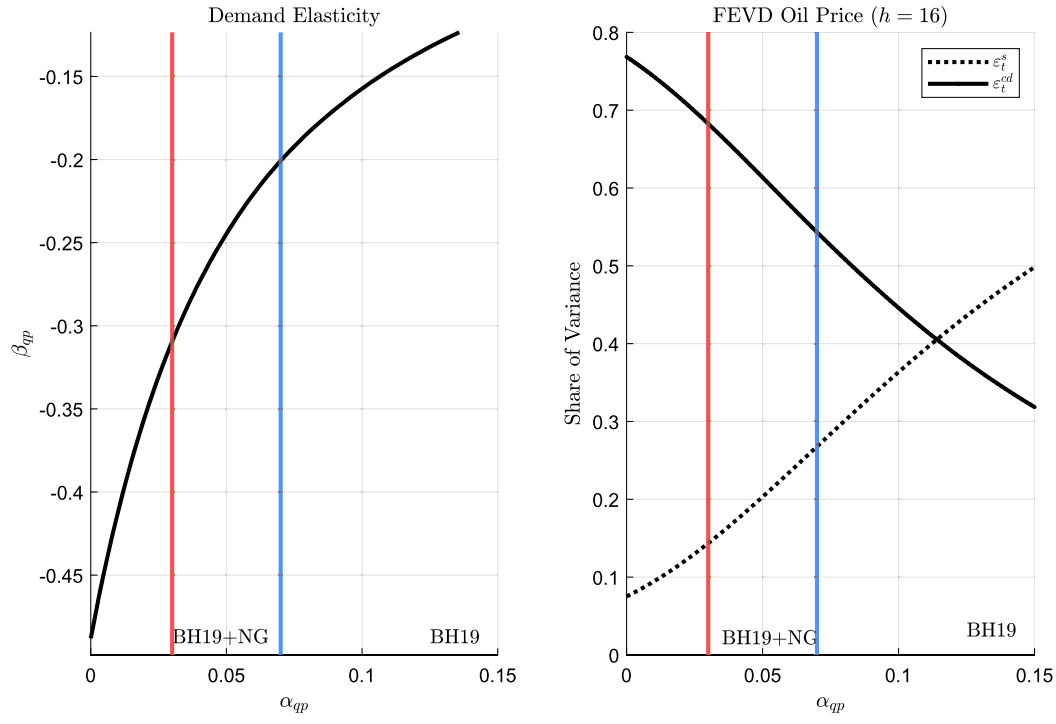

FIGURE S.11. Maximum a posteriori (MAP) estimates obtained under the Gaussian model when the short-run supply elasticity α_{qp} is fixed at the values shown on the x-axis. Here, the estimation sample is shorter and covers 1985M1–2019M12 (robustness exercise R2). Left panel: corresponding estimates obtained for β_{qp}. Right panel: estimated contributions of supply and consumption demand shocks to the variance of the real oil price (at $h = 16$ months forecast horizon).

ery in the empirical analysis. The second robustness check assesses the sensitivity to α_i, setting it to more conservative values, which favor one mixture component, and hence, Gaussian shocks. I set α_i such that $\mathrm{E}[k_i|T, \alpha_i] = 1$ for $i = 1, \ldots, 4$, which also implies a very low a priori standard deviation for the number of mixture components, given by $\mathrm{Var}[k_i|T, \alpha]^{1/2} = 0.02$. In this model, a shock must display strong non-Gaussianity to overrule the prior and become informative about the underlying structural parameters. Corresponding results will be labeled as R4.

First, I provide details on the estimated predictive distributions obtained under R3 and R4. Table S.2 displays posterior quantiles for η_i, the underlying degrees-of-freedom parameter of the student-t marginals (R3). Similar to the baseline results, strong non-Gaussianity is documented for the supply and economic activity shocks. For the consumption demand shock, the predictive seems slightly less heavy tailed, while there is little evidence for non-Gaussanity in the inventory demand shock.

Figure S.12 shows the posterior predictive distributions alongside 90% confidence sets for each shock obtained when the DPPM is set up such that it strongly favors Gaussian marginals a priori (R4). The results suggest that the data still favors non-Gaussian

TABLE S.2. Posterior distribution η_i (R3).

	5%	50%	95%
η_1	2.1	2.6	3.3
η_2	3.9	5.7	9.4
η_3	3.6	6.0	13.8
η_4	11.5	54.0	95.0

marginals for the first two shocks (ε_t^s and ε_t^{ea}). The posterior predictive distribution of the consumption and inventory demand shock coincide with that of a Gaussian.

Proceeding with structural analysis, Table S.3 revisits the posterior distribution of α_{qp}, β_{qp}, and the estimated importance of supply shocks for real oil price variation. Comparing the baseline DPMM results with that of R3 and R4, there is little difference in the posterior of the supply elasticity α_{qp}, concentrating most of the mass near zero. With respect to β_{qp}, the model using student-t errors implies a larger median estimate (-1.22) in absolute terms, than obtained in the baseline results (-0.94). Also, posterior uncertainty measured by the distance between the 5% and 95% quantiles is somewhat larger. On the other hand, for model R4, the posterior median elasticity is smaller (-0.79) compared to the baseline. This is to be expected, as it effectively shrinks the posterior

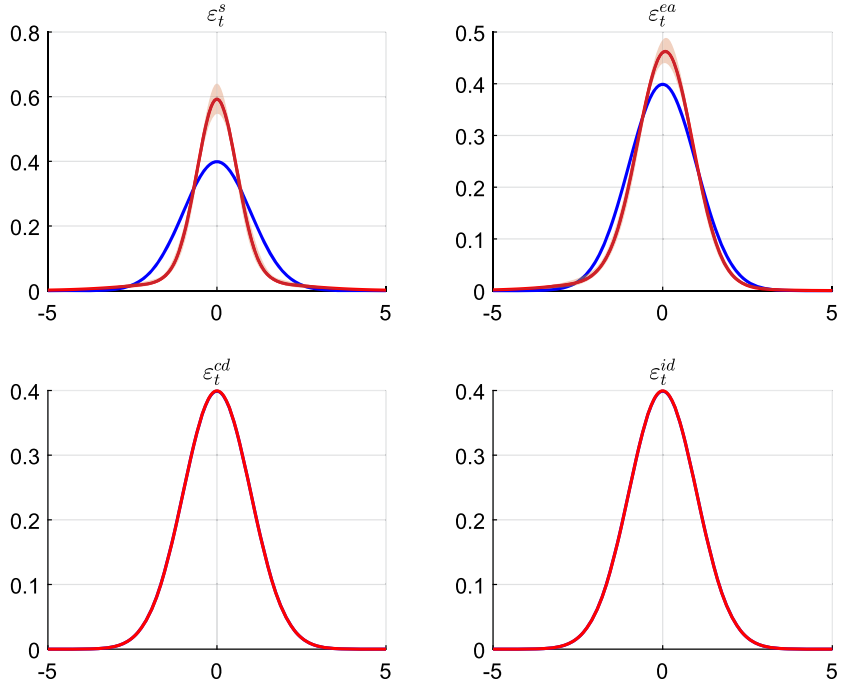

FIGURE S.12. Posterior predictive densities (90% credible interval) of standardized structural shocks $\tilde{\varepsilon}_{i,T+1} = \sigma_i^{-\frac{1}{2}}(\varepsilon_{i,T+1} - \mu_i)$ obtained in R4.

TABLE S.3. Further robustness analysis for the main empirical findings.

	5%	50%	95%
	Panel A: Posterior α_{qp}		
Baseline	0.002	0.013	0.027
R3	0.001	0.010	0.023
R4	0.003	0.015	0.030
	Panel A: Posterior β_{qp}		
Baseline	−1.58	−0.94	−0.63
R3	−1.88	−1.22	−0.75
R4	−1.22	−0.79	−0.55
	$h = 4$		$h = 16$
	Panel C: Contribution of ε_t^s to the FEVD of the real price of oil		
Baseline	0.05 (0.02, 0.10)		0.06 (0.03, 0.11)
R3	0.06 (0.02, 0.19)		0.07 (0.03, 0.22)
R4	0.07 (0.03, 0.13)		0.08 (0.04, 0.13)

Note: For robustness check R3, the non-Gaussian model is estimated with parametric student-t errors instead of nonparametric DPMMs. For robustness check R4, the non-Gaussian model is estimated with $\alpha_i = 1.0563\text{E}-04$ such that $\text{E}[k_i|\alpha_i, T] = 1$ and Variance $\text{Var}[k_i|T, \alpha]^{1/2} = 0.02$, placing a strong prior weight on Gaussian marginals.

toward that obtained under the Gaussian (BH19) model, which peaks near −0.3 (see Figure 5).

With respect to the variance decomposition of the real oil price, the model using student-t distribution instead of the DPMM (R3) points toward a similar importance of supply shocks than the baseline estimates. However, once more the uncertainty is considerably larger. 95% quantiles of 0.19 ($h = 4$) and 0.22 ($h = 4$) are twice as large than obtained by the baseline model. Hence, although the proposed SVAR–DPMM is nonparametric, there seem to be efficiency gains over the parametric student-t alternative.[5] Finally, the model strongly favoring Gaussian shocks a priori (R4) yields very similar results to the baseline results. This suggests that the variance decomposition robust results are robust with respect to the choice of α_i.

REFERENCES

Baumeister, Christiane and James D. Hamilton (2015), "Sign restrictions, structural vector autoregressions, and useful prior information." *Econometrica*, 83 (5), 1963–1999. [8, 9, 10, 12]

[5] More generally, FEVD and IRFs are estimated to be more uncertain in the student-t model. This holds particularly for the IRF of the economic activity shock, as well as the variance contributions of the supply shock. The complete results are available upon request.

Baumeister, Christiane and James D. Hamilton (2019), "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks." *American Economic Review*, 109 (5), 1873–1910. [15]

Caldara, Dario, Michele Cavallo, and Matteo Iacoviello (2019), "Oil price elasticities and oil price fluctuations." *Journal of Monetary Economics*, 103, 1–20. [15]

Chan, Joshua, Gary Koop, and Xuewen Yu (2023), "Large Order-Invariant Bayesian VARs with Stochastic Volatility." *Journal of Business & Economic Statistics*. [4]

Chan, Joshua C. C. (2020), "Large Bayesian vars: A flexible Kronecker error covariance structure." *Journal of Business & Economic Statistics*, 38 (1), 68–79. [16]

Escobar, Michael D. and Mike West (1995), "Bayesian density estimation and inference using mixtures." *Journal of the American Statistical Association*, 90 (430), 577–588. [3]

Geweke, John (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." In *Bayesian Statistics*, Vol. 4 (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), 169–193. [8]

Kass, Robert E. and Adrian E. Raftery (1995), "Bayes factors." *Journal of the american statistical association*, 90 (430), 773–795. [14]

Neal, Radford M. (2000), "Markov chain sampling methods for Dirichlet process mixture models." *Journal of computational and graphical statistics*, 9 (2), 249–265. [2]

Villani, Mattias (2009), "Steady-state priors for vector autoregressions." *Journal of Applied Econometrics*, 24 (4), 630–650. [4, 5]

Waggoner, Daniel F. and Tao Zha (2003), "A Gibbs sampler for structural vector autoregressions." *Journal of Economic Dynamics and Control*, 28 (2), 349–366. [4]

Co-editor Tao Zha handled this manuscript.

Manuscript received 8 February, 2022; final version accepted 9 June, 2023; available online 27 June, 2023.