



BANK OF ENGLAND

Staff Working Paper No. 957

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

Robin Braun

December 2021

Staff Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate. Any views expressed are solely those of the author(s) and so cannot be taken to represent those of the Bank of England or to state Bank of England policy. This paper should therefore not be reported as representing the views of the Bank of England or members of the Monetary Policy Committee, Financial Policy Committee or Prudential Regulation Committee.



BANK OF ENGLAND

Staff Working Paper No. 957

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

Robin Braun⁽¹⁾

Abstract

When quantifying the importance of supply and demand for oil price fluctuations, a wide range of estimates have been reported. Models identified via a sharp upper bound on the short-run price elasticity of supply, find supply shocks to be minor drivers. In turn, when replacing the upper bound with a fairly uninformative prior, supply shocks turn out to be quite important. In this paper, I revisit the evidence with a model identified by a combination of weakly informative priors and non-Gaussianity. For this purpose, a structural vector autoregressive (SVAR) model is developed where the distributions of the structural shocks are modelled non-parametrically. The empirical findings indicate that once non-Gaussianity is incorporated into the model, posterior mass of the short-run oil supply elasticity shifts towards zero and oil supply shocks become minor drivers of oil prices. In terms of contributions to the forecast error variance of oil prices, the model arrives at median estimates of just 6% over a 16-month horizon.

Key words: Oil market, SVAR, identification by non-Gaussianity, non-parametric Bayesian methods.

JEL classification: Q43, C32.

(1) Bank of England. Email: robin.braun@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. I thank Knut Aastveit, Christiane Baumeister, Dominik Bertsche, Ralf Brüggemann, Jamie Cross, Gabriele Fiorentini, Luca Gambetti, Lutz Kilian, Michele Piffer, Mikkel Plagborg Møller and Giovanni Ricco for many great comments and help. I'm also grateful to the many seminar participants at the University of Konstanz, the Bank of England, Queen Mary University, the CFE 2019, the ESWC 2020, the IAAE 2021, the Barcelona GSE Summer Forum 2021, the NASMES 2021 and the CEF 2021 for useful comments on this paper. Part of this work was conducted with financial support by the German Science Foundation (grant number: BR 2941/3-1) and the Graduate School of Decision Sciences at the University of Konstanz, which is gratefully acknowledged.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

© Bank of England 2021
ISSN 1749-9135 (on-line)

1 Introduction

Since Kilian (2009), an increasing number of papers have contributed to our understanding of the distinct role of supply and demand shocks in driving oil price fluctuations. However, when quantifying their relative importance, a wide range of estimates have been reported. One set of papers, including Kilian (2009, 2020a), Kilian & Murphy (2012, 2014), Juvenal & Petrella (2015), Antolín-Díaz & Rubio-Ramírez (2018), Zhou (2020) and Cross, Nguyen & Tran (2020) report that oil prices are mainly driven by demand and hence, that supply shocks are not important. Their estimates typically attribute less than 10% of the long term variability in oil prices to supply. On the other hand, recent papers of Baumeister & Hamilton (2019) and Caldara, Cavallo & Iacoviello (2019) point towards a substantially larger role of supply, estimating variance contributions of up to 37%. As discussed in Herrera & Rangaraju (2020), most of the disagreement can be attributed to differences in the identification strategy.¹ In particular, papers that impose very small upper bounds on the short-run price elasticity of supply find negligible effects of supply shocks. On the other hand, when the bound is replaced with a fairly diffuse prior that allows for larger elasticities, supply shocks can become quite important drivers of oil prices.

In this paper, I revisit the evidence based on an alternative identification strategy. In particular, I combine the prior distributions of Baumeister & Hamilton (2019) with identifying information from non-Gaussianity. The latter is based on the assumption that structural shocks are mutually independent and display some degree of non-Gaussianity. As documented in this paper, large deviations from Gaussianity characterize many oil market variables. My findings indicate that once non-Gaussianity is incorporated in the model, the posterior distribution of the short run oil-supply concentrates near zero and oil supply shocks are found to be minor drivers of oil prices.

To build up intuition on the identification strategy used in this paper, consider a stylized bivariate model for supply and demand:

$$\begin{array}{l} \text{supply:} \\ \text{demand:} \end{array} \quad \begin{array}{l} q_t = \alpha p_t + \sigma_1 \varepsilon_t^s \\ q_t = \beta p_t + \sigma_2 \varepsilon_t^d \end{array} \quad \begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^d \end{pmatrix} \sim (0, I_2),$$

where q_t and p_t are changes in the (log) quantity and (log) prices, α and β are the price elasticities of supply and demand, and $\sigma_{1/2}$ the standard deviations of the respective structural supply (ε_t^s) and demand shocks (ε_t^d). The model is not identified from the second moment of the bivariate dataset, as there are four structural parameters but only three reduced form covariance parameters. To achieve identification, the oil market literature has proceeded by imposing identifying restrictions which reflect their prior on the sign and magnitudes of the structural parameters. First and uncontroversially, the supply curve is assumed to be upward sloping ($\alpha > 0$), while the demand curve is downward sloping ($\beta < 0$). Second, since these

¹As noted in Aastveit, Bjørnland & Cross (2021), the disagreement is much less pronounced once a shorter sample is used for estimation, excluding the large oil price shocks of the 70s.

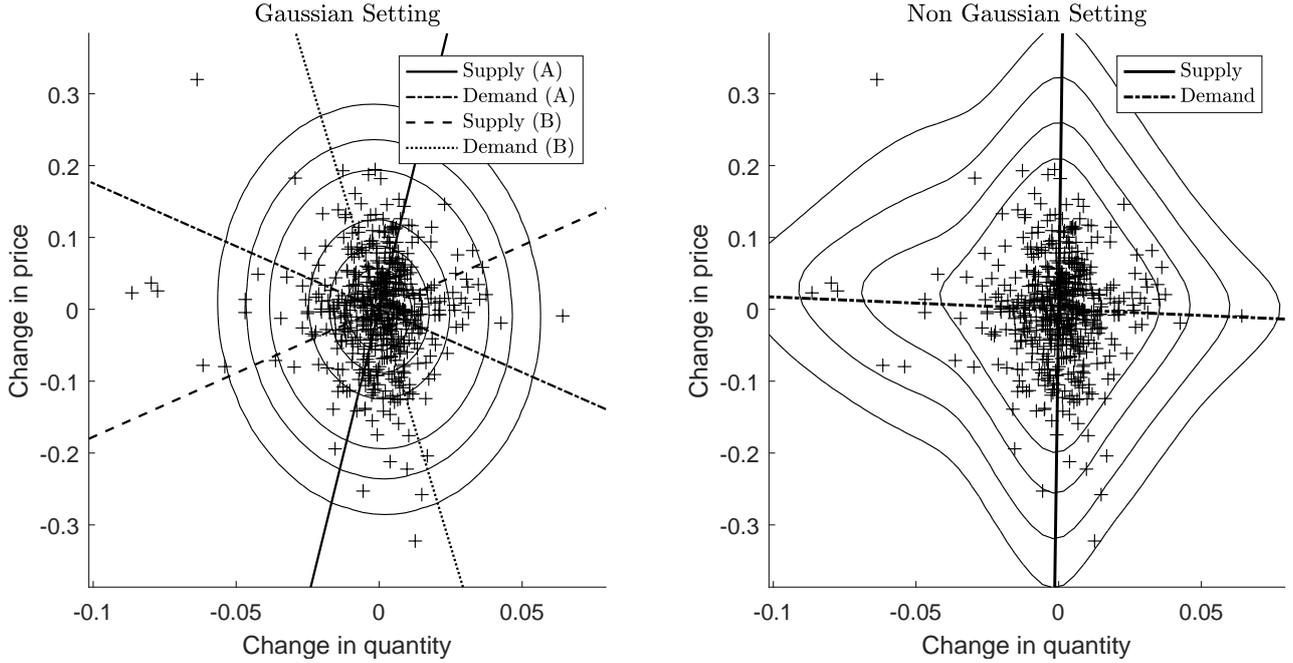


Figure 1: Identifying oil demand and supply curves by non-Gaussianity

sign-restrictions have been found to be fairly uninformative, restrictions on the magnitude of elasticities are included. For example, Kilian & Murphy (2014) impose a very inelastic short run supply curve via a tight upper bound on $\alpha \in (0, 0.025)$, while Baumeister & Hamilton (2019) allow for substantially larger values via a truncated student-t prior $\alpha \sim t_{0,\infty}(0.1, 0.2, 3)$ centred at 0.1 and with a scale of 0.2 and 3 degrees of freedom. The exact prior choice on the magnitudes has substantial implications for estimates of the relative importance of supply and demand shocks. The reason is that forecast errors in oil production and prices are fairly uncorrelated, yielding a large class of models identified by solely sign restrictions equally consistent with the data. In such an environment, minor differences in the prior for price elasticities have large effects on the posterior of important structural quantities such as the decomposition of oil price movements into supply and demand.

To illustrate this point and how non-Gaussianity can help circumventing this problem, consider the left panel of Figure 1. Here, a scatterplot is provided using forecast errors from a bivariate VAR for global (log) oil prices and (log) oil production. When the distribution is characterized by the second moments of the forecast errors ('Gaussian Setting'), many different models are conformable with the correlation structure. Consider two arbitrarily chosen supply and demand schedules (A and B), which are observationally equivalent but imply very different dynamics for the oil market. In model A, supply is fairly inelastic and demand is elastic. Consequently, in such a model oil production would be mainly driven by supply shocks while oil prices would be largely caused by demand shocks. In turn, for model B, the supply is more elastic and demand is fairly inelastic, implying the exact opposite for the driving forces of oil prices. Incorporating prior knowledge on elasticities ultimately boils down to picking a range of models from the class of observationally equivalent models, thereby shaping the answer

about the drivers of oil price fluctuations a priori. At this point, narrative analysis and external estimates from microeconomic approaches have been used to tune these priors with great care, e.g. based on elasticity estimates provided in Newell & Prest (2019), Bjørnland, Nordvik & Rohrer (2021) and Caldara et al. (2019). However, as pointed out in Kilian (2020b), the microeconomic evidence remains controversial and has not been sufficiently informative to yield conclusive evidence.

In the right panel of Figure 1 ('Non-Gaussian Setting'), I illustrate how the irregular joint-distribution of the reduced form errors can help discriminate among observationally equivalent models. The solid lines correspond to contour lines of the estimated joint density implied by the model developed in this paper. The density estimator captures the irregular distribution and comes up with a unique supply and demand schedule consistent with the data. The model rotates the curves such that the forecast errors cluster near the supply and demand schedule, in line with heavy tailed structural shocks. Hereby, the non-Gaussian shape makes certain shifts of the supply and demand curve more likely than others, this way working as a probabilistic instrument (see also Rigobon (2003) for a similar interpretation in SVARs identified by heteroskedasticity). For the data considered in the scatter plot, the statistical identifying information points towards a very steep supply curve and rather flat demand curve. Even though the Figure is based on a stylized bivariate model including oil price and production data, it anticipates the main findings of this paper: supply shocks contribute very little to fluctuations of oil prices.

As a statistical device, identification via non-Gaussianity yields a set of shocks which per se are not useful for economic analysis. Hence, in this paper I exploit non-Gaussianity only to sharpen identification in a weakly identified, economically meaningful model. Specifically, I incorporate non-Gaussianity into the framework of Baumeister & Hamilton (2019) (BH19 henceforth). BH19 identify oil market shocks based on carefully tuned prior distributions for structural parameters, some of them truncated to reflect economically sensible sign-restrictions. Among those is the relatively diffuse student- t prior for the oil supply elasticity ($\alpha \sim t_{0,\infty}(0.1, 0.2, 3)$). Given a Gaussian likelihood function, their posterior reflects information from a combination of the prior and covariance structure in the data. I illustrate that once the Gaussian likelihood is replaced with that of a non-Gaussian model, the posterior revises substantially reflecting the additional identifying information. The revision is in line with estimated predictive densities of oil market shocks pointing towards large degrees of skewness and excess kurtosis that can aid identification. My results suggest that the posterior median estimate of the short-run oil price elasticity of supply is substantially more muted in the non-Gaussian model ($\hat{\alpha} \approx 0.02$) as opposed to the Gaussian model ($\hat{\alpha} \approx 0.15$). Consequently, oil supply shocks are found to be less important in the statistically identified model. In terms of forecast error variance decomposition of the real price of oil, the posterior median estimates a share of 6% for supply shocks, as opposed to 32% in the Gaussian model. These findings are supportive of papers that impose a strong upper bound on the supply elasticity a priori.

To model the marginal distributions of each structural shock flexibly and in an automated fashion, I develop a novel non-Gaussian SVAR model. Specifically, each structural shock is assumed to be generated by a univariate Dirichlet process mixture model (DPMM) (Escobar & West; 1995). Much like kernel density estimators, DPMM are the workhorse model in Bayesian non-parametric statistics to model unknown density functions. There are various benefits from adopting this modeling approach. First, unlike many existing approaches, no prior knowledge is required on the underlying form of non-Gaussianity which is particularly appealing if structural parameters are to be identified by the distributional assumptions. Furthermore, and particularly important given the focus of this paper, the Bayesian estimates for variance decompositions are valid even under misspecification of the marginal distributions. As discussed in Fiorentini & Sentana (2020), consistent estimation of those quantities is generally not given in non-Gaussian SVARs if the marginals are misspecified. However, one exception arises when likelihood inference is based on finite Gaussian mixture models. Although a formal proof is beyond the scope of this paper, one would expect this argument to carry over to the model considered in this paper, given that DPMMs are closely related to finite Gaussian mixture models. In fact, as shown in Neal (2000) and illustrated later in this paper, they can be obtained as a special limiting case when the number of mixture components is growing to infinity. A further point in favour of the DPMM model is that assessing the amount of non-Gaussianity in the data is a straightforward task. A simple comparison of the posterior predictive density with the kernel of a standard normal gives an indication of how much identifying information one can expect from the statistical properties of the shocks. Finally, as pointed out above, the Bayesian perspective adopted in this paper yields a coherent framework to combine economically meaningful identifying information with statistical identification.

To conduct inference, I develop a novel MCMC algorithm that iteratively draws from the conditional distributions of the VAR parameters and those of the DPMMs. While most conditionals are well known and easy to draw from, the challenging part of the algorithm is drawing the matrix A which relates forecast errors u_t to structural shocks ε_t via $Au_t = \varepsilon_t$. Here, I make use of the algorithm proposed originally in Waggoner & Zha (2003) and generalized by Villani (2009) and Chan, Koop & Yu (2021). I further extend the algorithm to allow for non-zero normalizing constraints on A . This facilitates prior elicitation on elements in A , as it separates structural parameters from the scale of structural shocks. Furthermore, it allows the use of unrestricted DPMMs, as otherwise one would have to restrict the resulting predictive densities to unit variance.

With this methodology, I contribute broadly to the econometric literature on non-Gaussian SVARs. Among classical frequentist approaches, Lanne, Meitz & Saikkonen (2017) discuss Maximum Likelihood (ML) estimation and use a t -distribution in their empirical application involving monetary policy shocks. In turn, Gouriou, Monfort & Renne (2017) consider pseudo-ML inference for independent component analysis (ICA) in general and apply it to SVAR analysis. More recently, GMM estimation is considered by Lanne & Luoto (2019) and Herwartz (2018) uses non-parametric dependence measures to disentangle non-Gaussian shocks.

Closely related to this paper is also the estimator based on finite Gaussian mixture models considered in Fiorentini & Sentana (2020) and the kernel density ICA estimator considered in Boscolo, Pan & Roychowdhury (2004). To the best of my knowledge, the only Bayesian approach in this literature is that of Lanne & Luoto (2020), who rely on t -distributed error terms to capture deviations from normality.

This paper is not the first to consider a statistical identification approach to study the sources of oil price fluctuations (see also Lütkepohl & Netšunajev (2014), Herwartz & Plödt (2016) and Lanne & Luoto (2020)). However, this paper offers some improvements towards the current state of the art in the oil market literature. First, as opposed to the previous papers, I work with the four variable workhorse model that additionally incorporates oil-inventories. Furthermore, the combined identification approach adopted in this paper assures that each structural shock is meaningful at all stages of the analysis. In turn, the aforementioned papers rely on some kind of ex-post labelling of the shocks, which is not guaranteed to yield economically meaningful shocks. More in the spirit of this paper is a recent study of Carriero, Marcellino & Tornese (2021), who also consider combining the BH19 framework with statistical identification in one of the applications. However, their paper does not look at variance decompositions of the real oil price. Interestingly, despite relying on heteroskedasticity rather than non-Gaussianity, they arrive at very similar estimates for the underlying structural parameters.

The paper is structured as follows. In Section 2, the non-Gaussian SVAR model is introduced where structural shocks follow Dirichlet process mixture models (DPMM). In section 3, the methodology is used to revisit the importance of supply and demand shocks for fluctuations in the global crude oil market. Section 4 concludes.

2 Methodology

In this section, I introduce the non-Gaussian SVAR model endowed with Bayesian nonparametric density estimators for each structural shock. I start with a quick review of the identification problem of SVARs and the standard identification results that arise under independent and non-Gaussian shocks (section 2.1). In section 2.2, I proceed with a detailed description of the nonparametric methods used to model the marginal distributions of each structural shock. The SVAR with DPMMs is discussed in section 2.3 and Bayesian inference for this model is quickly outlined in section 2.4.

2.1 Non-Gaussian SVARs

The core of the model is a linear SVAR(p) specification for the conditional mean of a K -dimensional time series vector y_t :

$$y_t = c + \sum_{j=1}^p A_j y_{t-j} + u_t, \quad u_t \sim (0, \Sigma_u), \quad (2.1)$$

$$Au_t = \varepsilon_t, \quad \varepsilon_t \sim (0, \Sigma_\varepsilon), \quad (2.2)$$

where Σ_u is a full covariance matrix and Σ_ε is diagonal. In order to match the notation of the empirical analysis, the SVAR(p) is written as an A type of model in the terminology of Lütkepohl (2005), meaning that orthogonal structural shocks (ε_t) are modelled as a linear function of reduced form errors (u_t). In this model, the reduced form covariance matrix of the VAR forecast errors is linked to the structural parameters by $\Sigma_u = A^{-1}\Sigma_\varepsilon(A^{-1})'$. Throughout the paper, stationarity is assumed, that is:

$$\det A(z) = \det(I_K - A_1 z - \dots - A_p z^p) \neq 0, \quad \text{for } |z| \leq 1.$$

Therefore, the SVAR(p) has a MA(∞) representation given by $y_t = \mu_y + \sum_{j=1}^{\infty} \Theta_j \varepsilon_{t-j}$ where $\Theta_j = \Phi_j A^{-1}$, $\Phi_0 = I_K$, $\Phi_j = \sum_{i=1}^j \Phi_{j-i} A_i$ for $j \in \mathbb{N}$ with $A_i = 0$ for $i > p$. The ik -th entry of matrix Θ_j contains the impulse response, capturing the dynamic effect of structural shock k on the i -th variable in y_t , j periods after the shock.

Without additional assumptions, the covariance structure of the forecast errors jointly identifies A and Σ_ε only up to orthogonal rotations. To see this, note that an alternative structural model can be obtained yielding equivalent second moments by defining $\tilde{A} = Q'\Sigma_\varepsilon^{-1/2}A$ and $\tilde{\Sigma}_\varepsilon = I_K$ for any orthogonal matrix Q satisfying $QQ' = I_K$ and $Q^{-1} = Q'$. For this choice, the implied covariance matrix is equivalent in that $\tilde{A}^{-1}\tilde{\Sigma}_\varepsilon(\tilde{A}^{-1})' = A^{-1}\Sigma_\varepsilon^{1/2}QQ'\Sigma_\varepsilon^{1/2}(A^{-1})' = A^{-1}\Sigma_\varepsilon(A^{-1})'$. SVAR analysis proceeds by imposing additional restrictions to pin down the parameter values of the structural model. Effectively, this boils down to fixing a scale of the shocks Σ_ε and narrowing down a set of admissible matrices Q that allows for meaningful economic analysis. Among the most popular identification strategies are short- and long run restrictions on the effect of structural shocks (Bernanke; 1986; Blanchard & Quah; 1989), sign restrictions (e.g. Uhlig (2005)) or restrictions implied by external instruments (Stock & Watson; 2012; Mertens & Ravn; 2013). Alternatively, distributional assumptions have been introduced to aid identification in SVARs, e.g. exploiting heteroskedasticity (Rigobon; 2003; Lewis; 2021) or non-Gaussianity in the structural shocks (Lanne et al.; 2017; Gouieroux et al.; 2017).

In this paper, I will exploit statistical identifying information that arises under non-Gaussianity. Following Lanne et al. (2017), this entails imposing additional assumptions on the joint distribution of the structural shocks $\varepsilon_t = [\varepsilon_{1t}, \dots, \varepsilon_{Kt}]'$, which are given by:

- (a) The stochastic vector ε_t is independent and identically distributed (i.i.d) with zero mean and each component has finite positive variance $\text{Var}(\varepsilon_{it}) = \sigma_i^2, i = 1, \dots, K$.

- (b) At least $K - 1$ components of $\varepsilon_{it}, i = 1, \dots, K$ are mutually independent and have non-Gaussian marginal distributions.

As established in Lanne et al. (2017), assumptions (a) and (b) identify the SVAR model up to permutation, sign and scale. In other words, the set of orthogonal rotation matrices yielding observationally equivalent models reduces to $Q = PD$, where P is a K -dimensional permutation matrix and D a diagonal matrix with elements ± 1 .

At this point it is worth scrutinizing these identifying assumptions. First, note that the identification results are based on assuming mutual independence, which is stronger than (typically assumed) contemporaneous uncorrelatedness and rules out higher-order dependence among structural shocks. Generally, it seems very difficult to test this assumption and ultimately its credibility depends on the empirical application at hand. However, econometric refinements have evolved in this literature to relax independence. For example, Lanne & Lütkepohl (2010) exploit a specific mixture of two normals which doesn't require independence. Also, Lanne & Luoto (2019) show that one can replace independence with less restrictive co-kurtosis restrictions that allow for different types of higher order dependence in the structural shocks.² Second, assuming Non-Gaussian marginals is perfectly compatible with common forms of heteroskedasticity observed in structural shocks, induced e.g. by GARCH dynamics (Normandin & Phaneuf; 2004; Lanne & Saikkonen; 2007) or stochastic volatility (Bertsche & Braun; 2020). As long as the volatility model implies independent shocks, they are perfectly covered by the identification argument exploited in this paper. Finally, note that the identifying restrictions are of purely statistical nature and need to be combined with economic identifying information at some point of the analysis. Typical approaches are finding an economically plausible labeling ex post or testing economic identifying restrictions which are overidentifying in the statistically identified model (Kilian & Lütkepohl; 2017). In the empirical analysis conducted in this paper, I propose to instead rely on non-Gaussianity only within a model that is already weakly identified by economic restrictions. This allows interpretation of the structural shocks at all stages of the analysis and requires no pretesting for statistical identification. Furthermore, the degree of identifying information coming from non-Gaussianity can be easily assessed by comparing the results to those obtained under Gaussian errors.

To exploit non-Gaussianity as an identification device in SVARs, various econometric techniques have emerged over the last years. For this paper, I rely on Bayesian non-parametric density estimators to model the non-Gaussian marginals.

2.2 Dirichlet process mixture models for structural shocks

Before stating the non-Gaussian SVAR used in this paper, I review the univariate Dirichlet process mixture model (DPMM) used for each shock's marginal distribution. Readers familiar with Bayesian nonparametrics may want to skip this part. For ease of readability, I will drop

²See also Drautzburg & Wright (2021) who suggest to bound higher-order dependence within a set-identified model.

the i index during the subsection, and re-introduce it at a later stage. Each structural shock will be assumed to be independent and follow the following hierarchical model for $t = 1, \dots, T$:

$$\begin{aligned}\varepsilon_t | \theta_t &\sim F(\theta_t), \\ \theta_t &\sim G, \\ G &\sim \text{DP}(G_0, \alpha),\end{aligned}$$

where $F(\theta_t)$ is a probability distribution parametrized by θ_t and can be thought of a likelihood at time t with parameters θ_t . G is the corresponding prior distribution for θ_t and has the unusual characteristic of being random itself, following a Dirichlet process (DP) $G \sim \text{DP}(G_0, \alpha)$ (Ferguson; 1973). A Dirichlet process is uniquely characterized by a base distribution G_0 and a scalar concentration parameter $\alpha \in \mathbb{R}^+$. Realizations of a DP yield almost surely discrete priors for θ_t , which is why the model can be thought of a countably infinite mixture model. In order to facilitate understanding of the resulting model, I will review two instructive representations of the DPMM.

The first is named the Pólya Urn representation which goes back to Blackwell & MacQueen (1973). The idea is to marginalize out G , as it provides a more intuitive representation of the prior implied for θ . In particular, for $t = 1, \dots, T$, the distribution can be iteratively constructed as follows:

$$\begin{aligned}\theta_t | \theta_{t-1}, \dots, \theta_1 &\sim \frac{1}{t-1+\alpha} \sum_{j=1}^{t-1} \delta_{\theta_j} + \frac{\alpha}{t-1+\alpha} G_0, \\ &\sim \sum_{j=1}^k \frac{n_j}{t-1+\alpha} \delta_{\theta_j^*} + \frac{\alpha}{t-1+\alpha} G_0,\end{aligned}$$

where $\delta_{(\cdot)}$ is the Dirac measure and $\{\theta_j^*, j = 1, \dots, k\}$ are the distinct values (“clusters”) of $\{\theta_j, j = 1, \dots, t\}$ which have cluster size $n_j = \sum_{t=1}^{t-1} \mathbb{1}(\theta_t = \theta_j^*)$. In words, the first line states that at any point of time t , θ_t may take either the value of a previously drawn parameter or be sampled from the base distribution G_0 . The Pólya Urn scheme illustrates the main properties of the DPM prior of θ_t . First, the realizations are almost surely discrete. Second, there is a “richer get richer” property implied by the model which leads to heavy clustering of the mixing parameters θ . This is highlighted in the second line, where it becomes clear that the probability of θ_t joining a certain cluster θ_j^* increases in the cluster size n_j . Therefore, the model for ε_t can be interpreted as a flexible yet parsimonious mixture model where the number of components is random and increasing in the sample size T . The strength of clustering is governed by the concentration parameter α and lower values are associated with less mixture components (clusters) for a given sample size T . Finally, the choice of Base distribution G_0 will drive the location of the clusters.

A second convenient representation of the DPMM relates the model to more frequently used finite mixture models. As outlined in Neal (2000), a link can be established by casting a finite

mixture model of a certain form and letting the number of mixture components grow to infinity. In particular, for the following model with k mixture components:

$$\varepsilon_t | c_t, \theta^* \sim F(\theta_{c_t}^*) \quad (2.3)$$

$$c_t | p \sim \text{Discrete}(p_1, \dots, p_k) \quad (2.4)$$

$$p \sim \text{Dirichlet}(\alpha/k, \dots, \alpha/k) \quad (2.5)$$

$$\theta_j^* \sim G_0, \quad j = 1, 2, \dots \quad (2.6)$$

where c_t is a discrete assignment variable linking each observation to a certain mixture component. Each component is associated with a unique parameter θ_j^* which are drawn from the base distribution G_0 . If the mixing proportions $p = (p_1, \dots, p_k)$ are given a symmetric Dirichlet prior with concentration parameters α/k , a DPMM can be obtained when $k \rightarrow \infty$. Exploiting well known properties of the Dirichlet Multinomial distribution, the conditional probability of c_t given the sequence $\{c_{t-1}, \dots, c_1\}$ can be shown to be (Neal; 2000):

$$P(c_t = c | c_{t-1}, \dots, c_1) = P(c_{t-1}, \dots, c_1, c_t = c) / P(c_{t-1}, \dots, c_1) = \frac{n_{t,c} + \alpha/k}{t - 1 + \alpha},$$

where $n_{t,c}$ is the number of c_j for $j < t$ equal to c , that is the size of the clusters. Hence, when $k \rightarrow \infty$:

$$P(c_t = c | c_{t-1}, \dots, c_1) \rightarrow \frac{n_{t,c}}{t - 1 + \alpha_i}, \quad (2.7)$$

$$P(c_t \neq c_j \text{ for all } j < t | c_{t-1}, \dots, c_1) \rightarrow \frac{\alpha}{t - 1 + \alpha}, \quad (2.8)$$

where the first line gives the probability that the t -th shock ε_t is associated with cluster c , while the second line gives the residual probability that ε_t is associated with a cluster not observed in $\{c_{t-1}, \dots, c_1\}$. When compared with the Pólya Urn representation, these equations yield the same clustering behaviour and an equivalent model representation.

To get a fully operational DPMM, one needs to specify a density $F(\theta)$ and a corresponding base distribution G_0 . For this paper, I adopt a simple yet very popular specification pioneered by Escobar & West (1995), implying that $F(\theta_t)$ is a Gaussian distribution parametrized by mean μ_t and variance σ_t^2 , hence $\theta_t = (\mu_t, \sigma_t^2)'$. For computational convenience, a conjugate base distribution G_0 is chosen which is the normal inverse gamma $(\mu, \sigma^2) \sim \mathcal{NiG}(s/2, S/2, m, \tau) \sim p(\sigma^2)p(\mu|\sigma^2)$, where $p(\sigma^2) \sim \mathcal{iG}(s/2, S/2)$ is inverse Gamma and $\mu|\sigma^2 \sim \mathcal{N}(m, \tau\sigma^2)$ normal. Besides striking a fair balance between flexibility and computational complexity, the Gaussian mixture model benefits from the arguments outlined in Fiorentini & Sentana (2020) regarding consistent estimation of unconditional standard deviations under misspecification of the error term.

For the Gaussian DPMM, it is instructive to look at the implied predictive density conditional on a (prior or posterior) draw of the mixture parameters $\theta_{1:T} = \{\theta_T, \dots, \theta_1\}$:

$$\begin{aligned} p(\varepsilon_{T+1}|\theta_{1:T}) &= \int p(\varepsilon_{T+1}|\theta_{T+1})p(\theta_{T+1}|\theta_{1:T})d\theta_{T+1} \\ &= \frac{1}{\alpha + T} \sum_{t=1}^T \phi(\varepsilon_{T+1}; \mu_t, \sigma_t) + \frac{\alpha}{\alpha + T} T_s(\varepsilon_{T+1}; m, M), \\ &= \sum_{j=1}^k \frac{n_j}{t-1+\alpha} \phi(\varepsilon_{T+1}; \mu_j^*, \sigma_j^*) + \frac{\alpha}{\alpha + T} T_s(\varepsilon_{T+1}; m, M), \end{aligned}$$

where $\phi(\cdot; \mu, \sigma^2)$ denotes the density of the normal distribution and $T_s(\cdot; m, M)$ the density of a student t with mode m , scale $M^{1/2}$ for $M = (1 + \tau)S/s$ and s degrees of freedom. At first sight, the predictive density shares some similarities with the popular Gaussian kernel density estimator $p(\varepsilon_{T+1}|\varepsilon_{1:T}) \propto \sum_{t=1}^T \phi(\varepsilon_{T+1}; \varepsilon_t, H)$ where H is a global smoothing parameter. However, there are a few key differences worth mentioning. First, the fact that the DP induces heavy clustering in θ_t means the predictive is shrunk towards a finite set of k local modes $\{\mu_j^*, j = 1, \dots, k\}$. Furthermore, the component variances $\{\sigma_j^*, j = 1, \dots, k\}$ may differ allowing for local smoothing. Finally, the density is shrunk globally towards that of a t -distribution, with decreasing importance as sample size increases. The global smoothing parameter α governs both the strength of clustering (and hence sparsity) in $\theta_{1:T}$ as well as the strength of shrinkage towards the t - density. For more details and theoretical insights including consistency and convergence rates see e.g. Escobar & West (1995), Ghosal, Ghosh, Ramamoorthi et al. (1999) and Ghosh & Ramamoorthi (2003).

With respect to computational simplicity, adopting a conjugate base distribution facilitates MCMC inference on the mixing parameters $\theta_t, t = 1, \dots, T$ substantially. To see this, recall that the structural shocks ε_t are assumed to be independent and hence exchangeable, which yields the following prior based on the Pólya Urn representation:

$$\theta_t|\theta_{-t} \sim \frac{1}{T-1+\alpha} \sum_{j \neq t} \delta_{\theta_j} + \frac{\alpha}{T-1+\alpha} G_0,$$

where $\theta_{-t} = \{\theta_j, j \neq t\}$. Combined with the likelihood $F(\varepsilon_t|\theta_t)$, the posterior is given by the following mixture:

$$\theta_t|\theta_{-t}, \varepsilon_t \sim \sum_{t \neq j} q_{tj} \delta_{\theta_j} + r_t H_t, \tag{2.9}$$

where $q_{tj} = bF(\varepsilon_t|\theta_j)$, $r_t = b\alpha \int F(\varepsilon_t|\theta) dG_0(\theta)$ and H_t posterior of θ based on G_0 and ε_t . For the conjugate choice G_0 , the posterior is analytically tractable and of known form, implying that r_t can be computed in closed form and a random sample of H_t can be readily generated.

While cycling through the conditionals in (2.9) is certainly easy, it may lead to poor convergence. Hence, in this paper I rely on a refinement developed in Neal (2000) improving posterior

mixing. Akin to the finite mixture representation (equations (2.3)-(2.6)), the algorithm exploits that $\theta_t = \theta_{c_t}^*$ can be represented in terms of latent allocation variables c_t and unique cluster parameters θ_j^* . Combining the prior for c_t implicit in equations (2.7)-(2.8) with the likelihood conditional on cluster parameters, this yields the simple conditionals for $t = 1, \dots, T$:

$$P(c_t = c_j, j = 1, \dots, k | c_{-t}, \varepsilon_t) = b \frac{n_{-t, c_j}}{T - 1 + \alpha} F(\varepsilon_t | \theta_{c_t}^*), \quad (2.10)$$

$$P(c_t \neq c_j \text{ for all } j \neq t | c_{-t}, \varepsilon_t) = b \frac{\alpha}{T - 1 + \alpha} \int F(\varepsilon_t | \theta) dG_0(\theta), \quad (2.11)$$

where $c_{-t} = \{c_j, j \neq t\}$, $c_j, j = 1, \dots, k$ are unique values in c_{-t} of count n_{-t, c_j} and b is a normalizing constant. In a second step, conditional on the assignment variables and exploiting the conjugacy of G_0 , the (active) cluster parameters $\theta_j^*, j = 1, \dots, k$ can be drawn from known distributions in a straightforward manner. The resulting algorithm is reliable, easy to implement and widely used in DPMM.

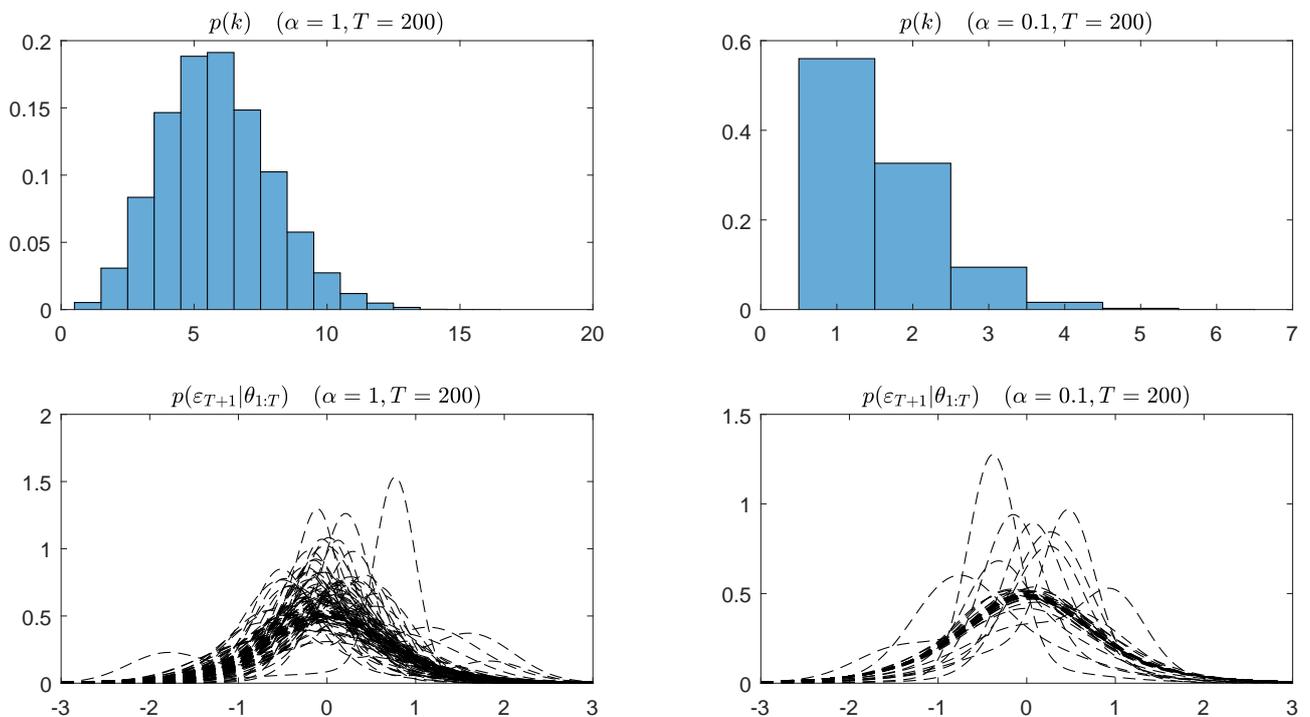


Figure 2: Top: implied prior for the number of clusters K . Base distribution is given by $\mathcal{NiG}(s/2 = 5, S/2 = 3/5, m = 0, \tau = 2)$. Bottom: $p(\varepsilon_{T+1} | \theta_{1:T})$ based on 50 prior draws from $\theta_{1:T}$.

At this stage, it is worth highlighting the crucial role of the smoothing parameter α . To illustrate the impact of α on the complexity of the model, note Figure 2. For a given value of α , the graph shows the implied distribution for the number of unique clusters k and a set of 50 arbitrary predictive densities obtained conditional on drawing $\theta_{1:T}$. The sample size underlying the Figure is set to $T = 200$, reflecting typical time series lengths in macroeconomics. For $\alpha = 1$ (left column, larger value), most of the prior probability mass for k concentrates at values below 10, with a mode between 5 and 6. The predictive densities illustrate the wide range of

distributions that can be generated under the DPMM, displaying all kinds of multimodality, skewness and fat tails. On the other hand, a smaller value $\alpha = 0.1$ (right column) implies that the prior mass for the number of clusters k concentrates at much lower values, with prior mode at just one component. Obviously, this translates into the prior predictive to be much more concentrated around unimodal shapes, although the variability remains high.

Hence, akin to the bandwidth parameter in kernel density estimation, α can be thought of as the global smoothing parameter that governs the flexibility of the underlying density estimator. To come up with a sensible choice for α , I follow standard practice and infer the value from the data (Escobar & West; 1995). This can be implemented by treating α as another random variable of the model. In this paper, the parameter is given a gamma prior $\alpha \sim \mathcal{G}(a_\alpha, b_\alpha)$, which allows for simple posterior inference. Similarly, τ and m are treated as random in order to let the prior adjust to different scales of the structural shocks. Conjugate hyperpriors are specified for simplicity, that is $\tau \sim i\mathcal{G}(a_\tau/2, b_\tau/2)$ and $m \sim \mathcal{N}(m_m, V_m)$. Overall, through the incorporation of these hyperpriors one obtains a fully automatic procedure, requiring minimal input by the researcher.

2.3 SVAR-DPMM

The next step is to embed the DPMMs into an SVAR model, which yields the methodology used in the empirical analysis. Let $x_t = [y'_{t-1}, \dots, y'_{t-p}, 1]'$ and stack the autoregressive coefficients into the $n \times np + 1$ matrix $A_+ = [A_1, A_2, \dots, A_p, c]$. Furthermore, denote by $\theta_{it} = [\mu_{it}, \sigma_{it}^2]$ and assume the availability of p fixed presample values y_0, \dots, y_{-p+1} . Then, the full hierarchy of the model reads:

$$A(y_t - A_+x_t) = \varepsilon_t, \tag{2.12}$$

$$\varepsilon_{it} | \theta_{it} \sim \mathcal{N}(\mu_{it}, \sigma_{it}^2), \tag{2.13}$$

$$\theta_{it} \sim G_i, \tag{2.14}$$

$$G_i \sim \text{DP}(G_{i0}, \alpha_i), \tag{2.15}$$

$$G_{i0} \sim \mathcal{NiG}(s_i/2, S_i/2, m_i, \tau_i), \tag{2.16}$$

for $i = 1, \dots, K$, $t = 1, \dots, T$. Here, equation (2.12) corresponds to the SVAR model (section 2.1) while equations (2.13)-(2.16) to the individual DPMM specified for each structural shock (section 2.2). Denote by $A_{i\bullet}$ the i th row of A . The following prior distributions are considered for the underlying model parameters, which completes the specification:

$$A_{i\bullet} \sim p(A_{i\bullet}), \tag{2.17}$$

$$\alpha_+ \sim \mathcal{N}(m_{\alpha_+}, V_{\alpha_+}), \tag{2.18}$$

$$\alpha_i \sim \mathcal{G}(a_\alpha, b_\alpha), \tag{2.19}$$

$$\tau_i \sim i\mathcal{G}(a_\tau, b_\tau), \tag{2.20}$$

$$m_i \sim \mathcal{N}(m_m, V_m), \tag{2.21}$$

for $i = 1, \dots, K$ and $\alpha_+ = \text{vec}(A_+)$. Similar to Baumeister & Hamilton (2015), the prior of the structural parameters in A is allowed to take an arbitrary form, enabling the researcher to incorporate any identifying information with high degree of flexibility. To facilitate efficient inference, however, I assume prior independence between different rows of A . As I discuss in Appendix A.1, this allows me to use an extension of the efficient algorithm of Waggoner & Zha (2003) to draw from the conditional posterior of A . For the vectorized reduced form slope parameters α_+ , a Gaussian prior is specified, a fairly common choice which allows for straightforward inference. The normal prior is widely used in VAR analysis and flexible enough to accommodate both non-informative priors as well as a variety of shrinkage priors including the popular Minnesota prior (Litterman; 1986). Finally, the K concentration parameters $\{\alpha_i, i = 1, \dots, K\}$ are given independent Gamma prior distributions, while the Base distribution parameters $\{\tau_i, m_i, i = 1, \dots, K\}$ have respective conjugate priors, yielding a fully automatic procedure as described in section 2.2.

When embedding DPMMs within the SVAR model, some care must be taken with respect to identifiability of location and scale of the shocks. First, unlike Gaussian errors, the marginals arising from DPMMs are not guaranteed to be mean zero. Hence, the intercept of the VAR model is not identified, and can be readily dropped. Alternatively, one may simply ignore the issue as usual quantities important for structural analysis remain unaffected, including impulse response functions or variance decompositions. With respect to scale, a similar problem arises. While in Gaussian SVARs the scale is often fixed to unity, doing so within DPMMs is rather involved, see e.g. the approach taken in an earlier version of this paper based on methodology developed in Yang, Dunson & Baird (2010). For this paper, I follow the model of BH19 and identify the scale of the shocks by normalizing certain elements in A to unity. This is particularly natural if the empirical model can be written as a simultaneous equation system, as is the case for the oil market model considered in this paper. Finally, recall that non-Gaussianity identifies shocks up to an arbitrary permutation (see section 2.1). In this paper, a unique labelling is obtained through economic restrictions reflected in the prior of A .

2.4 Posterior inference

In the following, I quickly describe posterior inference. Denote the collection of parameters $\varphi = \{A, \alpha_+, \alpha_i, \tau_i, m_i, i = 1, \dots, K\}$ and define the collection of auxiliary mixing parameters as $\theta = \{\theta_{it}, i = 1, \dots, K, t = 1, \dots, T\}$. The posterior distribution of φ based on observed data Y is proportional to prior times likelihood $p(\varphi|Y) \propto p(Y|\varphi)p(\varphi)$. Note that for DPMM models, the likelihood itself is not directly available, but must be obtained by integrating out the auxiliary parameters θ , that is $p(Y|\varphi) = \int p(Y|\theta, \varphi)p(\theta|\varphi)d\theta$. Since both likelihood and posterior are intractable, a full-scale MCMC algorithm is used in this paper to conduct posterior inference on the augmented set of parameters $\xi = \{\varphi, \theta\}$.

In the following, I will quickly sketch the algorithm at a high level, and refer to Appendix A.1 for a detailed description. Let ξ_{-x} be all parameters in ξ but x , and initialize the procedure by

choosing some arbitrary initial values. Then, the algorithm iteratively draws from the following blocks of conditionals:

- (1) Draw the SVAR structural parameters from $p(A|Y, \xi_{-\{A\}})$ via an extension of the Algorithm proposed in Waggoner & Zha (2003). This involves drawing from the conditional distribution of each row $A_{i\bullet}$ separately. Denote 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. In Appendix A.1, I show how to draw from $p(a_i|Y, \xi_{-\{a_i\}})$ when $w_i \neq 0$, using either a uniform or Gaussian prior for a_i . Under a more general prior, as for example those considered in the empirical application of this paper, a Metropolis Hastings step can be added to correct for the difference in prior density between proposed and current value of a_i . For the priors considered in the empirical application, the MH acceptance probabilities are very high and vary between 0.89 and 0.99 depending on the row of A .
- (2) Draw the VAR regression parameters from $p(\alpha_+|Y, \xi_{-\{\alpha_+\}})$ which is available in closed form.
- (3) Draw the DPMM parameters as in Neal (2000) and the hyperparameters from their conditionals as outlined in Escobar & West (1995).

In order to compute variance decompositions and historical decompositions in the SVAR-DPMM model, it is necessary to back out the unconditional variance of structural shocks. Within the MCMC algorithm, it is straightforward to recover these moments from the predictive density. Conditional on a draw of θ , it is:

$$p(\varepsilon_{i,T+1}|\theta_{i,1:T}) = \sum_{j=1}^k \frac{n_{ij}}{T-1+\alpha_i} \phi(\varepsilon_{i,T+1}; \mu_{ij}^*, \sigma_{ij}^*) + \frac{\alpha_i}{\alpha_i+T} T_{s_i}(\varepsilon_{i,T+1}; m_i, M_i),$$

where $M = (1 + \tau_i)S_i/s_i$. Effectively, this is a mixture of $k + 1$ distributions with component weights given by $w_{ij} = \frac{n_{ij}}{T-1+\alpha_i}$, $j \leq k$ and $w_{i,k+1} = \frac{\alpha_i}{\alpha_i+T}$. Corresponding component means are $\mu_{ij}^c = \mu_{ij}^*$, $j \leq k$ and $\mu_{i,k+1}^c = m_i$, while variances are given by $(\sigma_{ij}^c)^2 = (\sigma_{ij}^*)^2$, $j \leq k$ and $(\sigma_{i,k+1}^c)^2 = M_i \frac{s_i}{s_i-2}$. Hence, mean and variance of the predictive can be backed out by standard formulas for mixture models:

$$E(\varepsilon_{i,T+1}|\theta_{i,1:T}) = \mu_i = \sum_{j=1}^{k+1} w_{ij} \mu_{ij}^c, \quad (2.22)$$

$$\text{Var}(\varepsilon_{i,T+1}|\theta_{i,1:T}) = \sigma_i^2 = \sum_{j=1}^{k+1} w_{ij} ((\sigma_{ij}^c)^2 + (\mu_{ij}^c)^2 - \mu_i^2). \quad (2.23)$$

Given equations (2.22)-(2.23) and posterior draws of ξ , inference for mean and variance of the shocks is a straightforward by-product to obtain from the algorithm.

3 The importance of supply and demand for oil prices

In the following section, the methodology is applied to revisit the importance of supply and demand shocks for oil price fluctuations. The empirical strategy is kept simple. Throughout the analysis, I revisit the four variable oil market model considered in recent papers (Kilian & Murphy; 2014; Baumeister & Hamilton; 2019) and recover shocks with two different identification strategies. The first strategy closely follows the approach put forward by Baumeister & Hamilton (2019) (**BH19**). Specifically, BH19 impose a set of sign restrictions combined with weakly informative prior distributions on structural parameters, mainly oil price elasticities. Combined with a Gaussian likelihood, the resulting posterior distribution reflects information of the prior and covariance structure in the data. The second identification strategy relies on the same identifying information but in addition, exploits non-Gaussianity in the shocks (**BH19+NG**). Hence, any difference in the posteriors between the two identification approaches will reflect the additional statistical identifying information from non-Gaussianity.

3.1 Model and identification

The model is based on the following variables, exactly mimicking BH19:

$$y_t = [100 \times \Delta q_t, 100 \times \Delta y_t^a, 100 \times \Delta p_t, \Delta i_t]'$$

where q_t is the log of global crude oil production (in million barrels per day) and y_t^a is a measure of world economic activity proxied by the industrial production indices of OECD countries plus 6 major countries. Furthermore, p_t is the log of the real oil price defined as the US Refiner's Acquisition Cost of oil, deflated with the US consumer price index. Finally, Δi_t is a proxy for OECD oil inventories expressed as a fraction of global crude oil production. In order to allow for sufficient dynamics, the model includes $p = 12$ lags in the VAR. For estimation, monthly data from 1974m1 until 2019m12 is considered, which is slightly longer than the data in BH19. For more details on the model choices including a description of the dataset, I refer to the paper of BH19.

Abstracting from lags and Δ notation, the structural oil market model takes the form of the following simultaneous equation system:

$$\textit{Supply:} \quad q_t = \alpha_{qp} p_t + \varepsilon_t^s \quad (3.1)$$

$$\textit{Economic activity} \quad y_t^a = \alpha_{py} p_t + \varepsilon_t^{ad} \quad (3.2)$$

$$\textit{Consumption demand} \quad q_t - i_t^* = \beta_{qy} y_t^a + \beta_{qp} p_t + \varepsilon_t^{cd} \quad (3.3)$$

$$\textit{Inventory demand} \quad i_t^* = \psi_1 q_t + \psi_3 p_t + \varepsilon_t^{id} \quad (3.4)$$

$$\textit{Measurement error:} \quad i_t = \chi i_t^* + \varepsilon_t^{me} \quad (3.5)$$

where $\varepsilon_t = [\varepsilon_t^s, \varepsilon_t^{ea}, \varepsilon_t^{cd}, \varepsilon_t^{id}, \varepsilon_t^{me}]' \sim (0, \Sigma_\varepsilon)$ are uncorrelated structural shocks, which implies that Σ_ε is diagonal. There are five equations that summarize the contemporaneous relations across

the variables. First, consider equation (3.5), which reflects an assumption about additive measurement error in the observed inventories variable i_t . Specifically, it decomposes the variable into an unobserved “true” inventory series i_t^* and a measurement error ε_t^{me} . BH19 rationalize this approach by noting that inventory data is only available for OECD countries, which is arguably only a fraction χ of world inventories. Equation (3.1) characterizes the behaviour of global oil supply, relating production to real oil prices via the coefficient α_{qp} . Given that both variables are expressed in log deviations, the coefficient α_{qp} can be interpreted as the (short-run) price elasticity of oil supply. The third equation (3.2) characterizes global economic activity (EA), decomposing world industrial production into a component driven by oil prices and an EA shock ε_t^{ea} . Equation (3.3) models consumption demand, relating quantity consumed $q_t - i_t^*$ to world output and oil prices. Here, β_{qp} is the oil price elasticity of demand while β_{qy} characterizes the response of demand to increased economic activity. Finally, equation (3.4) captures residual demand for oil inventory which is related to quantity and prices via coefficients $\psi_{1/2}$.

The simultaneous equation model can be written as an A type structural VAR. To see this, define an augmented set of VAR forecast errors by $\tilde{u}_t = [u_t, u_t^{i^*}]'$, where $u_t = y_t - A_+ x_t$ are standard VAR forecast errors of the observables and $u_t^{i^*}$ is an unobserved prediction error for the “true” (latent) inventory series. Then, the model can be written as:

$$\underbrace{\begin{pmatrix} 1 & 0 & -\alpha_{pq} & 0 & 0 \\ 0 & 1 & -\alpha_{py} & 0 & 0 \\ 1 & -\beta_{qy} & -\beta_{qp} & 0 & -1 \\ -\psi_1 & 0 & -\psi_3 & 0 & 1 \\ 0 & 0 & 0 & 1 & -\chi \end{pmatrix}}_A \underbrace{\begin{pmatrix} u_t^q \\ u_t^y \\ u_t^p \\ u_t^i \\ u_t^{i^*} \end{pmatrix}}_{\tilde{u}_t} = \underbrace{\begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^{ea} \\ \varepsilon_t^{cd} \\ \varepsilon_t^{id} \\ \varepsilon_t^{me} \end{pmatrix}}_{\varepsilon_t}. \quad (3.6)$$

A simple counting exercise reveals that this model cannot be identified from the data, given that the fifth variable is unobserved. In particular, there are 12 structural parameters (7 elements in A plus 5 elements in Σ_ε) but there are only 10 reduced form parameters available in the covariance matrix of the observable prediction errors u_t .

To identify the model, the following two identification strategies are considered. In the first specification (**BH19**), structural shocks are modelled as Gaussian. Here, identification is obtained via a set of sign-restrictions and prior distributions. In the second specification, the same set of sign-restrictions and prior information is used, with the additional identifying assumption that all shocks but the measurement error (ε_t^{me}) are mutually independent and non-Gaussian (**BH19+NG**). Specifically, they are modelled by the non-parametric density estimators considered in section 2. Here, identification will come from a combination of statistical identification and the economically motivated prior information.

The exact priors used for each parameter are set out in Table 1. First, consider the priors for the structural parameters underlying A. Regarding the oil price elasticities of supply α_{qp} and demand β_{qp} , BH19 make use of truncated student- t distributions concentrated around 0.1 and -0.1 respectively, reflecting their view that they should be rather small. However, with scales

of 0.2 and 3 degrees of freedoms, the distributions are only weakly informative. As for the income elasticity β_{qy} , BH19 draw on external evidence from the literature to elicit a positively truncated student- t distribution with mode around 0.7, scale 0.2 and 3 degrees of freedom. The effect of oil prices on economic activity α_{yp} is judged to be rather small, reflected in a (negatively) truncated t -distribution with mode at just -0.05 . A smaller scale of 0.1 reflects more prior certainty than for the other parameters endowed with a t -prior, but the degrees of freedom are still set to 3, hence the distribution is relatively spread out. For the parameters of the inventory equation $\psi_{1/2}$, no prior knowledge is available, so uninformative student- t priors are used concentrated around 0 with scale of 0.5 and 3 degrees of freedom. With respect to χ , the fraction of inventories held by OECD countries, BH19 specify a Beta prior concentrated around 0.6, matching roughly the share of OECD countries in world oil consumption. The prior parameters are set in such a way that the standard deviation is equal to 0.1, reflecting a moderate degree of uncertainty for this number.

As for the diagonal elements of Σ_ε , in the Gaussian model they are given uninformative inverse Gamma priors for all shocks but the measurement error. In the non-Gaussian model, shock variances are indirectly parameterized. A series of hyperpriors are tabulated instead for the parameters underlying the base distribution and the concentration parameters ($\{\tau_i, m_i, \alpha_i, i = 1, \dots, 4\}$). The underlying priors are chosen such that they render a fairly automatic procedure which can adapt to different scales of each structural shock.

Finally, consider the measurement error, which in both identification strategies is assumed to be Gaussian, that is $\varepsilon_t^{me} \sim \mathcal{N}(0, \sigma_5^2)$. Instead of a prior for σ_5^2 , BH19 use a prior on the importance of the measurement error in a regression of u_t^i on u_t^p , given by $\rho = \frac{\chi^{-1}\sigma_5^2}{\sigma_3^2 + \chi^{-2}\sigma_5^2}$. This is motivated by the fact that, since u_t^{i*} is unobserved, the Algorithm developed in Baumeister & Hamilton (2015) cannot directly be applied to the oil market model. To get around this issue, BH19 rewrite the first four equations of (3.6) using observables. Algebraic manipulations yield $A^\dagger u_t = \varepsilon_t^\dagger$, for

$$A^\dagger = \begin{pmatrix} 1 & 0 & -\alpha_{qp} & 0 \\ 0 & 1 & -\alpha_{yp} & 0 \\ 1 & -\beta_{qy} & -\beta_{qy} & -\chi^{-1} \\ -\tilde{\psi}_1 & 0 & -\tilde{\psi}_3 & 1 \end{pmatrix},$$

$\psi_{1/2}^\dagger = \chi\psi_{1/2}$ and $\varepsilon_t^\dagger = [\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t} - \chi^{-1}\varepsilon_{5t}, \chi\varepsilon_{4t} + \varepsilon_{5t}]$. BH19 then show that premultiplying the system further by

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \rho & 1 \end{pmatrix}$$

yields orthogonal shocks $\varepsilon_t^* = \Gamma\varepsilon_t^\dagger$, hence allows the use of their standard algorithm for $A^* = \Gamma A^\dagger$. Since by construction $\rho \in (0, \chi)$, they use a Beta prior centred around 0.25χ as to reflect a moderate importance of the measurement error.

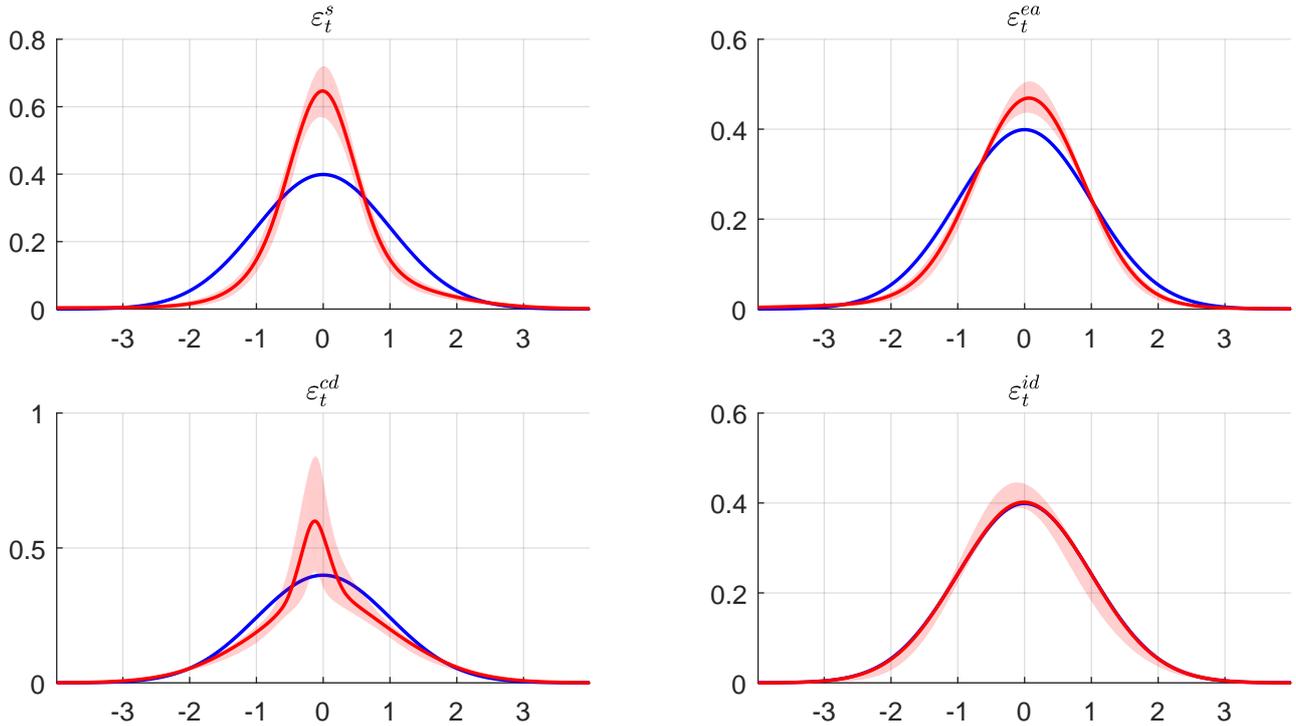


Figure 3: Posterior predictive densities (90% credible interval) of standardized structural shocks $\tilde{\varepsilon}_{i,T+1} = \sigma_i^{-\frac{1}{2}}(\varepsilon_{i,T+1} - \mu_i)$.

In the non-Gaussian model, however, this workaround cannot be applied since by construction the residuals of the transformed model (ε_t^*) are not independent invalidating the identification strategy. Therefore, I incorporate an additional step into the MCMC as to infer the latent inventory series u_t^{i*} . Furthermore, I include σ_5^2 into the last row of A and normalize $\varepsilon_{5t} \sim \mathcal{N}(0, 1)$, yielding $A_{5\bullet} = [0, 0, 0, \sigma_5^{-1}, -\chi\sigma_5^{-1}]'$. Finally, instead of a prior for σ_5^2 , I use a Beta prior on the fraction of variance in inventories explained by the measurement error. The resulting coefficient is given by $\rho^* = \frac{\sigma_5^2}{\chi^{-2} \text{var}(u_t^i) + \sigma_5^2} \in (0, 1)$, and I use pre-1974 inventory data to set $\text{var}(u_t^i) \approx 1.3$.³ Reflecting a moderate degree of importance, I set the Beta prior such that $E(\rho^*) = 0.25$ with standard deviation 0.12. To maintain comparability with the Gaussian model, I use the same prior for both identification strategies. The modified algorithm that includes inference on u_t^{i*} is presented in Appendix A.2. Also, a description of the algorithm used for the modified Gaussian model is described in Appendix A.3.

Finally, the prior for the autoregressive coefficients α_+ are set to be uninformative centred at zero with a fairly large variance.

Table 1: Summary of prior distributions

Student t distribution					
		location	scale	dof	sign restriction
α_{qp}	Oil supply elasticity	0.1	0.2	3	$\alpha_{qp} > 0$
α_{yp}	Effect of p on activity	-0.05	0.1	3	$\alpha_{yp} < 0$
β_{qy}	Income elasticity of oil demand	0.7	0.2	3	$\beta_{qy} > 0$
β_{qp}	Oil demand elasticity	-0.1	0.2	3	$\beta_{qp} < 0$
ψ_1	Effect of q on inventories	0	0.5	3	none
ψ_3	Effect of p on inventories	0	0.5	3	none
Beta distribution					
		mean	standard deviation		sign restriction
χ	Fraction of inventories	0.6	0.1		$0 \leq \chi \leq 1$
ρ^*	Importance of measurement error in u_t^{ix}	0.25	0.12		$0 \leq \rho^* \leq 1$
Normal distribution					
		mean	variance		sign restriction
α_+	vector of autoregressive parameters	0	100		none
Inverse Gamma distribution (only Gaussian model)					
		mean	variance		sign restriction
$\sigma_i^2, i = 1, 2, 3, 4$	shock variances	1	10		$\sigma_i^2 > 0$
Gamma distribution (only non-Gaussian model)					
		mean	variance		sign restriction
$\alpha_i, i = 1, 2, 3, 4$	concentration parameter	1	1		$\alpha_i > 0$
$\tau_i, i = 1, 2, 3, 4$	scale of base distribution	1	1		$\tau_i > 0$
Normal distribution (only non-Gaussian model)					
		mean	variance		sign restriction
$m_i, i = 1, 2, 3, 4$	location of base distribution	0	10		none

3.2 Empirical results

Before discussing standard results from structural analysis obtained under the two identification approaches, it is useful to check if the structural shocks display non-Gaussianity in the first place. According to the identification conditions, non-Gaussianity is required if the statistical properties of the shocks are to be exploited for identification. In Figure 3, posterior median estimates of the predictive densities are provided for the standardized structural shocks, alongside 90% posterior confidence sets (shaded area). Furthermore, for comparison, the density of a standard normal distribution is drawn in blue. Clearly, three out of four structural shocks are displaying large degrees of non-Gaussianity in some regions of the predictive density. This hints towards considerable identifying information that can be exploited in the context of the oil-market application.

³A prior on ρ would be difficult to implement in the non-Gaussian framework since the implied prior for $A_{5\bullet}$ would depend on the DPMM model of the third shock. To see this, note that solving for σ_5^2 yields $\sigma_5^2 = \frac{\rho\sigma_3^2}{\chi^{-1} - \rho\chi^{-2}}$ and σ_3^2 is a function of the DPMM mixture parameters (see section 2.4).

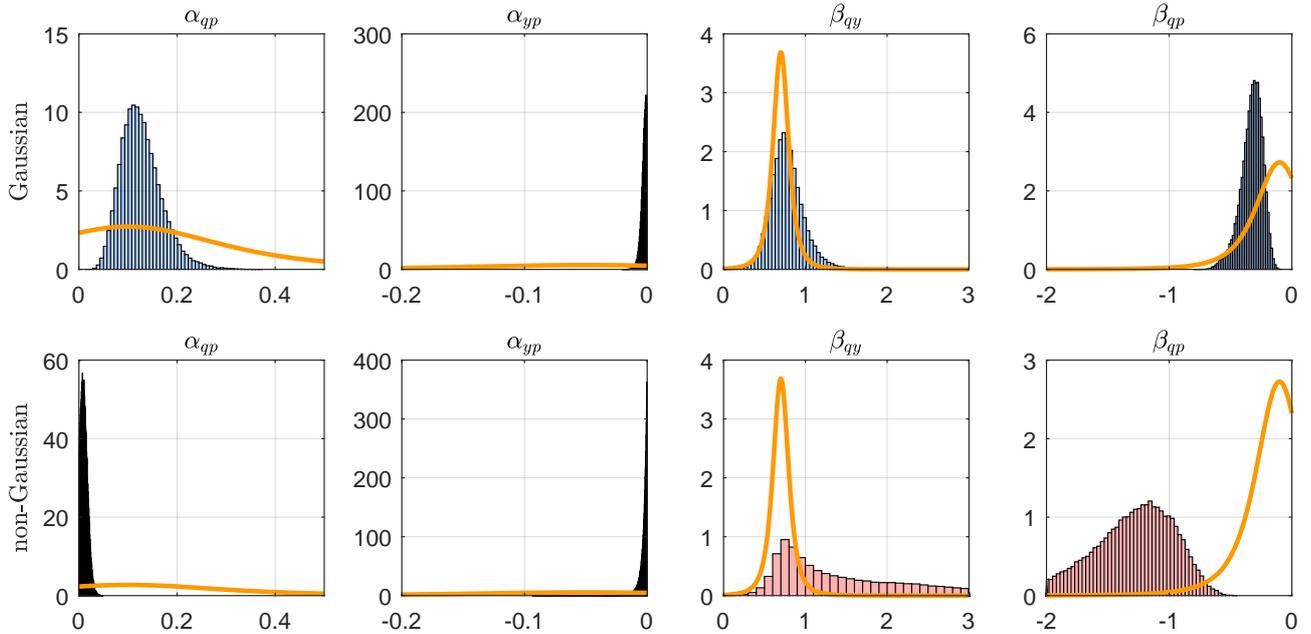


Figure 4: Prior (orange line) and posterior density of key structural parameters. Top panel: Gaussian model. Bottom panel: non-Gaussian model.

Given the large degree of non-Gaussianity in the structural shocks, one may expect that the posterior distributions of the structural parameters differ across the two identification schemes. In Figure 4, I plot prior and posterior distributions for key structural parameters in the model and indeed, find considerable differences. The first column compares prior and posterior distribution of α_{qp} , the short-run oil price elasticity of supply. Under a Gaussian likelihood (top), the prior distribution is peaking close to the prior mode, although uncertainty decreased substantially. In contrast, in the non-Gaussian model, posterior mass is concentrated very close to zero. Interestingly, the estimates obtained under non-Gaussianity supports very sharp upper-bound restrictions used previously in the literature (Kilian & Murphy; 2014; Herrera & Rangaraju; 2020). With respect to the effect of oil prices on activity, α_{yy} , the posterior distributions are very similar across both identification approaches. Compared to the prior, both posteriors concentrate strongly around values close to zero. Stronger differences in the posteriors are visible in the parameters underlying the consumption demand equation. With respect to the income elasticity of oil demand (β_{qy}), prior and posterior coincide in the Gaussian model, which may reflect that there is very little information in the covariance structure of the data to learn about this parameter. In the non-Gaussian model, however, the prior distribution is updated to some extent. While the modal value is still below one, a remarkable degree of posterior mass is attached to larger values. A similar picture arises for the oil demand elasticity (β_{qp}). In the Gaussian model, the prior is only slightly revised but posterior mass still concentrates around high density regions of the prior. Instead, in the non-Gaussian model, the posterior is revised to a much larger extent. The posterior mode indicates that the demand elasticity is estimated to be much larger than indicated by the prior, with a modal value slightly above unity. However, posterior uncertainty remains high in the non-Gaussian model. Readers interested in the

remaining parameters underlying the simultaneous equation model are referred to Appendix C. Overall, the posterior plots indicate that there is substantial identifying information that can be exploited from the statistical properties of oil market shocks. The results suggest that once this information is taken into account, (short-run) oil supply is estimated to be considerably more inelastic while consumption demand is found to be more elastic.

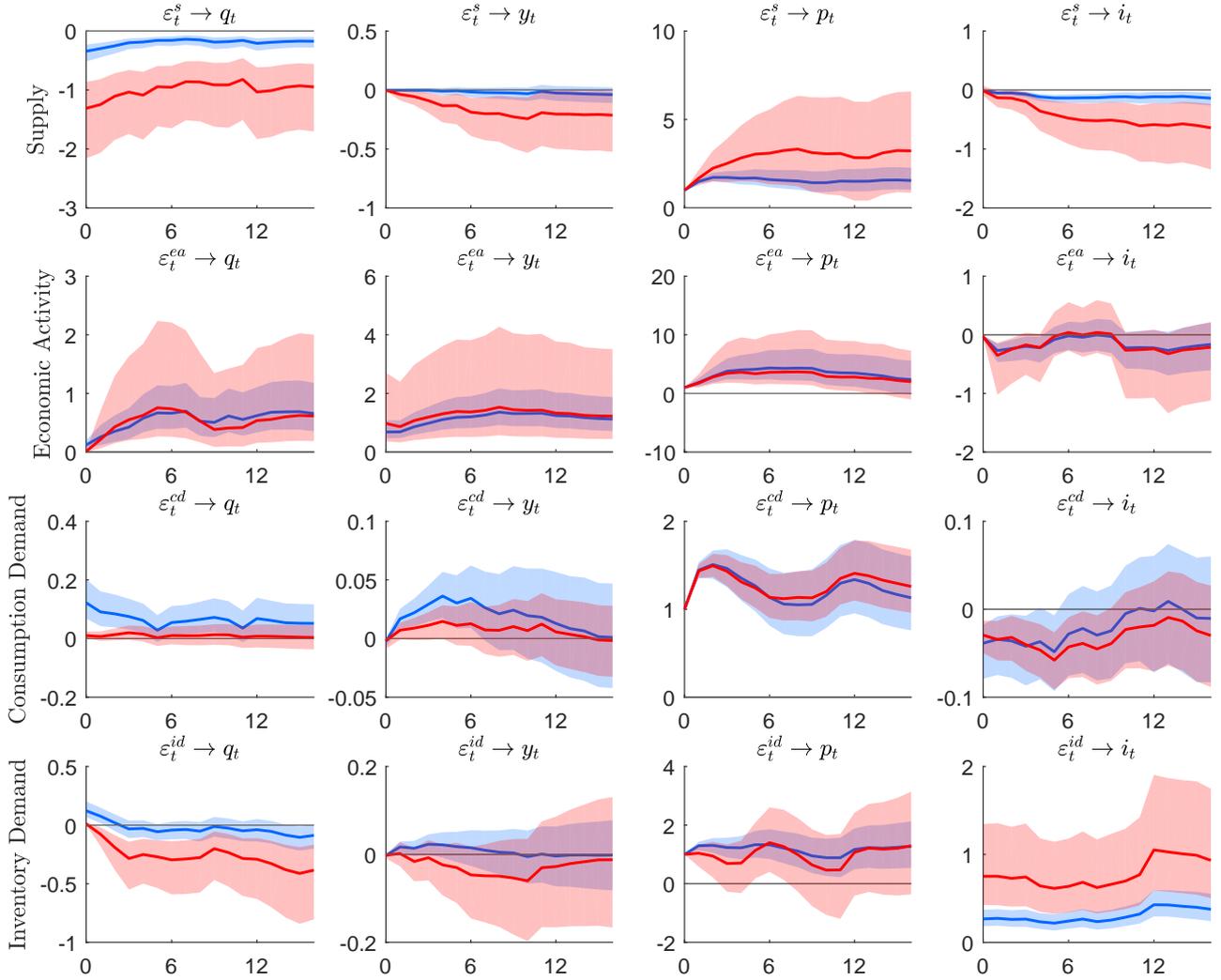


Figure 5: Posterior median IRFs with 90% credible intervals (shaded areas). Blue: Gaussian model. Red: non-Gaussian model.

Differences in the posterior distribution of key structural parameters have direct implications for structural analysis. In Figure 5, I provide (point-wise) posterior medians and 90% credible sets for impulse response functions (IRFs) up to 16 months, each standardized to increase oil prices by 1% on impact.⁴ The IRFs track the dynamic response of structural innovations on the level of the four endogenous variables. First, consider the effects of the oil supply shock (first row). In the non-Gaussian model, a much larger disruption in supply is required to achieve a price increase of the same magnitude. In turn, this leads to a considerably stronger response of

⁴For an alternative approach involving joint inference on impulse response functions see Inoue & Kilian (2021).

Table 2: Forecast Error Variance Decomposition (FEVD) of the real price of oil

horizon	ε_t^s	ε_t^{ea}	ε_t^{cd}	ε_t^{id}	ε_t^{me}
Gaussian Model					
4	0.32 (0.17, 0.51)	0.04 (0.02, 0.08)	0.58 (0.38, 0.74)	0.04 (0.02, 0.1)	0.01 (0, 0.02)
16	0.29 (0.16, 0.47)	0.06 (0.03, 0.09)	0.55 (0.36, 0.68)	0.04 (0.02, 0.1)	0.06 (0.03, 0.09)
Non-Gaussian Model					
4	0.05 (0.02, 0.1)	0.03 (0.01, 0.08)	0.89 (0.81, 0.94)	0.02 (0, 0.04)	0.01 (0, 0.02)
16	0.06 (0.03, 0.11)	0.04 (0.02, 0.09)	0.82 (0.74, 0.88)	0.03 (0.01, 0.07)	0.03 (0.01, 0.07)

The table gives posterior median estimates of the contribution of each shock to the forecast error variance of the real oil price at 4 and 16 months horizon. Values in brackets indicate corresponding 90% posterior credibility sets.

global economic activity and draw-down of inventories compared to the Gaussian specification. The opposite can be found for the consumption demand shock. Here, the estimated increase in oil produced is considerably more muted in the non-Gaussian model. No significant difference can be observed across the identification schemes for the response of global economic activity and oil inventories. In line with the literature, economic activity may slightly increase while inventories are drawn-up to mitigate some of the price increase.

IRFs to an Economic Activity (EA) shock are virtually indistinguishable across both identification approaches (BH19 and BH19+NG). An EA shock that increases oil prices by 1% is associated with a slowly increasing production, increase in global activity and decrease in inventories. With respect to the inventory demand shock, some subtle differences are found. First, note that the response of world activity, oil price and inventory are quite similar across specifications. For both models, oil prices and inventories display a positive co-movement, while global activity is barely affected. However, the impact response of global oil production differs. While in BH19, oil production is estimated to increase for a few months before gradually decreasing, the impact in the non-Gaussian model is virtually zero and is estimated to decrease afterwards. Hence, in BH19+NG, the shock behaves similar to the oil supply news shock discussed in Känzig (2021). These shocks reflect an anticipated decrease in oil production, which is associated with a sudden precautionary build-up of oil inventories and strong increase in oil prices. Note, however, that the effect on oil prices is more muted than documented in Känzig (2021).

Table 2 contains the forecast error variance decomposition of the real price of oil at 4 and 16 months horizon. Once more, the main difference across the two identification approaches is found along the effects of supply- and consumption demand shocks (highlighted in bold). As for the supply shocks, they are found to be much more important in the Gaussian model than

the non-Gaussian model. In particular, posterior median estimates indicate that in the BH19 model, supply shock explain around one third of the variance observed in real oil prices, with 90% credible sets covering anything between 16% and 51%. On the contrary, if non-Gaussianity is exploited as additional identification device, posterior median estimates suggest that supply shocks explain only a very small fraction of oil price movements, with median estimates at just 6%. In this identification scheme, posterior credibility sets are substantially more narrow and indicate that supply shocks are unlikely to explain more than 11% of the variation. Hence, estimates within the BH19+NG model are more in line with previously reported values (Kilian & Murphy; 2012, 2014). My findings suggest that even with a relatively diffuse prior for the oil supply elasticity, one may end up with similar results if non-Gaussianity is exploited. As for demand shocks, the opposite effect can be documented. Here, 90% posterior credible sets suggest that in the Gaussian model, consumption demand shocks explain between 36% and 74% of the variation. This contrasts sharply with much larger estimates associated with the non-Gaussian model. Specifically, posterior credible sets cover values between 74% and 94%.

3.3 Robustness

It is fair to say that for many experts in the oil market literature, the posterior of the short-run price elasticity of demand (β_{qp}) obtained under non-Gaussianity seems too large in absolute terms. Typically, values below -0.8 have been considered unreasonable in various papers, a value that corresponds to estimates of the long-run elasticity of demand (see for example Hausman & Newey (1995)). In this case, the prior used in Table 1 is not appropriate, since it attaches considerable mass to large negative values. A priori, the truncated t prior of table 1 implies $p(\beta_{qp} < -0.8) \approx 0.03$. Also, the low degrees of freedom in the student- t prior is not overly informative at the tails. To address this concern, I follow Kilian & Murphy (2012) and Baumeister & Peersman (2013) and restrict β_{qp} from below at -0.8 , effectively truncating the prior to lie on the interval $(-0.8, 0)$. To maintain comparability, both Gaussian and non-Gaussian model are re-estimated under the double truncated prior for β_{qp} . Under the label R1, posterior estimates for α_{qp} , β_{qp} and the contribution of ε_t^s to the FEVD of the real price of oil are reported in Table 3. First, note that the estimates for the Gaussian model are not affected under the alternative prior. This is not a surprise given that the bulk of the posterior mass of β_{qp} already lies above -0.8 in the baseline results (see Figure 4). In the non-Gaussian model, 90% posterior credibility sets of β_{qp} are now between -0.6 and -0.79 reflecting the additional constraint. Other than that, imposing the alternative prior does not materially affect the posterior of α_{qp} nor estimates of the supply shocks contribution to the forecast error variance of the oil price. While estimates are slightly higher than reported under the baseline results, supply shocks still play a minor role. Point estimates suggest that only 10% of price variation is driven by supply shocks.

In a second robustness check, I assess the sensitivity of the results to the large outliers observed in crude oil production in the earlier part of the sample (see Figure C.13 for a time series plot of the data). Therefore, I re-estimate the model using data from January 1985 to

December 2019. Corresponding results are labeled as R2 in Table 3. Under the shorter sample, the posterior of the Gaussian model point towards smaller elasticities (in absolute terms) of supply and demand, which resembles findings of Aastveit et al. (2021). A similar pattern applies to the results obtained by the non-Gaussian model. While the posterior for α_{qp} is still very close to zero, 90% credibility sets of β_{qp} lie between -0.19 and -0.44 . This is much lower than observed in the full sample, and confirm earlier findings in the literature regarding a possible break around 1985 (Baumeister & Peersman; 2013). Regarding the contribution of ε_t^s to the FEVD of the real price, point estimates of around 18% obtained under the non-Gaussian model are considerable higher than in the baseline specification. However, the same pattern is observed in that once non-Gaussianity is introduced into the model, supply shocks become less important than in the Gaussian model.

Table 3: Robustness analysis for the main empirical findings.

	Gaussian			Non-Gaussian		
Panel A: Posterior α_{qp}						
	5%	50%	95%	5%	50%	95%
R1	0.07	0.12	0.20	0.01	0.02	0.04
R2	0.03	0.07	0.13	0.01	0.03	0.05
Panel A: Posterior β_{qp}						
	5%	50%	95%	5%	50%	95%
R1	-0.50	-0.32	-0.20	-0.79	-0.74	-0.60
R2	-0.28	-0.17	-0.10	-0.44	-0.29	-0.19
Panel C: Contribution of ε_t^s to the FEVD of the real price of oil						
	Gaussian		Non-Gaussian			
	$h = 4$	$h = 16$	$h = 4$	$h = 16$		
R1	0.31 (0.17, 0.5)	0.29 (0.16, 0.46)	0.09 (0.06, 0.15)	0.1 (0.07, 0.16)		
R2	0.32 (0.16, 0.54)	0.29 (0.16, 0.47)	0.18 (0.1, 0.29)	0.18 (0.1, 0.28)		

For robustness check R1, the model is re-estimated based on a student-t prior of β_{qy} truncated on the interval $(-0.8, 0)$. For robustness check R2, the model is re-estimated based on a shortened sample covering January 1985 to December 2019. In panel C, the values in brackets give 90% posterior confidence sets.

4 Conclusion

In this paper, new evidence is provided on the relative importance of supply and demand shocks for oil price fluctuations. To identify their effects, non-Gaussianity is exploited in addition to a set of sign restrictions and weakly informative prior distributions spelled out directly on structural parameters (Baumeister & Hamilton; 2019). The empirical findings indicate that once the statistical properties of oil-market shocks are exploited for identification, oil supply

shocks become minor drivers of oil prices. The results are able to confirm estimates obtained previously in the literature (Kilian; 2009; Kilian & Murphy; 2012, 2014; Antolín-Díaz & Rubio-Ramírez; 2018), however, without the need of very strong identifying restrictions on underlying structural parameters.

From an econometric point of view, this paper offers a novel Bayesian estimator for non-Gaussian SVAR models. Specifically, each structural shock's marginal density is modeled non-parametrically using Bayesian infinite mixture models. The benefit from pursuing a non-parametric approach is that one ends up with a fully automatic procedure, requiring no prior knowledge on the form of non-Gaussianity a priori. The flexible density estimators are perfectly able to exploit deviations from normality at any region of the sample space, and hence capture excess kurtosis, skewness or other type of non-Gaussianity often documented in structural shocks.

References

- Aastveit, K. A., Bjørnland, H. C. & Cross, J. L. (2021). Inflation Expectations and the Pass-Through of Oil Prices, *The Review of Economics and Statistics* pp. 1–26.
- Antolín-Díaz, J. & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs, *American Economic Review* **108**(10): 2802–29.
- Baumeister, C. & Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information, *Econometrica* **83**(5): 1963–1999.
- Baumeister, C. & Hamilton, J. D. (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.
- Baumeister, C. & Peersman, G. (2013). The role of time-varying price elasticities in accounting for volatility changes in the crude oil market, *Journal of Applied Econometrics* **28**(7): 1087–1109.
- Bernanke, B. S. (1986). Alternative explanations of the money-income correlation, *Carnegie-Rochester Conference Series on Public Policy*, Vol. 25, Elsevier, pp. 49–99.
- Bertsche, D. & Braun, R. (2020). Identification of structural vector autoregressions by stochastic volatility, *Journal of Business & Economic Statistics* pp. 1–14.
- Bjørnland, H. C., Nordvik, F. M. & Rohrer, M. (2021). Supply flexibility in the shale patch: Evidence from north dakota, *Journal of Applied Econometrics* **36**(3): 273–292.
- Blackwell, D. & MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes, *The Annals of Statistics* **1**(2): 353–355.

- Blanchard, O. J. & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *The American Economic Review* **79**(4): 655–673.
- Boscolo, R., Pan, H. & Roychowdhury, V. P. (2004). Independent component analysis based on nonparametric density estimation, *IEEE Transactions on Neural Networks* **15**(1): 55–65.
- Caldara, D., Cavallo, M. & Iacoviello, M. (2019). Oil price elasticities and oil price fluctuations, *Journal of Monetary Economics* **103**: 1–20.
- Carriero, A., Marcellino, M. & Tornese, T. (2021). Blended identification in structural vars, *Technical report*, Working Paper.
- Chan, J., Koop, G. & Yu, X. (2021). Large order-invariant bayesian vars with stochastic volatility, *Technical report*, Working Paper.
- Cross, J., Nguyen, B. H. & Tran, T. D. (2020). The role of precautionary and speculative demand in the global market for crude oil.
- Drautzburg, T. & Wright, J. H. (2021). Refining set-identification in vars through independence, *Technical report*, National Bureau of Economic Research.
- Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association* **90**(430): 577–588.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics* **1**(2): 209–230.
- Fiorentini, G. & Sentana, E. (2020). *Discrete Mixtures of Normals Pseudo Maximum Likelihood Estimators of Structural Vector Autoregressions*, Centro de estudios monetarios y financieros.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds., Bayesian Statistics 4*: 169–193.
- Ghosal, S., Ghosh, J. K., Ramamoorthi, R. et al. (1999). Posterior consistency of Dirichlet mixtures in density estimation, *Ann. Statist* **27**(1): 143–158.
- Ghosh, J. K. & Ramamoorthi, R. (2003). *Bayesian nonparametrics*, Springer Science & Business Media.
- Gourieroux, C., Monfort, A. & Renne, J. P. (2017). Statistical inference for independent component analysis: Application to structural VAR models, *Journal of Econometrics* **196**(1): 111–126.
- Hausman, J. A. & Newey, W. K. (1995). Nonparametric estimation of exact consumers surplus and deadweight loss, *Econometrica: Journal of the Econometric Society* pp. 1445–1476.

- Herrera, A. M. & Rangaraju, S. K. (2020). The effect of oil supply shocks on us economic activity: What have we learned?, *Journal of Applied Econometrics* **35**(2): 141–159.
- Herwartz, H. (2018). Hodges–Lehmann detection of structural shocks: An analysis of macroeconomic dynamics in the euro area, *Oxford Bulletin of Economics and Statistics* .
- Herwartz, H. & Plödt, M. (2016). The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach, *Journal of International Money and Finance* **61**: 30–44.
- Inoue, A. & Kilian, L. (2021). Joint Bayesian inference about impulse responses in VAR models, *Journal of Econometrics* .
- Juvenal, L. & Petrella, I. (2015). Speculation in the oil market, *Journal of Applied Econometrics* **30**(4): 621–649.
- Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements, *American Economic Review* **111**(4): 1092–1125.
- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, *American Economic Review* **99**(3): 1053–1069.
- Kilian, L. (2020a). Facts and fiction in oil market modeling, *Technical report*, Federal Reserve Bank of Dallas.
- Kilian, L. (2020b). Understanding the estimation of oil demand and oil supply elasticities, *Technical report*, Federal Reserve Bank of Dallas.
- Kilian, L. & Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.
- Kilian, L. & Murphy, D. P. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models, *Journal of the European Economic Association* **10**(5): 1166–1188.
- Kilian, L. & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil, *Journal of Applied Econometrics* **29**(3): 454–478.
- Lanne, M. & Luoto, J. (2019). GMM estimation of non-gaussian structural vector autoregression, *Journal of Business & Economic Statistics* **0**(0): 1–13.
- Lanne, M. & Luoto, J. (2020). Identification of economic shocks by inequality constraints in bayesian structural vector autoregression, *Oxford Bulletin of Economics and Statistics* **82**(2): 425–452.
- Lanne, M. & Lütkepohl, H. (2010). Structural vector autoregressions with nonnormal residuals, *Journal of Business & Economic Statistics* **28**(1): 159–168.

- Lanne, M., Meitz, M. & Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions, *Journal of Econometrics* **196**(2): 288–304.
- Lanne, M. & Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model, *Journal of Business & Economic Statistics* **25**(1): 61–75.
- Lewis, D. J. (2021). Identifying Shocks via Time-Varying Volatility, *The Review of Economic Studies* .
- Litterman, R. (1986). Forecasting with Bayesian vector autoregressions - five years of experience, *Journal of Business & Economic Statistics* **4**(1): 25–38.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer Science & Business Media.
- Lütkepohl, H. & Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs, *Journal of Applied Econometrics* **29**(3): 479–496.
- Mertens, K. & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**(4): 1212–1247.
- Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models, *Journal of computational and graphical statistics* **9**(2): 249–265.
- Newell, R. G. & Prest, B. C. (2019). The unconventional oil supply boom: Aggregate price response from microdata, *The Energy Journal* **40**(3).
- Normandin, M. & Phaneuf, L. (2004). Monetary policy shocks: Testing identification conditions under time-varying conditional volatility, *Journal of Monetary Economics* **51**(6): 1217–1243.
- Rigobon, R. (2003). Identification through heteroskedasticity, *Review of Economics and Statistics* **85**(4): 777–792.
- Stock, J. H. & Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* **43**(1): 81–156.
- Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**: 381–419.
- Villani, M. (2009). Steady-state priors for vector autoregressions, *Journal of Applied Econometrics* **24**(4): 630–650.
- Waggoner, D. F. & Zha, T. (2003). A Gibbs sampler for structural vector autoregressions, *Journal of Economic Dynamics and Control* **28**(2): 349–366.

Yang, M., Dunson, D. B. & Baird, D. (2010). Semiparametric Bayes hierarchical models with mean and variance constraints, *Computational Statistics & Data Analysis* **54**(9): 2172–2186.

Zhou, X. (2020). Refining the workhorse oil market model, *Journal of Applied Econometrics* **35**(1): 130–140.

Appendix A Bayesian inference

A.1 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 $\alpha_+ = \text{vec}(A_+)$, $A_{i\bullet}$ the i -th 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 $i = 1, \dots, K$ and $t = 1, \dots, T$:

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

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

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

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

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

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

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

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

$$\tau_i \sim i\mathcal{G}(a_\tau/2, b_\tau/2), \tag{A.9}$$

$$m_i \sim \mathcal{N}(m_m, V_m), \tag{A.10}$$

where $x_t = [y'_{t-1}, \dots, y'_{t-p}, 1]'$ and $A_+ = [A_1, A_2, \dots, A_p, c]$.

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 parameter by $\xi = \{\varphi, \Theta\}$, and denote by ξ_{-x} all parameter 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 conditionals distributions of subsets in ξ . The algorithm involves the following steps:

1. For $i = 1, \dots, K$, draw from the mixture parameters θ_{it} , $t = 1, \dots, T$. To achieve better mixing properties of the Markov Chain, this step is performed using Algorithm 2 of Neal (2000). Neal further splits the mixing parameters into two components: $\theta_{it} = \theta_{i,cit}^*$, where

c_{it} are latent discrete assignment variables and θ_{ij}^* are unique cluster parameters. The algorithm involves drawing from two separate conditionals:

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

$$P(c_{it} = c_{ij}, j = 1, \dots, k_i | c_{i,-t}, \varepsilon_t) = b \frac{n_{-t, c_{ij}}}{T - 1 + \alpha_i} F(\varepsilon_{it} | \theta_{i, c_{ij}}^*),$$

$$P(c_{it} \neq c_{ij} \text{ for all } j \neq t | c_{i,-t}, \varepsilon_t) = b \frac{\alpha}{T - 1 + \alpha_i} \int F(\varepsilon_{it} | \theta) dG_0(\theta),$$

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. Given the conjugate Base distribution G_0 , the integral $\int F(\varepsilon_{it} | \theta) dG_0(\theta)$ is tractable and given in closed form. Hence, drawing from the distribution is straightforward.

- (b) The second conditional is that of the (active) cluster parameters $p(\theta_{ij}^* | Y, \xi_{-\theta_{ij}^*}), j = 1, \dots, k_i$, which are given by:

$$\sigma_{ij}^{*2} \sim i\mathcal{G}(\bar{a}_{ij}, \bar{b}_{ij})$$

$$\mu_{ij}^* \sim \mathcal{N}(\bar{m}_{ij}, \sigma_{ij}^{*2} \bar{V}_{ij})$$

with moments defined as follows:

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

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

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

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

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

$$p(d_i | \alpha_i) \sim \text{Beta}(\alpha_i + 1, T),$$

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

$$+ (1 - \pi_{d_i}) \mathcal{G}(a_\alpha + k_i - 1, b_\alpha - \log(d_i)),$$

where π_{d_i} is defined as:

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

- (b) Draw $p(m_i|Y, \xi_{-m_i}) \sim \mathcal{N}(\bar{m}_{m,i}, \bar{V}_{m,i})$ where $\bar{V}_{m,i} = \tau_i x_{\sigma_i^*} V_{\sigma_i^*}$, $\bar{m}_{m,i} = (1 - x_{\sigma_i^*})m_m + x_{\sigma_i^*} V_{\sigma_i^*} \left(\sum_{j=1}^{k_i} \sigma_{ij}^{*-2} \mu_{ij}^* \right)$ for $V_{\sigma_i^*}^{-1} = \sum_{j=1}^{k_i} \sigma_{ij}^{*-2}$, and $x_{\sigma_i^*} = V_m / (m_m + \tau_i V_{\sigma_i^*})$
- (c) Draw $p(\tau_i|Y, \xi_{-\tau_i}) \sim \mathcal{G}(\bar{a}_{\tau,i}, \bar{b}_{\tau,i})$ where $\bar{a}_{\tau,i} = \frac{a_\tau + k_i}{2}$ and $\bar{b}_{\tau,i} = \frac{b_\tau + \sum_{j=1}^{k_i} (\mu_{ij}^* - m_i) / \sigma_{ij}^{*2}}{2}$.

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 that the prior was uniform, that is $p^*(a_i) \propto c$. Let $U = [u_1 : \dots : u_T]'$ for $u_t = y_t - A_+ x_t$, $\mu_i = [\mu_{i1}, \dots, \mu_{iT}]'$ and $\Sigma_i = \text{diag}([\sigma_{i1}^2, \dots, \sigma_{iT}^2])$. Then, the conditional posterior is proportional to:

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

where $\Omega_{a_i}^{-1} = T^{-1} W_i' U \Sigma_i^{-1} U W_i$, $\mu_{a_i} = (W_i' U \Sigma_i^{-1} U W_i)^{-1} W_i' Y' (\mu_i - U w_i)$. Chan et al. (2021) derive an efficient way to sample from $p^*(a_i|Y, \xi_{-a_i})$ for $w_i = 0$, which builds on previous work of Waggoner & Zha (2003) and Villani (2009). In the following, I generalize the sampling scheme for w_i containing non-zero elements. Hereby, I closely follow the exposition and notation of Villani (2009):

Definition 1. A random variable X follows the generalized absolute normal distribution $GAN(a, b, \mu, \rho)$ if it has density function:

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

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 i th column deleted, B_\perp the orthogonal complement of B , and $\text{chol}(B)$ the Choleski decomposition of B such that $\text{chol}(B)\text{chol}(B)' = B$. Also, denote by $\|\cdot\|$ the Euclidean norm and $\stackrel{d}{=}$ equality in distribution.

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

$$a_i \stackrel{d}{=} R_i \sum_{j=1}^{r_i} \gamma_j v_j, \tag{A.11}$$

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 $j = 2, \dots, r_i$, $\hat{\gamma}_j = \mu'_{a_i} R_i'^{-1} v_j$, $v_1 = R_i W_i'(A)_{-i\perp} / \|R_i W_i'(A)_{-i\perp}\|$, $(v_2, \dots, v_{r_i}) = v_{1\perp}$, $\hat{a} = \det([A'_{1\bullet}, \dots, w_i, \dots, A'_{K\bullet}])$ and $\hat{b} = \det([A'_{1\bullet}, \dots, W_i R_i v_1, \dots, A'_{K\bullet}])$.

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

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

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

$$\begin{aligned} |A| &= \det \left[A'_{1\bullet} | \cdots | w_i + W_i R_i \sum_{j=1}^{r_i} \gamma_j v_j | \cdots | A'_{K\bullet} \right] \\ &= \det [A'_{1\bullet} | \cdots | w_i | \cdots | A'_{K\bullet}] + \sum_{j=1}^{r_i} \gamma_j \det \left[A'_{1\bullet} | \cdots | W_i R_i \sum_{j=1}^{r_i} v_j | \cdots | A'_{K\bullet} \right] \\ &= \underbrace{\det [A'_{1\bullet} | \cdots | w_i | \cdots | A'_{K\bullet}]}_{\hat{a}} + \underbrace{\det \left[A'_{1\bullet} | \cdots | W_i R_i \sum_{j=1}^{r_i} v_j | \cdots | A'_{K\bullet} \right]}_{\hat{b}} \gamma_1 \end{aligned}$$

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

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

□

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

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

Corresponding curvature is given by:

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

Hence the following normal approximation:

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

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 $w = \frac{p_{GAN}(\mu_1; a, b, \mu, \rho)}{\sum_{j=1}^2 p_{GAN}(\mu_j; a, b, \mu, \rho)}$ 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.

In any case, such a step is necessary when working with a more general prior for $p(a_i)$ than the uniform used to derive $p^*(a_i|Y, \xi_{a_i})$. In most cases, it will suffice to use a Metropolis Hastings step that corrects for the fact that $p^*(a_i|Y, \xi_{a_i})$ is missing the information from a non-uniform prior $p(a_i)$. Denote by $a_i^{(l-1)}$ the current state of the Markov chain and by $a'_i \sim p^*(a_i|Y, \xi_{a_i})$ the proposed value under a uniform prior. Then, the MH algorithm proceeds setting $a_i^{(l)} = a'_i$ with probability $\alpha_{MH} = \min \left\{ 1, \frac{p(a'_i)}{p(a_i^{(l-1)})} \right\}$. If the proposed draw is not accepted, $a_i^{(l)} = a_i^{(l-1)}$.⁵

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

$$p(\alpha_+ | Y, \xi_{-\alpha_+}) \sim \mathcal{N}(\bar{\mu}_A, \bar{V}_A), \quad (\text{A.12})$$

where

$$\bar{V}_{\alpha_+} = \left(V_{\alpha_+}^{-1} + \sum_{t=1}^T (x_t \otimes I_K) (A' \Sigma_t^{-1} A) (x_t' \otimes I_K) \right)^{-1}, \quad (\text{A.13})$$

$$\bar{\mu}_{\alpha_+} = \bar{V}_{\alpha_+} \left(V_{\alpha_+}^{-1} m_{\alpha_+} + \sum_{t=1}^T (x_t \otimes I_K) (A' \Sigma_t^{-1} A) \tilde{y}_t \right), \quad (\text{A.14})$$

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

A.2 Adjustments to inference 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*} , the forecast error of the scaled up oil inventories, is an unobserved latent variable. To get around this problem, I include u_t^{i*} into the set of latent variables and infer it from the data within the MCMC algorithm. Specifically, the forth block is altered as to draw from $p(\alpha_+, u^{i*} | Y, \xi_{\{-\alpha_+, -u^{i*}\}})$, where $u^{i*} = [u_1^{i*}, \dots, u_T^{i*}]'$. Specifically, I

⁵The 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.

make use of the possibility to marginalize over u^{i^*} when sampling α_+ . The adjusted forth block of the MCMC algorithm draws from:

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

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

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

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 $\tilde{\Sigma}_t = \text{diag}([v'_t, \sigma_5^2]')$. Manipulating the equation, the reduced form can be obtained:

$$\begin{pmatrix} y_t \\ u_t^{i^*} \end{pmatrix} = \begin{pmatrix} A_+ x_t \\ 0 \end{pmatrix} + A^{-1} \tilde{\mu}_t + \tilde{\eta}_t, \quad \varepsilon_t \sim \mathcal{N}(0, A^{-1} \tilde{\Sigma}_t A^{-1'}), \quad (\text{A.15})$$

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

$$p(Y | \alpha_+, \xi_{\{-\alpha_+, -u^{i^*}\}}) \propto |\Omega_t|^{-T/2} \exp \left(\sum_{t=1}^T (\tilde{y}_t - A_+ x_t)' \Omega_t^{-1} (\tilde{y}_t - A_+ x_t) \right) \quad (\text{A.16})$$

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

$$\bar{V}_A = \left(V_A^{-1} + \sum_{t=1}^T (x_t \otimes I_K) \Omega_t^{-1} (x_t' \otimes I_K) \right)^{-1}, \quad (\text{A.17})$$

$$\bar{\mu}_A = \bar{V}_A \left(V_A^{-1} m_{\alpha_+} + \sum_{t=1}^T (x_t \otimes I_K) \Omega_t^{-1} \tilde{y}_t \right), \quad (\text{A.18})$$

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

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

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

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

A.3 Markov chain Monte Carlo algorithm for the Gaussian oil market model

The MCMC algorithm for the Gaussian model is derived in the following. First, note the model which is given as follows:

$$\tilde{A} \tilde{u}_t = \varepsilon_t, \quad \varepsilon_t \sim (0, I_{K+1}),$$

where $\tilde{A} = A \Sigma_\varepsilon^{-1/2}$, $\tilde{u}_t = [u_t', u_t^{i*}]'$ and $u_t = y_t - A_+ x_t$. Let γ_A be the underlying parameters of \tilde{A} with prior distribution $p(\gamma_A)$. As previously, the autoregressive parameters $\alpha_+ = \text{vec}(A_+)$ are given a normal prior $\alpha_+ \sim \mathcal{N}(m_{\alpha_+}, V_{\alpha_+})$. Then, the MCMC algorithm generates random draws from the following posterior distribution:

$$p(\gamma_A, \alpha_+, u^{i*} | Y) = \underbrace{p(\gamma_A | Y)}_{\text{unknown} \rightarrow \text{RW-MH}} \underbrace{p(\alpha_+ | \gamma_A, Y)}_{\text{normal}} \underbrace{p(u^{i*} | \gamma_A, \alpha_+, Y)}_{\text{normal}}.$$

As in Baumeister & Hamilton (2019), draws from $p(\gamma_A | Y)$ are generated via an Random Walk Metropolis Hastings algorithm, while conditional on these draws, $p(\alpha_+ | \gamma_A, Y)$ and then $p(u^{i*} | \gamma_A, \alpha_+, Y)$ are simple normal distributions and readily available.

To run the Random Walk MH algorithm, it is necessary to evaluate $p(\gamma_A | Y)$. This is derived in the following. First, note the likelihood is given as a special case of equation (A.16):

$$p(Y | \gamma_A, \alpha_+) \propto |\Omega|^{-T/2} \exp \left(\sum_{t=1}^T (y_t - A_+ x_t)' \Omega^{-1} (y_t - A_+ x_t) \right).$$

for $\Omega = J \tilde{A}^{-1} \tilde{A}^{-1'} J'$ is a function of γ_A . Based on Bayes theorem, the marginal posterior $p(Y | \gamma_A)$ can be obtained as follows:

$$p(\gamma_A | Y) \propto p(\gamma_A) \times p(Y | \gamma_A) \tag{A.19}$$

$$\propto p(\gamma_A) \times \frac{p(Y | \gamma_A, \alpha_+) p(\alpha_+)}{p(\alpha_+ | \gamma_A, Y)}, \tag{A.20}$$

which can be evaluated at any α_+ . The missing ingredient is $p(\alpha_+|\gamma_A, Y)$, the conditional posterior of α_+ . As in the non-Gaussian model, this is a normal $p(\alpha_+|\gamma_A, Y) \sim \mathcal{N}(\bar{\mu}_{\alpha_+}, \bar{V}_{\alpha_+})$ with moments given as:

$$\bar{V}_{\alpha_+} = \left(V_{\alpha_+}^{-1} + \sum_{t=1}^T (x_t \otimes I_K) \Omega^{-1} (x_t' \otimes I_K) \right)^{-1}, \quad (\text{A.21})$$

$$\bar{\mu}_{\alpha_+} = \bar{V}_{\alpha_+} \left(V_{\alpha_+}^{-1} m_{\alpha_+} + \sum_{t=1}^T (x_t \otimes I_K) \Omega^{-1} y_t \right), \quad (\text{A.22})$$

In practice, I evaluate the right hand side of (A.20) at $\alpha_+ = \bar{\mu}_A$, yielding

$$p(\gamma_A|Y) \propto p(\gamma_A) \frac{|\Omega|^{-T/2} |V_{\alpha_+}|^{-1/2}}{|\bar{V}_A|^{-1/2}} \exp \left(-0.5 \sum_{t=1}^T \hat{u}_t' \Omega^{-1} \hat{u}_t - 0.5 (\bar{\mu}_A - \mu_A)' V_{\alpha_+}^{-1} (\bar{\mu}_A - \mu_A) \right).$$

for $\hat{u}_t = y_t - (x_t' \otimes I_K) \bar{\mu}_A$. For large systems of equations or VAR models with many lags, substantial computational gains can be obtained by using a prior of the form $p(\alpha_+|\gamma_A) \sim \mathcal{N}(\mu_A, \Omega \otimes V_{\alpha_+})$, circumventing the need of inverting a $K(Kp+1)$ square matrix \bar{V}_A^{-1} to obtain $\bar{\mu}_A$.

Finally, $p(u_t^{i*}|\gamma_A, \alpha_+, Y)$ is derived as in Appendix A.2. That is, defining

$$\tilde{A}^{-1} \tilde{A}^{-1'} = \tilde{\Omega} = \begin{pmatrix} \tilde{\Omega}_{11} & \tilde{\Omega}_{12} \\ \tilde{\Omega}_{21} & \tilde{\Omega}_{22} \end{pmatrix}$$

the conditional posterior is normal given as follows:

$$\begin{aligned} p(u_t^{i*}|\gamma_A, \alpha_+, Y) &\sim \mathcal{N}(\bar{u}_t^{i*}, \bar{V}_{u_t^{i*}}), \\ \bar{u}_t^{i*} &= \tilde{\Omega}_{21} \tilde{\Omega}_{11}^{-1} (y_t - A_+ x_t), \\ \bar{V}_{u_t^{i*}} &= \tilde{\Omega}_{22} - \tilde{\Omega}_{21} \tilde{\Omega}_{11}^{-1} \tilde{\Omega}_{12}. \end{aligned}$$

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 13 seconds using a standard *i5* Laptop processor.⁶ To contrast the results to those of a Gaussian model, the model is also estimated using the methodology of Baumeister & 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 B.6 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 capture well the strong non-Gaussian shape in the data. Particularly the second shock has strong outliers leading to very heavy tails.

Second, Figure B.7 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 same efficiency than drawing iid 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 suggest that similar to the algorithm of Baumeister & Hamilton (2015), one should consider a relatively large Markov chain of 100000 to obtain comparably precise results of at least 1000 iid draws.

Finally, Figure B.8 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 towards the true value of α_{qp} , and further narrows down the value of β_{qp} .

⁶For the computations in this paper, a Intel(R) Core(TM) i5-6300U CPU with 2.40GHz was used.

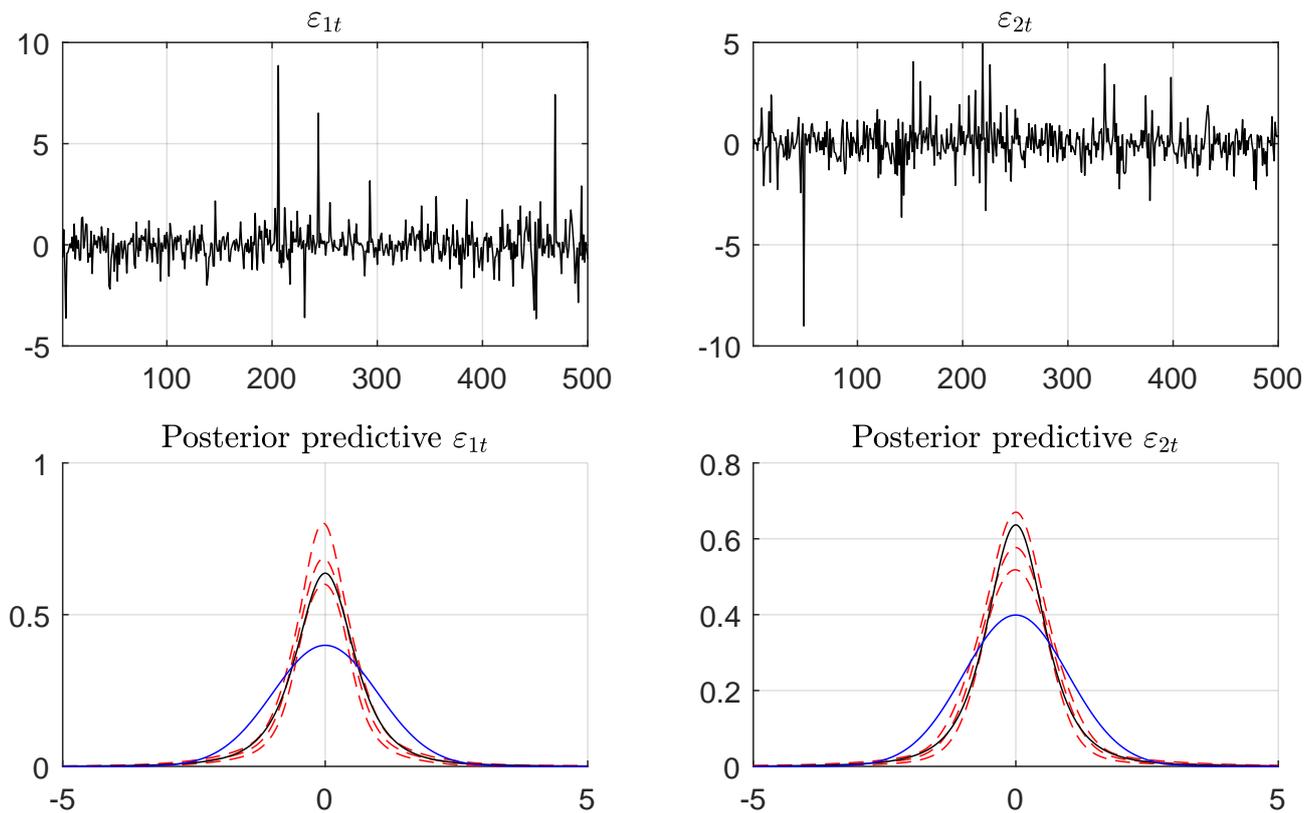


Figure B.6: Simulated structural shocks (top panel) and estimated posterior predictive densities under the non-Gaussian model. Red dashed lines indicate 90% posterior credibility sets, the black line that of a unit variance standardized t_3 distribution, and the blue line gives the standard normal density.

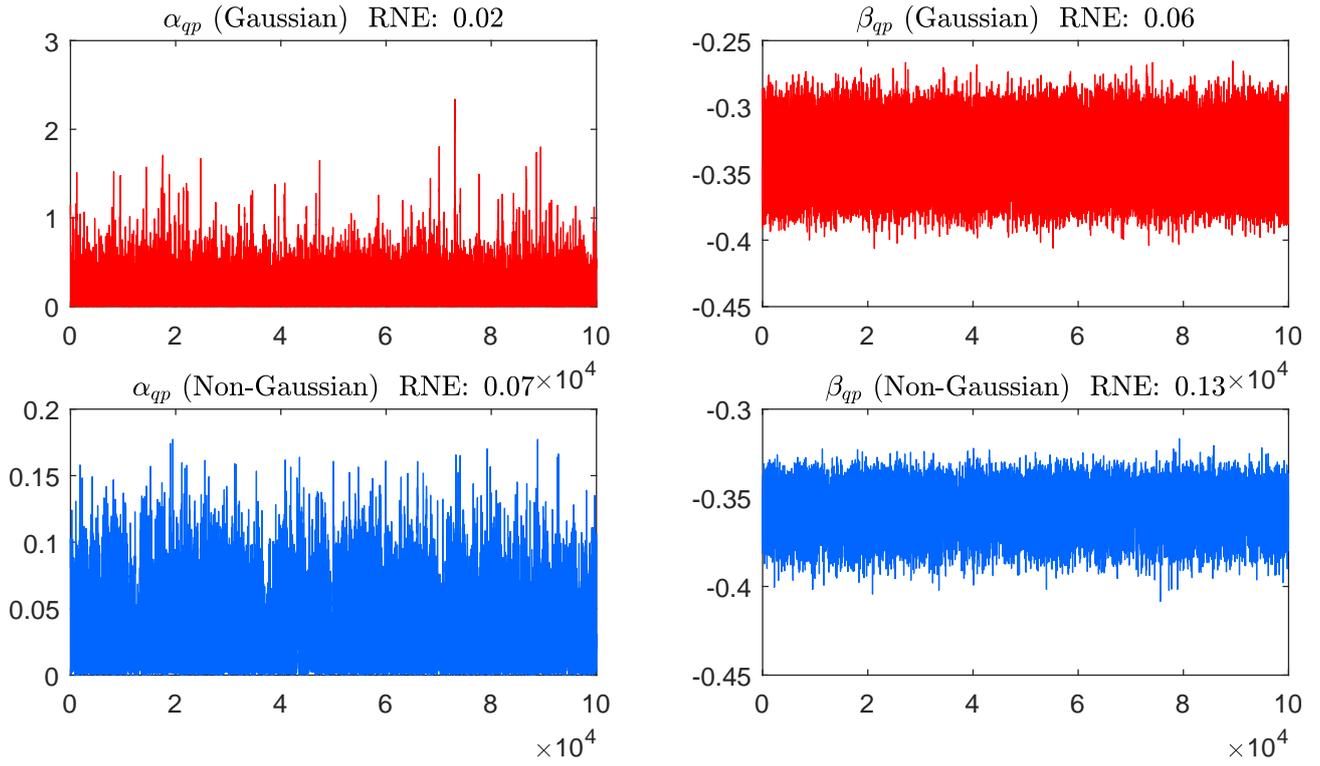


Figure B.7: Markov Chain Monte Carlo output of length 100'000. Top panel: Gaussian model with MCMC as in Baumeister & Hamilton (2015). Bottom panel: MCMC of non-Gaussian model as described in Appendix A.1.

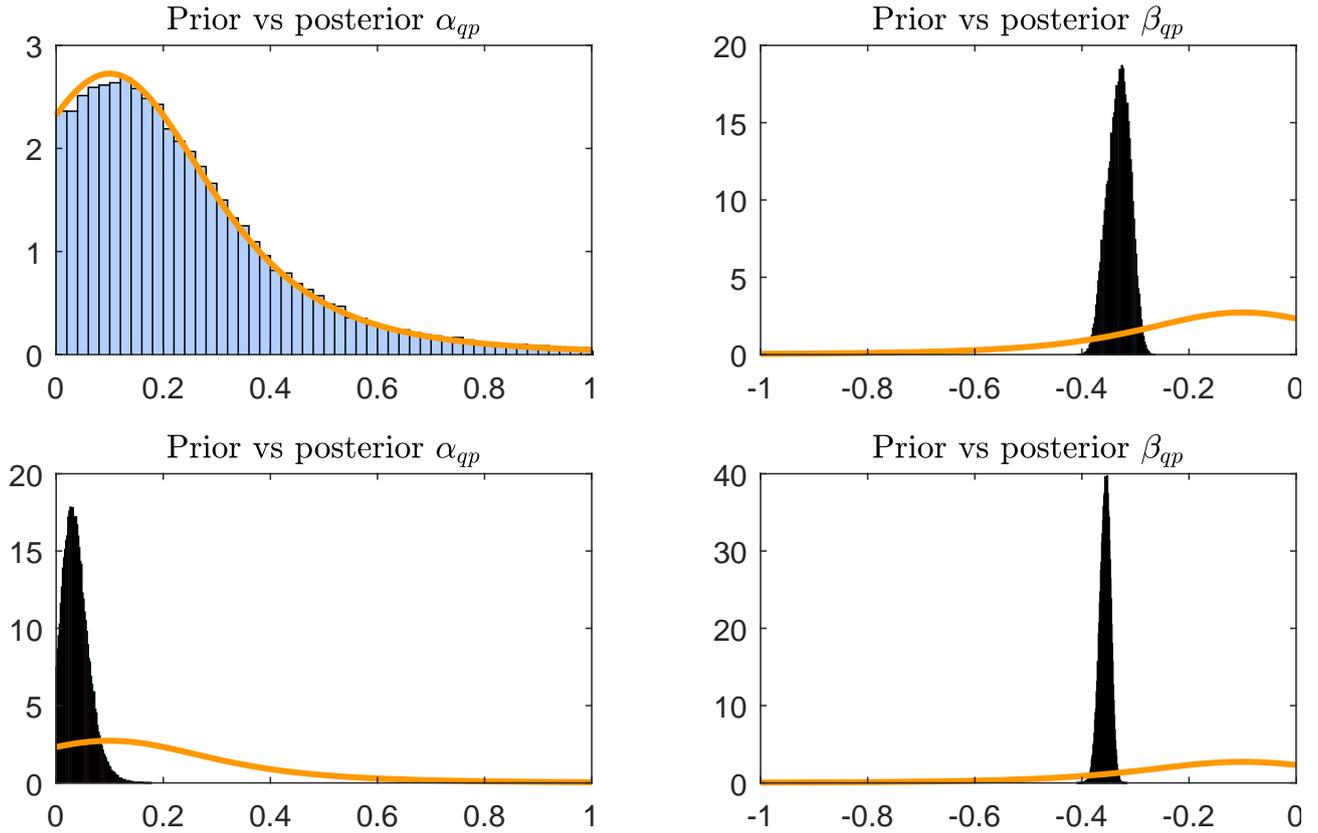


Figure B.8: Prior (orange line) and posterior density of the two structural parameters α_{qp} and β_{qp} . Top panel: Gaussian model. Bottom panel: non-Gaussian model.

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 B.9, 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 B.10 suggests no apparent problem with the MCMC. However, the RNE values deteriorates somewhat, which is to be expected for Gibbs sampler type MCMC algorithms under weak identification. Finally, Figure B.11 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.

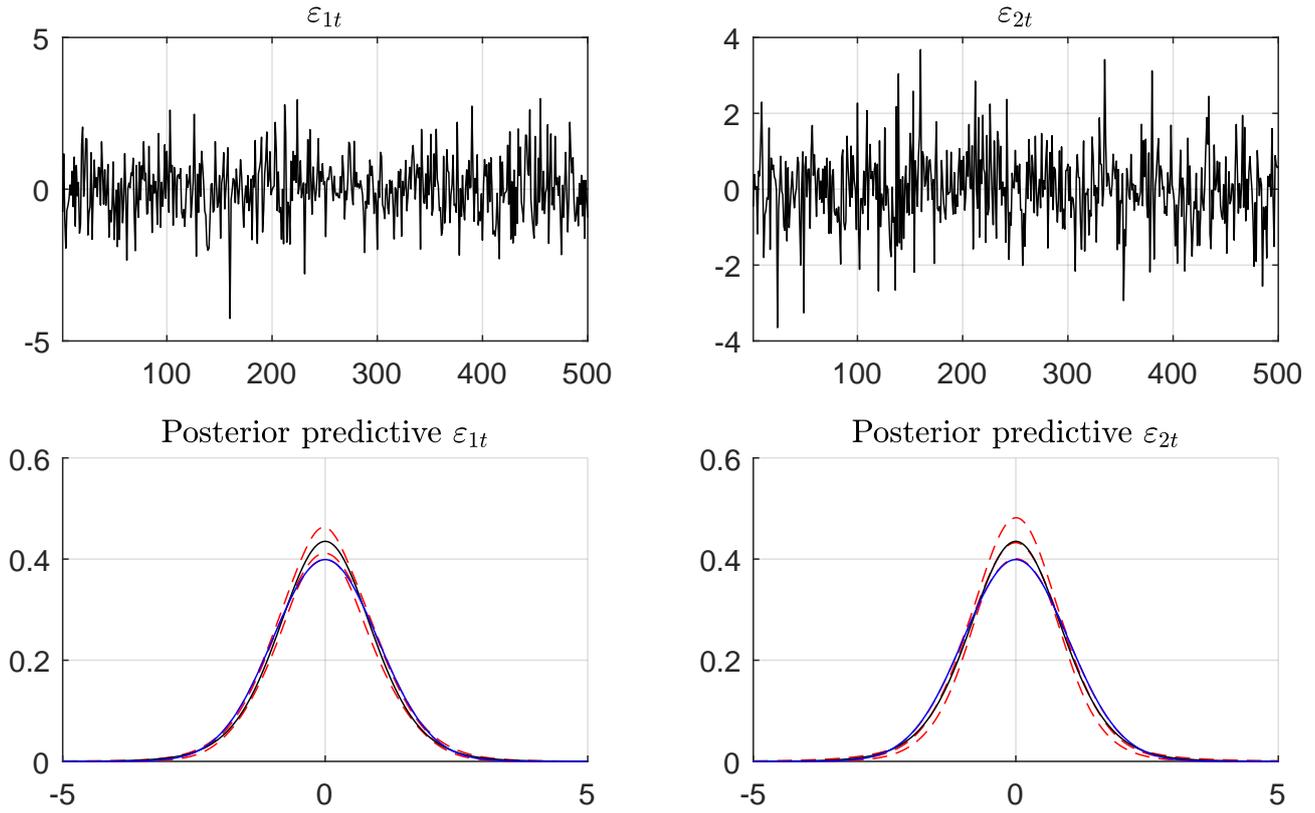


Figure B.9: Simulated structural shocks (top panel) and estimated posterior predictive densities under the non-Gaussian model. Red dashed lines indicate 90% posterior credibility sets, the black line that of a unit variance standardized t_{10} distribution, and the blue line gives the standard normal density.

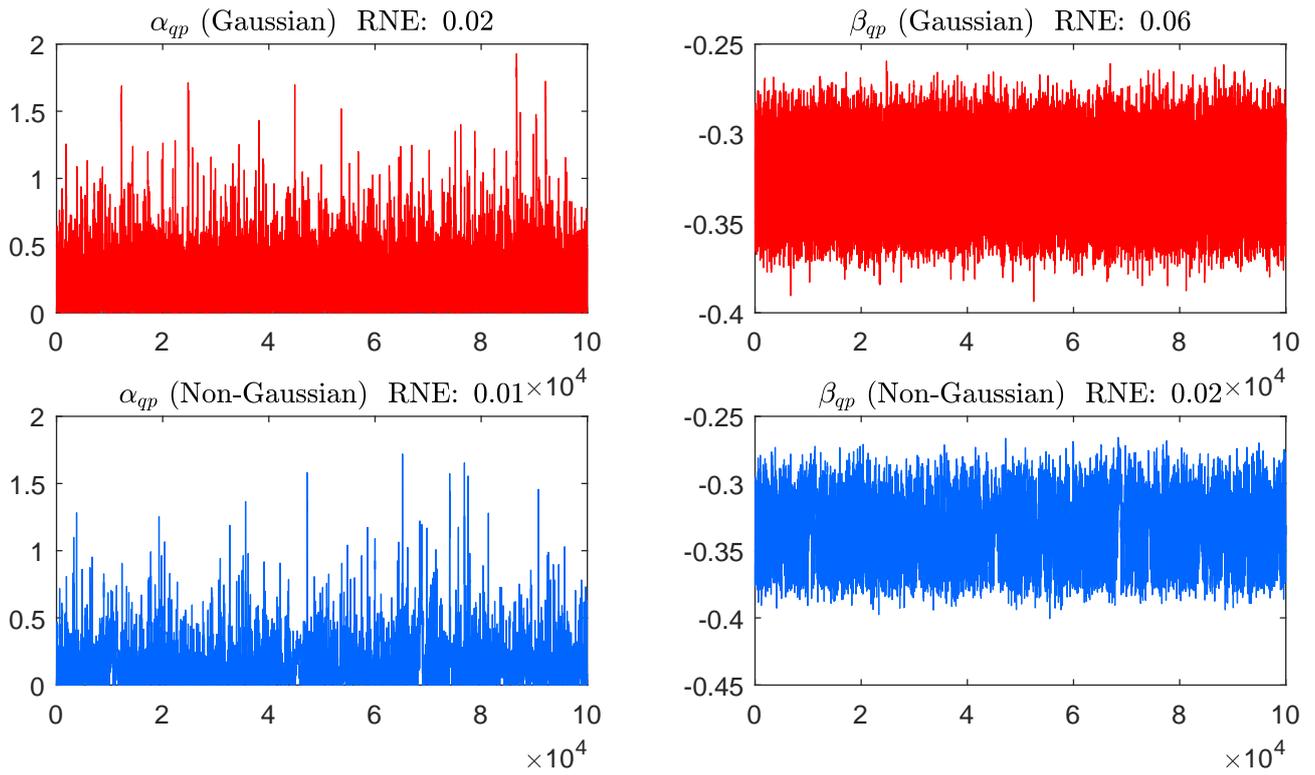


Figure B.10: Markov Chain Monte Carlo output of length 100'000. Top panel: Gaussian model with MCMC as in Baumeister & Hamilton (2015). Bottom panel: MCMC of non-Gaussian model as described in Appendix A.1.

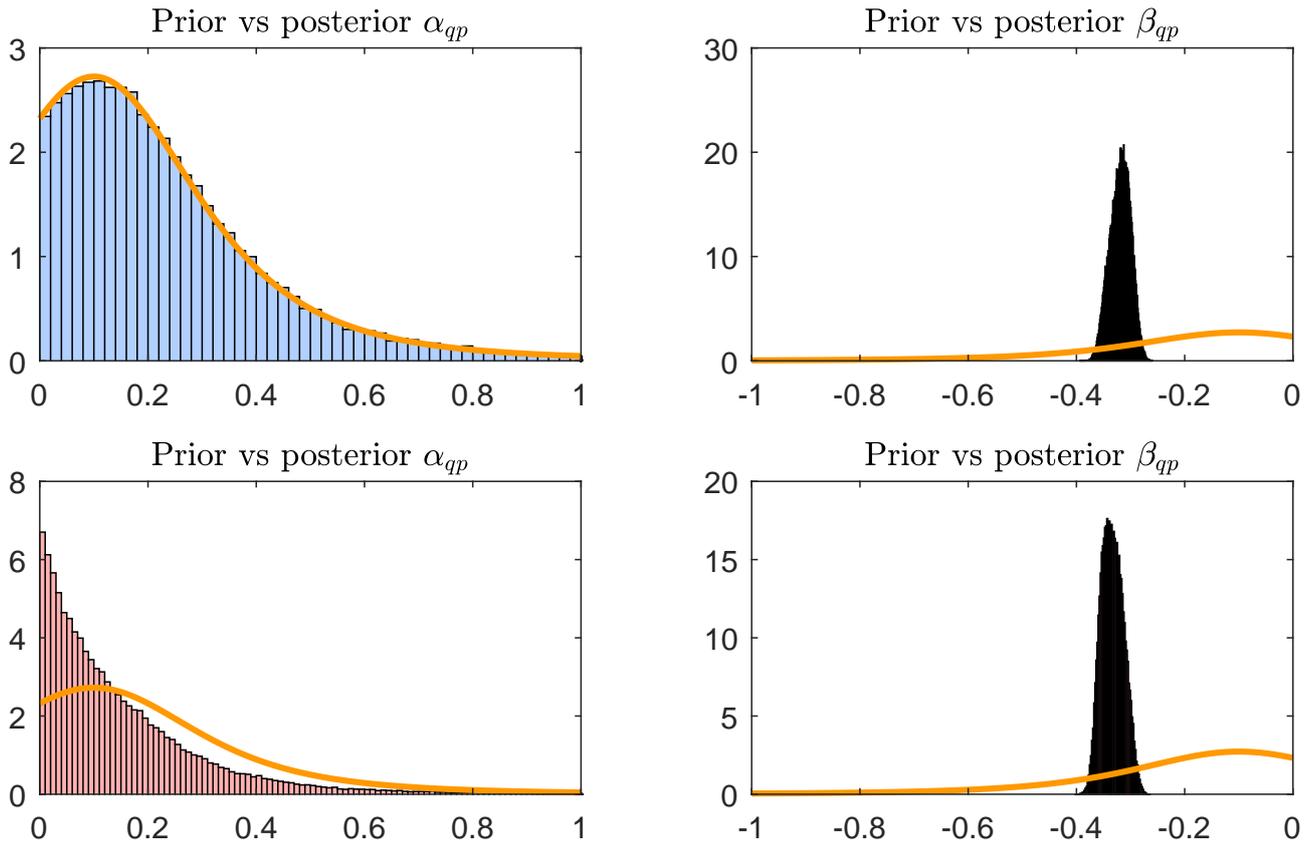


Figure B.11: Prior (orange line) and posterior density of the two structural parameters α_{qp} and β_{qp} . Top panel: Gaussian model. Bottom panel: non-Gaussian model.

B.3 Empirical application

As a last exercise, Figure B.12 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 suggest a fairly high autocorrelation in the draws justifying the use of very long Markov Chain.

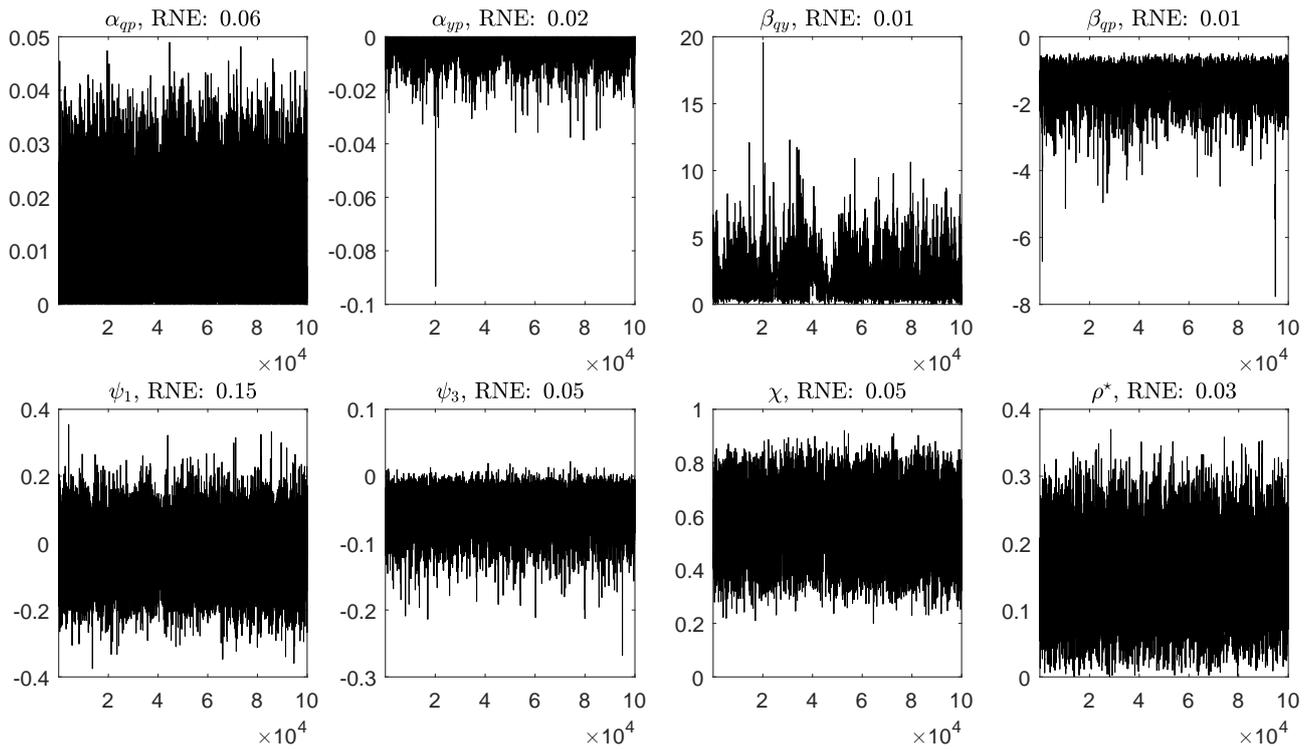


Figure B.12: Markov Chain Monte Carlo output of each element of A obtained under the non-Gaussian model of section 3.

Appendix C Supplementary Figures

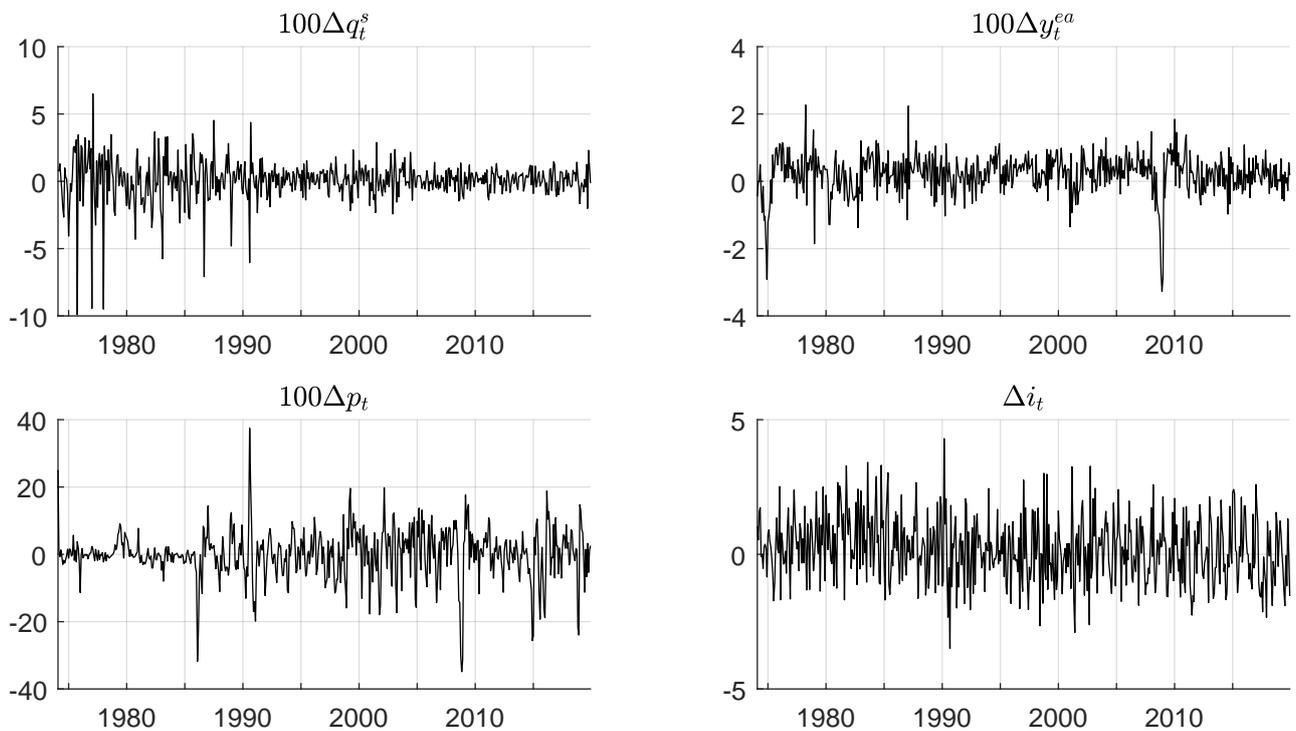


Figure C.13: Oil market dataset

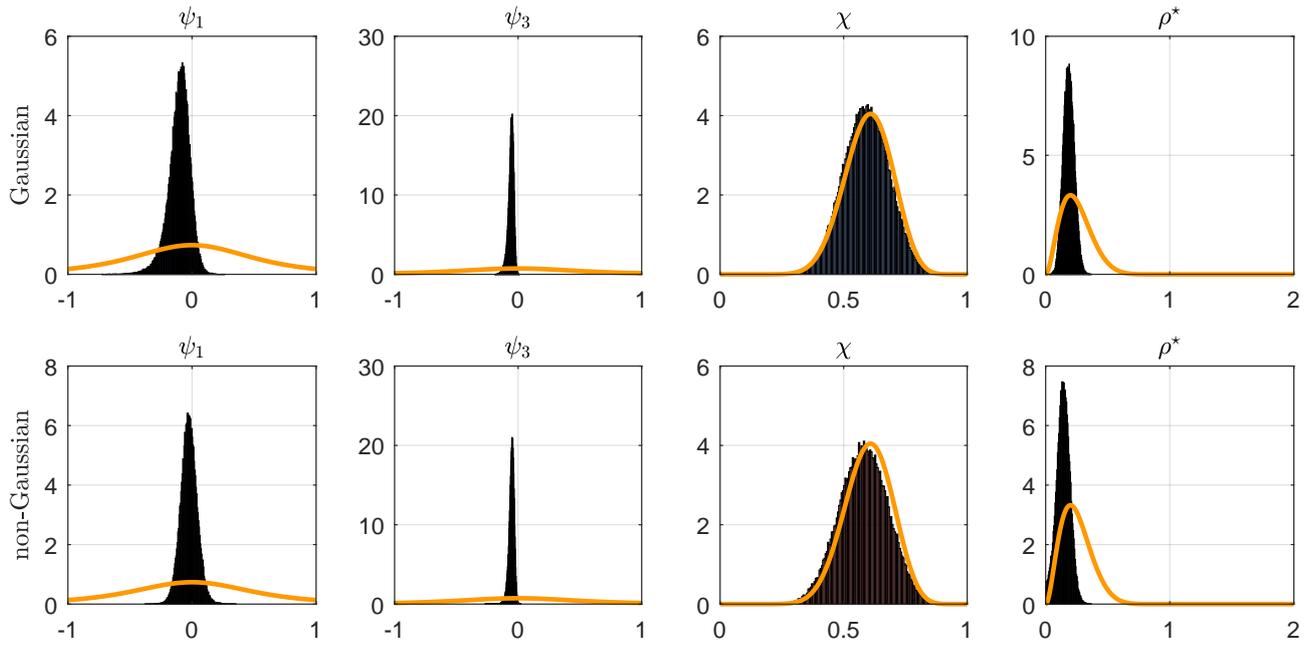


Figure C.14: Prior (orange line) and posterior density of the remaining structural parameters. Top panel: Gaussian model. Bottom panel: non-Gaussian model.