Online Appendix for Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks

Christiane Baumeister

James D. Hamilton

Appendix D. Details of Bayesian algorithm.

For any numerical value of $\mathbf{A}$ we can calculate $\zeta_i^*(\mathbf{A})$ and $\tau_i^*(\mathbf{A})$ using equations (7) and (6) from which we can calculate the log of the target

$$q(\mathbf{A}) = \log(p(\mathbf{A})) + (T/2)\log\left[\det\left(\mathbf{A}\hat{\boldsymbol{\Omega}}_T\mathbf{A}'\right)\right] \tag{D1}$$
$$- \sum_{i=1}^{n}\kappa_i^*\log[(2/T)\tau_i^*(\mathbf{A})] + \sum_{i=1}^{n}\kappa_i\log\tau_i(\mathbf{A}).$$

We can improve the efficiency of the algorithm by using information about the shape of this function calculated as follows. Collect elements of $\mathbf{A}$ that are not known with certainty in an $(n_\alpha \times 1)$ vector $\boldsymbol{\alpha}$, and find the value $\hat{\boldsymbol{\alpha}}$ that maximizes (D1) numerically. This value $\hat{\boldsymbol{\alpha}}$ offers a reasonable guess for the posterior mean of $\boldsymbol{\alpha}$, while the matrix of second derivatives (again obtained numerically) gives an idea of the curvature of the posterior distribution:

$$\hat{\boldsymbol{\Lambda}} = -\left.\frac{\partial^2 q(\mathbf{A}(\boldsymbol{\alpha}))}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'}\right|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}}.$$

We then use this guess to inform a random-walk Metropolis Hastings algorithm to generate candidate draws of $\boldsymbol{\alpha}$ from the posterior distribution, as follows. As a result of step ℓ we have generated a value of $\boldsymbol{\alpha}^{(\ell)}$. For step $\ell + 1$ we generate

$$\tilde{\boldsymbol{\alpha}}^{(\ell+1)} = \boldsymbol{\alpha}^{(\ell)} + \xi\left(\hat{\mathbf{Q}}^{-1}\right)'\mathbf{v}_t$$

for $\mathbf{v}_t$ an $(n_\alpha \times 1)$ vector of Student t variables with 2 degrees of freedom, $\hat{\mathbf{Q}}$ the Cholesky factor of $\hat{\boldsymbol{\Lambda}}$ (namely $\hat{\mathbf{Q}}\hat{\mathbf{Q}}' = \hat{\boldsymbol{\Lambda}}$ with $\hat{\mathbf{Q}}$ lower triangular), and ξ a tuning scalar to be described shortly. If $q(\mathbf{A}(\tilde{\boldsymbol{\alpha}}^{(\ell+1)})) < q(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)}))$, we set $\boldsymbol{\alpha}^{(\ell+1)} = \boldsymbol{\alpha}^{(\ell)}$ with probability $1 - \exp\left[q(\mathbf{A}(\tilde{\boldsymbol{\alpha}}^{(\ell+1)})) - q(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)}))\right]$; otherwise, we set $\boldsymbol{\alpha}^{(\ell+1)} = \tilde{\boldsymbol{\alpha}}^{(\ell+1)}$. The parameter ξ is chosen so that about 30 percent of the newly generated $\tilde{\boldsymbol{\alpha}}^{(\ell+1)}$ get retained. The algorithm can be started by setting $\boldsymbol{\alpha}^{(1)} = \hat{\boldsymbol{\alpha}}$, and the values after the first D burn-in draws $\{\boldsymbol{\alpha}^{(D+1)}, \boldsymbol{\alpha}^{(D+2)}, ..., \boldsymbol{\alpha}^{(D+N)}\}$ represent a sample of size N drawn from the posterior distribution $p(\boldsymbol{\alpha}|\mathbf{Y}_T)$; in our applications we have used $D = N = 10^6$.

For each of these N final values for $\boldsymbol{\alpha}^{(\ell)}$ we further generate $\delta_{ii}^{(\ell)} \sim \Gamma(\kappa_i^*, \tau_i^*(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)})))$ for

$i = 1, ..., n$ and take $\mathbf{D}^{(\ell)}$ to be a diagonal matrix whose row i, column i element is given by $1/\delta_{ii}^{(\ell)}$. From these we also generate $\mathbf{b}_i^{(\ell)} \sim N(\mathbf{m}_i^*(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)})), d_{ii}^{(\ell)}\mathbf{M}_i^*)$ for $i = 1, ..., n$ and take $\mathbf{B}^{(\ell)}$ the matrix whose ith row is given by $\mathbf{b}_i^{(\ell)\prime}$. The triple $\{\mathbf{A}(\boldsymbol{\alpha}^{(\ell)}), \mathbf{D}^{(\ell)}, \mathbf{B}^{(\ell)}\}_{\ell=D+1}^{D+N}$ then represents a sample of size N drawn from the posterior distribution $p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T)$.

Appendix E. Data sources.

The data sets used in the original studies by Kilian (2009) and Kilian and Murphy (2012) are available from the public data archives of the *Journal of the European Economic Association* (http://onlinelibrary.wiley.com/doi/10.1111/j.1542-4774.2012.01080.x/suppinfo) and we used these exact same data for the statistical analysis reported in Sections II.A and II.B. We also reconstructed these data sets from the original sources ourselves as part of the process of assembling extended time series as described below.

Monthly world oil production data measured in thousands of barrels of oil per day were obtained from the U.S. Energy Information Administration's (EIA) *Monthly Energy Review* for the period January 1973 to December 2016. Monthly data for global production of crude oil for the period 1958:M1 to 1972:M12 were collected from the weekly *Oil and Gas Journal* (issue of the first week of each month) as in Baumeister and Peersman (2013b).

The nominal spot oil price for West Texas Intermediate (WTI) was retrieved from the Federal Reserve Economic Data (FRED) database maintained by the St. Louis Fed (OILPRICE). Prior to 1982 this equals the posted price. This series was discontinued in July 2013. From August 2013 onwards data are obtained from the EIA website (http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=rwtc&f=m). To deflate the nominal spot oil price, we use the U.S. consumer price index (CPIAUCSL: consumer price index for all urban consumers: all items, index 1982-1984 = 100) which was taken from the FRED database.

For the extended data set our measure for global economic activity is the industrial production index for OECD countries and six major non-member economies (Brazil, China, India, Indonesia, the Russian Federation and South Africa) obtained from the OECD Main Economic Indicators (MEI) database in 2011. The index covers the period 1958:M1 to 2011:M10 and was subsequently discontinued. To extend the data set after October 2011, we applied the same methodology used by the OCED. Specifically, we use OECD industrial production and industrial production for the individual non-member countries which are available in the MEI database and apply the weights reported by the OECD to aggregate those series into a single index. The source of the weights data is the International Monetary Fund's World Economic Outlook (WEO) database. The weights are updated on a yearly basis and a link to a document containing the weights can be found at http://www.oecd.org/std/compositeleadingindicatorsclifrequentlyaskedquestionsfaqs.htm#11.

Monthly U.S. crude oil stocks in millions of barrels (which include the Strategic Petro-

leum Reserve) are available from EIA for the entire period 1958:M1-2016:M12. We obtain an estimate for global stocks as in Kilian and Murphy (2012) by multiplying the U.S. crude oil inventories by the ratio of OECD inventories of crude petroleum and petroleum products to U.S. inventories of petroleum and petroleum products. Given that OECD petroleum inventories only start in January 1988, we assume that the ratio before January 1988 is the same as in January 1988. To calculate our proxy for Δi_t, the change in OECD inventories as a fraction of last period's oil production, we convert the production data into millions of barrels per month by multiplying the million barrels of crude oil produced per day by 30.

Appendix F. Adapting the algorithms in Baumeister and Hamilton (2015) to allow for measurement error.

Rewriting (34) in the form of (1). The variance matrix for the structural shocks in (34) is given by

$$\tilde{\mathbf{D}} = E(\tilde{\mathbf{u}}_t \tilde{\mathbf{u}}_t') = \begin{bmatrix} d_{11}^* & 0 & 0 & 0 \\ 0 & d_{22}^* & 0 & 0 \\ 0 & 0 & d_{33}^* + \chi^{-2}\sigma_e^2 & -\chi^{-1}\sigma_e^2 \\ 0 & 0 & -\chi^{-1}\sigma_e^2 & \chi^2 d_{44}^* + \sigma_e^2 \end{bmatrix}. \tag{F1}$$

It's not hard to see that $\mathbf{\Gamma}\tilde{\mathbf{D}}\mathbf{\Gamma}' = \mathbf{D}$ is diagonal for

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \rho & 1 \end{bmatrix} \tag{F2}$$

with ρ given by (37). Thus if we premultiply (34) by $\mathbf{\Gamma}$ we arrive at a system in the form of (1) for which $\mathbf{A} = \mathbf{\Gamma}\tilde{\mathbf{A}}$, $\mathbf{B} = \mathbf{\Gamma}\tilde{\mathbf{B}}$, and

$$\mathbf{u}_t = \mathbf{\Gamma}\tilde{\mathbf{u}}_t = \begin{bmatrix} u_{1t}^* \\ u_{2t}^* \\ u_{3t}^* - \chi^{-1}e_t \\ \chi u_{4t}^* + \rho u_{3t}^* + (1 - \rho/\chi)e_t \end{bmatrix} \tag{F3}$$

whose variance matrix we denote $\mathbf{D} = \text{diag}(d_{11}, d_{22}, d_{33}, d_{44})$. This is exactly in the form of the general class of models discussed in Section I with the elements of the matrix $\mathbf{A}$ determined by $\boldsymbol{\theta}_{\mathbf{A}} = (\alpha_{qp}, \alpha_{yp}, \beta_{qy}, \beta_{qp}, \chi, \psi_1, \psi_3, \rho)'$. Given a prior distribution for these parameters, we can then draw from the posterior distribution $p(\boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T)$ as described in Appendix D.

Given a draw from $p(\boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T)$ we immediately have a draw for $\tilde{\mathbf{A}} = \mathbf{\Gamma}^{-1}\mathbf{A}$, $\tilde{\mathbf{B}} = \mathbf{\Gamma}^{-1}\mathbf{B}$,

and $\tilde{\mathbf{D}} = \boldsymbol{\Gamma}^{-1}\mathbf{D}(\boldsymbol{\Gamma}^{-1})'$. The lower-right (2×2) block of the last equation is

$$
\begin{bmatrix} \tilde{d}_{33} & \tilde{d}_{34} \\ \tilde{d}_{43} & \tilde{d}_{44} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} d_{33} & 0 \\ 0 & d_{44} \end{bmatrix} \begin{bmatrix} 1 & -\rho \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} d_{33} & -\rho d_{33} \\ -\rho d_{33} & d_{44} + \rho^2 d_{33} \end{bmatrix}. \tag{F4}
$$

Historical decompositions. Recall $\boldsymbol{\epsilon}_t = \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{u}}_t$ for $\tilde{\mathbf{u}}_t$ defined in (36). Taking expectations of both sides conditional on the data and on a draw of the parameters, the following equation holds exactly for every t,

$$
\boldsymbol{\epsilon}_t = \tilde{\mathbf{A}}^{-1} \begin{bmatrix} u_{1t}^* \\ u_{2t}^* \\ E(u_{3t}^*|\tilde{\mathbf{u}}_t, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) - \chi^{-1}E(e_t|\tilde{\mathbf{u}}_t, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) \\ \chi E(u_{4t}^*|\tilde{\mathbf{u}}_t, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) + E(e_t|\tilde{\mathbf{u}}_t, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) \end{bmatrix}, \tag{F5}
$$

where

$$
\begin{bmatrix} E(u_{3t}^*|\tilde{\mathbf{u}}_t, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) \\ E(u_{4t}^*|\tilde{\mathbf{u}}_t, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) \\ E(e_t|\tilde{\mathbf{u}}_t, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) \end{bmatrix} = \begin{bmatrix} E(u_{3t}^*|\tilde{u}_{3t}, \tilde{u}_{4t}, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) \\ E(u_{4t}^*|\tilde{u}_{3t}, \tilde{u}_{4t}, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) \\ E(e_t|\tilde{u}_{3t}, \tilde{u}_{4t}, \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D}) \end{bmatrix}
$$

$$
= E\left(\begin{bmatrix} u_{3t}^*\tilde{u}_{3t} & u_{3t}^*\tilde{u}_{4t} \\ u_{4t}^*\tilde{u}_{3t} & u_{4t}^*\tilde{u}_{4t} \\ e_t\tilde{u}_{3t} & e_t\tilde{u}_{4t} \end{bmatrix} \middle| \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D} \right) \left\{ E\left(\begin{bmatrix} \tilde{u}_{3t}^2 & \tilde{u}_{3t}\tilde{u}_{4t} \\ \tilde{u}_{4t}\tilde{u}_{3t} & \tilde{u}_{4t}^2 \end{bmatrix} \middle| \boldsymbol{\theta}_{\mathbf{A}}, \mathbf{D} \right) \right\}^{-1} \begin{bmatrix} \tilde{u}_{3t} \\ \tilde{u}_{4t} \end{bmatrix}. \tag{F6}
$$

The second matrix in (F6) is known from (F4) while the first matrix can be calculated from the equations

$$
\tilde{u}_{3t} = u_{3t}^* - \chi^{-1}e_t
$$

$$
\tilde{u}_{4t} = \chi u_{4t}^* + e_t
$$

and the fact that the three disturbances u_{3t}^*, u_{4t}^*, and e_t are mutually uncorrelated:

$$
E(\tilde{u}_{3t}u_{4t}^*) = E(\tilde{u}_{4t}u_{3t}^*) = 0 \tag{F7}
$$

$$
E(\tilde{u}_{3t}e_t) = -\chi^{-1}\sigma_e^2 = E(\tilde{u}_{3t}\tilde{u}_{4t}) \tag{F8}
$$

$$
E(\tilde{u}_{4t}e_t) = \sigma_e^2 = -\chi E(\tilde{u}_{3t}\tilde{u}_{4t}) \tag{F9}
$$

$$
E(\tilde{u}_{3t}u_{3t}^*) = E(\tilde{u}_{3t})(\tilde{u}_{3t} + \chi^{-1}e_t) = E(\tilde{u}_{3t}^2) + \chi^{-1}E(\tilde{u}_{3t}\tilde{u}_{4t}) \tag{F10}
$$

$$
\chi E(\tilde{u}_{4t}u_{4t}^*) = E(\tilde{u}_{4t})(\tilde{u}_{4t} - e_t) = E(\tilde{u}_{4t}^2) + \chi E(\tilde{u}_{3t}\tilde{u}_{4t}). \tag{F11}
$$

Collecting (F7)-(F11) into a matrix equation,

$$
E\left(\left[\begin{array}{cc} u_{3t}^*\tilde{u}_{3t} & u_{3t}^*\tilde{u}_{4t} \\ u_{4t}^*\tilde{u}_{3t} & u_{4t}^*\tilde{u}_{4t} \\ e_t\tilde{u}_{3t} & e_t\tilde{u}_{4t} \end{array}\right] \middle| \boldsymbol{\theta_A}, \mathbf{D}\right) = \left[\begin{array}{cc} \tilde{d}_{33} + \chi^{-1}\tilde{d}_{34} & 0 \\ 0 & (\tilde{d}_{44} + \chi\tilde{d}_{34})/\chi \\ \tilde{d}_{34} & -\chi\tilde{d}_{34} \end{array}\right]
$$

$$
= \left[\begin{array}{cc} d_{33}(1 - \rho/\chi) & 0 \\ 0 & \frac{d_{44} + \rho(\rho - \chi)d_{33}}{\chi} \\ -\rho d_{33} & \rho\chi d_{33} \end{array}\right] \tag{F12}
$$

where the last equation follows from (F4). Substituting (F4) and (F12) into (F6),

$$
\left[\begin{array}{c} E(u_{3t}^*|\tilde{\mathbf{u}}_t, \boldsymbol{\theta_A}, \mathbf{D}) \\ E(u_{4t}^*|\tilde{\mathbf{u}}_t, \boldsymbol{\theta_A}, \mathbf{D}) \\ E(e_t|\tilde{\mathbf{u}}_t, \boldsymbol{\theta_A}, \mathbf{D}) \end{array}\right] = \left[\begin{array}{cc} d_{33}(1 - \frac{\rho}{\chi}) & 0 \\ 0 & \frac{d_{44} + \rho(\rho - \chi)d_{33}}{\chi} \\ -\rho d_{33} & \rho\chi d_{33} \end{array}\right]\left[\begin{array}{cc} d_{33} & -\rho d_{33} \\ -\rho d_{33} & d_{44} + \rho^2 d_{33} \end{array}\right]^{-1}\left[\begin{array}{c} \tilde{u}_{3t} \\ \tilde{u}_{4t} \end{array}\right]
$$

$$
= \left[\begin{array}{cc} h_{33} & h_{34} \\ h_{43} & h_{44} \\ h_{e3} & h_{e4} \end{array}\right]\left[\begin{array}{c} \tilde{u}_{3t} \\ \tilde{u}_{4t} \end{array}\right] \tag{F13}
$$

Historical decompositions can thus be calculated as follows. The value of $\mathbf{y}_t$ can be written as the r-period-ahead forecast plus a known function of the forecast errors between $t - r$ and t:

$$
\mathbf{y}_t = \hat{\mathbf{y}}_{t|t-r} + \sum_{s=0}^{r-1}\boldsymbol{\Psi}_s\boldsymbol{\epsilon}_{t-s}. \tag{F14}
$$

We have inferred values for $\tilde{\mathbf{u}}_t$ for each date from $\tilde{\mathbf{u}}_t = \tilde{\mathbf{A}}\mathbf{y}_t - \tilde{\mathbf{B}}\mathbf{x}_{t-1}$. Hence

$$
\mathbf{y}_t = \hat{\mathbf{y}}_{t|t-r} + \sum_{s=0}^{r-1}\boldsymbol{\Psi}_s\tilde{\mathbf{A}}^{-1}\left[\begin{array}{c} u_{1,t-s}^* \\ u_{2,t-s}^* \\ \hat{u}_{3,t-s}^* - \chi^{-1}\hat{e}_{t-s} \\ \chi\hat{u}_{4,t-s}^* + \hat{e}_{t-s} \end{array}\right]
$$

where $(u_{1t}^*, u_{2t}^*)' = (\tilde{u}_{1t}, \tilde{u}_{2t})'$ and the vector $(\hat{u}_{3t}^*, \hat{u}_{4t}^*, \hat{e}_t)'$ is calculated from (F13). With this expression we can calculate the contribution of each of the shocks $(u_1^*, u_2^*, u_3^*, u_4^*, e)$ to the historical value of $\mathbf{y}_t$. For example, the historical contribution of inventory demand shocks (u_{4t}^*) to $\mathbf{y}_t$ is found from $\sum_{s=0}^{r-1}\boldsymbol{\Psi}_s\tilde{\mathbf{A}}^{-1}\mathbf{h}_{t-s}^{4*}$ for $\mathbf{h}_t^{4*} = (0, 0, 0, \chi(h_{43}\tilde{u}_{3t} + h_{44}\tilde{u}_{4t}))'$, while the contribution of measurement error is

$$
\sum_{s=0}^{r-1}\boldsymbol{\Psi}_s\tilde{\mathbf{A}}^{-1}\mathbf{h}_{t-s}^e \tag{F15}
$$

for $\mathbf{h}_t^e = (0, 0, -\chi^{-1}(h_{e3}\tilde{u}_{3t} + h_{e4}\tilde{u}_{4t}), h_{e3}\tilde{u}_{3t} + h_{e4}\tilde{u}_{4t})'$.