Not-for-Publication Appendix to: The Role of Time-Varying Price Elasticities in Accounting for Volatility Changes in the Crude Oil Market

Christiane Baumeister
Bank of Canada
cbaumeister@bankofcanada.ca

Gert Peersman
Ghent University
gert.peersman@ugent.be

February 2012

This appendix provides a detailed description of the data used in the empirical analysis; of the setup of the econometric model, including the choice of priors and the estimation algorithm; and of the computation of impulse responses and the implementation of the sign restrictions. We demonstrate by conducting a Monte Carlo simulation exercise that our econometric model is able to capture abrupt changes in a satisfactory manner, but show that the data favor a drifting-coefficient model by incorporating a mixture innovation model into our benchmark specification. We then provide evidence on the significance of changes over time in the volatilities of the real price of crude oil and world oil production and in the short-run price elasticities of oil supply and oil demand. The main findings of the paper are shown to be robust to changes in the model specification and to different assumptions underlying the identification of the structural shocks. Finally, some additional results on the time-varying contribution of the structural shocks to the variances of the real price of oil and world oil production are presented.

1 Data description

World oil production data measured in thousands of barrels of oil per day starting in January 1973 were obtained from the Energy Information Administration's (EIA) *Monthly Energy Review*. Monthly data for global production of crude oil for the period $1953M4$ to $1972M12$ were taken from the weekly *Oil & Gas Journal* (issue of the first week of each month). For the period $1947M1$ to $1953M3$, monthly data were obtained by interpolation of yearly world oil production data by means of the Litterman (1983) methodology with U.S. monthly oil production from the EIA used as an

indicator variable.[1] Annual oil production data were retrieved from *World Petroleum* (1947−1954). Quarterly data are averages of monthly observations.

The nominal U.S. refiners' acquisition cost of imported crude oil was taken from the *Monthly Energy Review*.[2] Since this series is only available from January 1974, it was backcast until 1947Q1 with the quarterly growth rate of the producer price index (PPI) for crude oil obtained from the Bureau of Labor Statistics (BLS) database (WPU0561). Data were converted to quarterly frequency before backcasting by taking averages over months. Monthly seasonally adjusted data for the U.S. consumer price index (CPIAUCSL: consumer price index for all urban consumers: all items, index $1982 - 1984 = 100$) taken from the Federal Reserve Bank of St. Louis FRED database were used to deflate the nominal refiners' acquisition cost for crude oil imports.

The index of world industrial production was taken from the United Nations *Monthly Bulletin of Statistics* (MBS). The index numbers are reported on a quarterly basis and span the period 1947Q1 to 2008Q3. The index covers industrial activities in mining and quarrying; manufacturing; and electricity, gas and water supply. The index indicates trends in global value added in constant U.S. dollars. The measure of value added is the national accounts concept, which is defined as gross output less the cost of materials, supplies, fuel and electricity consumed and services received. Each series was compiled using the Laspeyres formula, i.e. indices are base-weighted arithmetic means. The production series of individual countries are weighted by their value-added contribution, generally measured at factor costs, to gross domestic product of the given industry during the base year. For most countries the estimates of value added used as weights are derived from the results of national industrial censuses (census of production) or similar inquiries. A new set of weights is introduced every five years to account for structural changes in the composition of production in each industry over time, and the index series are chain-linked by the technique of splicing at overlapping years. These data in national currencies were converted into U.S. dollars by means of official or free-market exchange rates. The weights were most recently updated in 2000, which is also the base year for the index ($2000 = 100$). The index was constructed to shift the whole series to this base. Since the majority of national indices have not been adjusted for fluctuations due to seasonal factors, we applied the census X12 ARIMA procedure to the reconstructed series in order to obtain a seasonally adjusted index for the entire period. Given that according to the August 2009 issue of the MBS the world industrial production index will not be updated until the implementation of ISIC Rev. 4, we extended the series from 2008Q3 to 2010Q4

[1] Since this part of the data is only needed for the training sample to initialize the priors based on the estimation of a fixed-coefficient VAR, the use of interpolated data as opposed to actual ones is of minor importance.

[2] The refiners' acquisition cost of crude oil imports (IRAC) is a volume-weighted average price of all kinds of crude oil imported into the U.S. over a specified period. Since the U.S. imports more types of crude oil than any other country, it may represent the best proxy for a true "world oil price" among all published crude oil prices. The IRAC is also similar to the OPEC basket price.

using the average quarterly growth rate of monthly industrial production of OECD countries and six major non-member economies obtained from the OECD's Main Economic Indicators (MEI) database. The index of global real economic activity in industrial commodity markets introduced in Kilian (2009) was retrieved for the period $1968M1 - 2011M4$ from Lutz Kilian's web page (http://www-personal.umich.edu/~lkilian/reaupdate.txt). This monthly series was converted to quarterly frequency by taking the average.

Daily data on the total number of WTI futures contracts (volume) traded on the New York Mercantile Exchange (NYMEX) are available from the commercial provider Price-data.com (End of Day Futures) from March 30, 1983 onwards. For each daily observation, the trading date and the contract's date of delivery are reported. For each futures contract, we calculated the number of business days to delivery and kept only those observations that were fewer than 260 business days (one year) to maturity. We sum the volume of all remaining observations recorded within the same quarter, regardless of the date of delivery, to obtain a measure of total quarterly trading activity in the oil futures market.

Annual data on global spare capacity of oil production for the period $1974 - 2010$ were taken from the August 2006 IMF *World Economic Outlook* and the June 2011 Department of Energy *Short-Term Energy Outlook*. Spare capacity refers to production capacity that can be brought online within 30 days and sustained for 90 days. Global capacity utilization rates were calculated as a percentage of total potential annual world oil production, which is the sum of actual oil production taken from the *Annual Energy Review* and available spare capacity. Monthly data on worldwide rig counts starting in January 1975 were obtained from Baker & Hughes Inc. (http://investor.shareholder.com/bhi/rig_counts/rc_index.cfm).

2 A Bayesian SVAR with time-varying parameters and stochastic volatility

Model setup. The observation equation of our state space model is

$$y_t = X_t'\theta_t + u_t \tag{1}$$

where y_t is a 3×1 vector of observations of the dependent variables, X_t is a matrix including lags ($p = 4$) of all of the dependent variables and a constant term, and θ_t is a $3(3p+1) \times 1$ vector of states that contains the time-varying parameters. The u_t of the measurement equation are heteroscedastic disturbance terms that follow a normal distribution with a zero mean and a time-varying covariance matrix Ω_t which can be decomposed in the following way: $\Omega_t = A_t^{-1} H_t \left(A_t^{-1}\right)'$. A_t is a lower triangular matrix that models the contemporaneous interactions among the endogenous variables,

and H_t is a diagonal matrix that contains the stochastic volatilities:

$$A_t = \begin{bmatrix} 1 & 0 & 0 \\ a_{21,t} & 1 & 0 \\ a_{31,t} & a_{32,t} & 1 \end{bmatrix} \qquad H_t = \begin{bmatrix} h_{1,t} & 0 & 0 \\ 0 & h_{2,t} & 0 \\ 0 & 0 & h_{3,t} \end{bmatrix} \tag{2}$$

Let α_t be the vector of non-zero and non-one elements of the matrix A_t (stacked by rows), $\alpha_t = [a_{21,t}, a_{31,t}, a_{32,t}]'$, and h_t be the vector containing the diagonal elements of H_t, $h_t = [h_{1,t}, h_{2,t}, h_{3,t}]'$. As in Primiceri (2005), the three driving processes of the system are postulated to evolve as follows:

$$\theta_t = \theta_{t-1} + \nu_t \qquad \nu_t \sim N(0, Q) \tag{3}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t \qquad \zeta_t \sim N(0, S) \tag{4}$$

$$\ln h_{i,t} = \ln h_{i,t-1} + \sigma_i \eta_{i,t} \qquad \eta_{i,t} \sim N(0, 1) \tag{5}$$

The time-varying parameters θ_t and α_t are modeled as driftless random walks.[3] The elements of the vector of volatilities h_t are assumed to evolve as geometric random walks independent of each other.[4] The error terms of the transition equations are independent of each other and of the innovations in the observation equation. In addition, we impose a block-diagonal structure for S of the following form:

$$S \equiv Var(\zeta_t) = Var\left(\begin{bmatrix} \zeta_{21,t} \\ \zeta_{31,t} \\ \zeta_{32,t} \end{bmatrix}\right) = \begin{bmatrix} S_1 & 0_{1x2} \\ 0_{2x1} & S_2 \end{bmatrix} \tag{6}$$

which implies independence across the blocks of S with $S_1 \equiv Var(\zeta_{21,t})$ and $S_2 \equiv Var\left(\left[\zeta_{31,t}, \zeta_{32,t}\right]'\right)$ so that the covariance states can be estimated equation by equation.

Prior distributions and initial values. The priors of the regression coefficients, the covariances and the log volatilities, $p(\theta_0)$, $p(\alpha_0)$ and $p(\ln h_0)$ respectively, are assumed to be normally distributed, independent of each other and independent of the hyperparameters which are the elements of Q, S and σ_i^2 for $i = 1, 2, 3$. The priors are calibrated on the point estimates of a constant-coefficient VAR(4) estimated over the period 1947Q2-1972Q2.

[3] As pointed out by Primiceri (2005), the random walk assumption has the desirable property of focusing on permanent parameter shifts and of reducing the number of parameters to be estimated.

[4] Stochastic volatility models are typically used to infer values for unobservable conditional volatilities. The main advantage of modeling the heteroscedastic structure of the innovation variances by a stochastic volatility model as opposed to the more common GARCH specification lies in its parsimony and in the independence of the conditional variance and the conditional mean. Put differently, changes in the dependent variable are driven by two different random variables since the conditional mean and the conditional variance evolve separately. Implicit in the random walk assumption is the view that the volatilities evolve smoothly.

We set $\theta_0 \sim N\left[\widehat{\theta}_{OLS}, \widehat{P}_{OLS}\right]$ where $\widehat{\theta}_{OLS}$ corresponds to the OLS point estimates of the training sample and $\widehat{P}_{OLS}$ to four times the covariance matrix $\widehat{V}\left(\widehat{\theta}_{OLS}\right)$. With regard to the prior specification of α_0 and h_0, we follow Primiceri (2005) and Benati and Mumtaz (2007). Let $P = AD^{1/2}$ be the Choleski factor of the time-invariant variance-covariance matrix $\widehat{\Sigma}_{OLS}$ of the reduced-form innovations from the estimation of the fixed-coefficient VAR(4) where A is a lower triangular matrix with ones on the diagonal, and $D^{1/2}$ denotes a diagonal matrix whose elements are the standard deviations of the residuals. Then the prior for the log volatilities is set to $\ln h_0 \sim N\left(\ln \mu_0, 10 \times I_3\right)$ where μ_0 is a vector that contains the diagonal elements of $D^{1/2}$ squared, and the variance-covariance matrix is arbitrarily set to ten times the identity matrix to make the prior only weakly informative. The prior for the contemporaneous interrelations is set as $\alpha_0 \sim N\left[\widetilde{\alpha}_0, \widetilde{V}\left(\widetilde{\alpha}_0\right)\right]$, where the prior mean for α_0 is obtained by taking the inverse of A and stacking the elements below the diagonal row by row in a vector in the following way: $\widetilde{\alpha}_0 = [\widetilde{\alpha}_{0,21}, \widetilde{\alpha}_{0,31}, \widetilde{\alpha}_{0,32}]'$. The covariance matrix, $\widetilde{V}\left(\widetilde{\alpha}_0\right)$, is assumed to be diagonal with each diagonal element set to ten times the absolute value of the corresponding element in $\widetilde{\alpha}_0$. While this scaling is obviously arbitrary, it accounts for the relative magnitude of the elements in $\widetilde{\alpha}_0$ as noted by Benati and Mumtaz (2007).

With regard to the hyperparameters, we make the following assumptions along the lines of Benati and Mumtaz (2007). We postulate that Q follows an inverse-Wishart distribution: $Q \sim IW\left(\overline{Q}^{-1}, T_0\right)$, where T_0 is the prior degrees of freedom and is set equal to the length of the training sample, which is sufficiently long (25 years of quarterly data) to guarantee a proper prior. Following Primiceri (2005), we adopt a relatively conservative prior for the time variation in the parameters by setting the scale matrix to $\overline{Q} = (0.01)^2 \cdot \widehat{V}\left(\widehat{\theta}_{OLS}\right)$ multiplied by the prior degrees of freedom. This is a weakly informative prior, and the particular choice for its starting value is not expected to influence the results substantially since the prior is dominated by the sample information as time progresses. We have experimented with different initial conditions that induce different amounts of time variation in the coefficients to test whether our results are sensitive to the choice of the prior specification. We follow Cogley and Sargent (2005) in setting the prior degrees of freedom alternatively to the minimum value allowed for the prior to be proper, $T_0 = \dim\left(\theta_t\right) + 1$, together with a different value of the scale matrix, $\overline{Q} = 3.5e^{-4} \cdot \widehat{V}\left(\widehat{\theta}_{OLS}\right)$, which together put as little weight as possible on our prior belief about the drift in θ_t. We have also investigated the opposite assumption by choosing $\overline{Q} = 0.01 \cdot \widehat{V}\left(\widehat{\theta}_{OLS}\right)$, which allows for a substantial amount of time variation in the parameters. Our results are not affected by these different choices for the initial values of this prior. The two blocks of S are postulated to follow inverse-Wishart distributions, with the prior degrees of freedom set equal to the minimum value required for the prior to be proper: $S_1 \sim IW\left(\overline{S}_1^{-1}, 2\right)$ and $S_2 \sim IW\left(\overline{S}_2^{-1}, 3\right)$. As for the scale matrices, they are calibrated on the absolute values of the respective elements in $\widetilde{\alpha}_0$ as in Benati and Mumtaz (2007). Given the univariate feature of the law of motion of the stochastic volatilities, the variances of the innovations

to the univariate stochastic volatility equations are drawn from an inverse-Gamma distribution as in Cogley and Sargent (2005): $\sigma_i^2 \sim IG\left(\frac{10^{-4}}{2}, \frac{1}{2}\right)$.

MCMC algorithm (Metropolis within Gibbs sampler): Simulating the posterior distribution. Since sampling from the joint posterior is complicated, we simulate the posterior distribution by sequentially drawing from the conditional posteriors of the four blocks of parameters: the coefficients θ^T, the simultaneous relations A^T, the variances H^T, where the superscript T refers to the whole sample, and the hyperparameters – the elements of Q, S, and σ_i^2 for $i = 1, 2, 3$ – collectively referred to as M. Posteriors for each block of the Gibbs sampler are conditional on the observed data Y^T and the rest of the parameters drawn at previous steps.

Step 1: Drawing coefficient states

Conditional on A^T, H^T, M, and Y^T, the measurement equation is linear and has Gaussian innovations with known variance. Therefore, the conditional posterior is a product of Gaussian densities, and θ^Tcan be drawn using a standard simulation smoother (see Carter and Kohn 1994; Cogley and Sargent 2005) that produces a trajectory of parameters:

$$p\left(\theta^T \mid Y^T, A^T, H^T\right) = p\left(\theta_T \mid Y^T, A^T, H^T\right) \prod_{t=1}^{T-1} p\left(\theta_t \mid \theta_{t+1}, Y^T, A^T, H^T\right)$$

From the terminal state of the forward Kalman filter, the backward recursions produce the required smoothed draws which take the information of the whole sample into account. More specifically, the last iteration of the filter provides the conditional mean $\theta_{T|T}$ and conditional variance $P_{T|T}$ of the posterior distribution. A draw from this distribution provides the input for the backward recursion at $T-1$ and so on until the beginning of the sample according to:

$$\begin{aligned} \theta_{t|t+1} &= \theta_{t|t} + P_{t|t} P_{t+1|t}^{-1} \left(\theta_{t+1} - \theta_t\right) \\ P_{t|t+1} &= P_{t|t} - P_{t|t} P_{t+1|t}^{-1} P_{t|t} \end{aligned}$$

Step 2: Drawing covariance states

Similarly, the posterior of A^Tconditional on θ^T, H^T, and Y^T is a product of normal densities and can be calculated by applying the same algorithm as in step 1 as a consequence of the block diagonal structure of the variance-covariance matrix S. More specifically, a system of unrelated regressions based on the following relation: $A_t u_t = \varepsilon_t$, where ε_t are orthogonalized innovations with known time-varying variance H_t and $u_t = y_t - X_t'\theta_t$ are observable residuals, can be estimated to recover A^T according to the following transformed equations where the residuals are independent

standard normal:

$$\begin{aligned} u_{1,t} &= \varepsilon_{1,t} \\ \left(h_{2,t}^{-\frac{1}{2}} u_{2,t}\right) &= -\alpha_{2,1}\left(h_{2,t}^{-\frac{1}{2}} u_{1,t}\right) + \left(h_{2,t}^{-\frac{1}{2}} \varepsilon_{2,t}\right) \\ \left(h_{3,t}^{-\frac{1}{2}} u_{3,t}\right) &= -\alpha_{3,1}\left(h_{3,t}^{-\frac{1}{2}} u_{1,t}\right) - \alpha_{3,2}\left(h_{3,t}^{-\frac{1}{2}} u_{2,t}\right) + \left(h_{3,t}^{-\frac{1}{2}} \varepsilon_{3,t}\right) \end{aligned}$$

Step 3: Drawing volatility states

Conditional on θ^T, A^T, and Y^T, the orthogonalized innovations $\varepsilon_t \equiv A_t\left(y_t - X_t'\theta_t\right)$, with $Var\left(\varepsilon_t\right) = H_t$, are observable. However, drawing from the conditional posterior of H^T is more involved because the conditional state-space representation for $\ln h_{i,t}$ is not Gaussian. The log-normal prior on the volatility parameters is common in the stochastic volatility literature, but such a prior is not conjugate. Following Cogley and Sargent (2005, Appendix B.2.5) and Benati and Mumtaz (2007), we apply the univariate algorithm developed by Jacquier, Polson, and Rossi (1994) that draws the volatility states $h_{i,t}$ one at a time.

Step 4: Drawing hyperparameters

The hyperparameters M of the model can be drawn directly from their respective posterior distributions since the disturbance terms of the transition equations are observable given θ^T, A^T, H^T and Y^T.

We perform $50,000$ iterations of the Gibbs sampler but keep only every 10^{th} draw in order to mitigate the autocorrelation among the draws. After an initial "burn-in" period of $50,000$ iterations, the sequence of draws of the four blocks from their respective conditional posteriors converges to a sample from the joint posterior distribution $p\left(\theta^T, A^T, H^T, M \mid Y^T\right)$. Following Primiceri (2005) and Benati and Mumtaz (2007), we ascertain that our Markov chain has converged to the ergodic distribution by computing the draws' inefficiency factors which are the inverse of the relative numerical efficiency measure (RNE) introduced by Geweke (1992),

$$RNE = (2\pi)^{-1}\frac{1}{S(0)}\int_{-\pi}^{\pi} S(\omega)d\omega \tag{7}$$

where $S(\omega)$ is the spectral density of the retained draws from the Gibbs sampling replications for each set of parameters at frequency ω.[5] Figure 1A displays the inefficiency factors for the states and the hyperparameters of the model which are all far below the value of 20 designated as an upper bound by Primiceri (2005). Thus, the autocorrelation across draws is modest for all elements, providing evidence of convergence to the ergodic distribution. In total, we have $5,000$ simulated values from the Gibbs chain on which we base our structural analysis.

[5] See Benati and Mumtaz (2007) for details on the implementation.

3 Impulse responses and sign restrictions

Here we describe the Monte Carlo integration procedure we use to compute the path of impulse response functions to our three structural shocks. In the spirit of Koop, Pesaran, and Potter (1996) we compute the generalized impulse responses as the difference between two conditional expectations with and without exogenous shocks:

$$IRF_{t+k} = E\left[y_{t+k} \mid \varepsilon_t, \omega_t\right] - E\left[y_{t+k} \mid \omega_t\right] \tag{8}$$

where y_{t+k} contains the forecasts of the endogenous variables at horizon k, ω_t represents the current information set, and ε_t is a vector of current disturbance terms. At each point in time, the information set upon which we condition the forecasts contains the actual values of the lagged endogenous variables and a random draw of the model parameters and hyperparameters. More specifically, in order to calculate the conditional expectations, we simulate the model in the following way. We randomly draw one possible state of the economy at time t from the Gibbs sampler output represented by the time-varying lagged coefficients and the elements of the variance-covariance matrix. Starting from this random draw from the joint posterior including hyperparameters, we stochastically simulate the future paths of the coefficient vector as well as the components of the variance-covariance matrix based on the transition laws for up to 20 quarters into the future. By projecting the evolution of the system into the future in this way, we account for all of the potential sources of uncertainty deriving from the additive innovations, variations in the lagged coefficients and changes in the contemporaneous relations among the variables in the system.

To obtain the time-varying structural impact matrix $B_{0,t}$, we implement the procedure proposed by Rubio-Ramírez, Waggoner, and Zha (2010). Given the current state of the economy, let $\Omega_t = P_t D_t P_t'$ be the eigenvalue-eigenvector decomposition of the VAR's time-varying covariance matrix Ω_t at time t. Draw an $N \times N$ matrix, K, from the $N(0,1)$ distribution, take the QR decomposition of K, where R is a diagonal matrix whose elements are normalized to be positive, and Q is a matrix whose columns are orthogonal to each other, and compute the time-varying structural impact matrix as $B_{0,t} = P_t D_t^{\frac{1}{2}} Q'$. Given this contemporaneous impact matrix, we compute the reduced-form innovations based on the relationship $u_t = B_{0,t}\varepsilon_t$, where ε_t contains three structural shocks obtained by drawing from a standard normal distribution. Impulse responses are then computed by comparing the effects of a shock on the evolution of the endogenous variables to the benchmark case without a shock, where in the former case the shock is set to $\varepsilon_{i,t} + 1$, while in the latter we only consider $\varepsilon_{i,t}$. The reason for this is to allow the system to be hit by other disturbances during the propagation of the shocks of interest. From the set of impulse responses derived in this way, we select only those impulse responses that on impact satisfy the whole set of sign restrictions, i.e. jointly display the effects on the endogenous variables associated with the structural shocks we wish to identify; all others are discarded. Within this loop, we also compute from all accepted draws

of the impulse responses the implied price elasticities of oil supply and oil demand as the ratio of the impact responses of oil production and of the real oil price after demand-side and supply-side shocks respectively. We repeat this procedure until 100 iterations fulfil the sign restrictions and then calculate the mean responses of our three endogenous variables over these accepted simulations as well as the average price elasticities. For each point in time, we randomly draw 500 current states of the economy which provide the distribution of impulse responses and impact elasticities taking into account possible developments in the structure of the economy. In a second step, following Kilian and Murphy (2010, 2011), we select only those impulse responses and corresponding elasticity estimates that jointly satisfy the upper bound on the oil supply elasticity of 0.6 and the lower bound on the oil demand elasticity of -0.8. This results in a set of admissible structural models that can change in size at each point in time.

4 Evidence on the nature of parameter evolution

4.1 A Monte Carlo study

To explore whether our econometric model with smooth transitions is well suited to capture abrupt changes in the data, we carry out a Monte Carlo exercise based on simulated data where the underlying data-generating process is characterized by a one-time break. Given that our benchmark model is too complex to be amenable to a Monte Carlo study, we assess the performance of its main building blocks by conducting two separate experiments. We generate data from (1) an AR(1) model with one exogenous regressor that features a one-off regime shift in its coefficients, and (2) a bivariate version of our benchmark VAR(4) model with an abrupt break in the variance. These two simpler models provide a parsimonious way to assess the appropriateness of modeling structural change in a smoothly evolving way as opposed to a regime switch.

4.1.1 A regression model with a break in the coefficients

To illustrate the effects of incorrectly assuming a smooth process for the evolution of the coefficients, we simulate data from the following stationary AR(1) model with one exogenous regressor, written in demeaned form:

$$y_t = \alpha_i y_{t-1} + \beta_i x_{t-1} + \epsilon_t \qquad \epsilon_t \sim N(0,1) \tag{9}$$

where we set $\alpha_1 = 0.2$ and $\beta_1 = 0.5$ for the first half of the sample and $\alpha_2 = 0.6$ and $\beta_2 = 1.5$ for the second half of the sample. For each sample generated with this parameterization, we estimate a model that postulates that the coefficient vector $\delta = [\alpha \quad \beta]'$ evolves smoothly according to a driftless random walk process:

$$\delta_t = \delta_{t-1} + \eta_t \qquad \eta_t \sim N(0, Q) \tag{10}$$

We estimate the state-space model in equations (9) and (10) by Bayesian methods described in Kim and Nelson (1999). The unrestricted prior for the initial state is Gaussian:

$$\delta_0 \sim N(\widehat{\delta}_{OLS}, 4 \cdot \widehat{V}(\widehat{\delta}_{OLS})) \tag{11}$$

where $\widehat{\delta}_{OLS}$ and $\widehat{V}(\widehat{\delta}_{OLS})$ are the OLS point estimate and asymptotic variance based on a training sample as in our benchmark model. For the variance σ^2 in the observation equation, we postulate an inverse-gamma distribution:

$$\sigma^2 \sim IG\left(\frac{\lambda}{2}, \frac{\nu}{2}\right) \tag{12}$$

with scale parameter $\lambda = 0.01$ and degrees-of-freedom parameter $\nu = 2$. The prior for Q is assumed to be inverse Wishart:

$$Q \sim IW\left(\overline{Q}^{-1}, \xi\right) \tag{13}$$

where $\overline{Q} = 0.01 \cdot \xi$ and $\xi = 3$. The starting values for the coefficients are set to the OLS estimates, $\sigma_0 = 1$, and $Q_0 = 0.1 \cdot I_2$ where I_2 is a 2×2 identity matrix. The time-varying coefficients are drawn using the Carter and Kohn (1994) algorithm outlined above. We constrain α_t to be less than one in absolute value at all dates t. The first $2,000$ draws in the Gibbs simulation process are discarded to ensure convergence. The posterior mean of $\widehat{\delta}_t$ is computed based on the remaining $1,000$ generated values. To evaluate how well this model can pick up the break imposed in the data-generating process, we also obtain an estimate of $\widehat{\delta}_t^{dummy}$ from a model that includes a dummy variable that takes a value of 0 before the break and 1 thereafter. In this way, we can construct error bands that capture the parameter uncertainty in estimating the true model.

We construct sample sizes of $T_1 = 200$ and $T_2 = 600$ after discarding the first $1,000$ periods to remove the influence of the initial values. A sample size of 200 can be considered the equivalent of the typical sample length for quarterly time series available for the post-WWII period, and 600 is representative of such a dataset at monthly frequency. There are $T/2$ data points on each side of the break date. We carry out $1,000$ Monte Carlo replications for each model and sample size.

Figure 2A reports the mean of the estimates for the exogenous coefficient and for the AR coefficient for the smooth-transition model and for the discrete-break model together with the 68% and 90% posterior credible sets for the two sample sizes. The estimation results show that the drifiting coefficient model locates the break in a satisfactory manner and moves relatively swiftly to the new regime.

4.1.2 A bivariate VAR model with a break in the variance

In the second experiment, the data-generating process is a bivariate VAR(4) model similar to equation (1) in the main text:

$$y_t = X_t'\theta + \varepsilon_t \tag{14}$$

where y_t denotes a vector of variables; X_t is a matrix including four lags of y_t and a constant; θ is a coefficient matrix, and $\varepsilon_t \sim N(0, \Omega_i)$, $i = 1, 2$ with the following variance-covariance matrices for two subperiods:

$$\Omega_1 = \begin{bmatrix} 20 & 5 \\ 5 & 30 \end{bmatrix} \qquad \Omega_2 = \begin{bmatrix} 1.5 & -4 \\ -4 & 300 \end{bmatrix} \tag{15}$$

To obtain a realistic parameterization for Ω_1 and Ω_2, we take guidance from the estimation of a bivariate VAR model for oil production and the real price of oil over two subsamples. The data generated from this model mimic a specific feature of the observed oil production and oil price series, namely a considerable decrease in oil production volatility and an increase in oil price volatility after the break in the variance. Figure 3A, panel A illustrates this behavior in one such random sample. The length of each sample generated from this model is 750, and the initial 500 periods are removed to yield a sample similar in size to that used in the empirical analysis.

For each sample, we estimate the time-varying VAR model with stochastic volatility presented in section 2. We retain the same priors as in the benchmark model and obtain initial values from the estimation of a constant-coefficient VAR(4) over a twenty–five–year training sample. This leaves us with 150 observations for the actual estimation, and the regime switch occurs at $t = 68$. Given the greater complexity of this model, we can only perform 250 Monte Carlo replications, and the results should consequently be viewed as suggestive. It should, however, be sufficient to examine the speed of transition from one regime to the other, which is the main feature of interest.

Figure 3A, panel B displays the time profile of the average of the variance estimates over the Monte Carlo simulations together with both the 16^{th} and 84^{th} and the 5^{th} and 95^{th} percentiles of the posterior distribution. The results indicate that our approach has the power to detect the regime shift to a satisfactory degree even in a relatively short sample.

4.2 Additional statistical evidence for the TVP model

To provide additional evidence on the question of whether any parameter changes have been gradual or abrupt, we incorporate into our time-varying model a mixture innovation model along the lines suggested by Koop, Leon-Gonzalez, and Strachan (2009) that allows estimating the number of breakpoints in a data-driven way. Such a model nests the two extreme cases of models with few (possibly large) breaks and those with many (possibly small) breaks and lets the data characterize the nature of time variation in the parameters. Specifically, instead of imposing a break every period as in our benchmark model, we modify the three transition equations by introducing the mixture innovation variables K_{1t}, K_{2t} and K_{3t} for $t = 1, ..., T$ where the associated transition probabilities p_1, p_2 and p_3 endogenously determine the probability of change in each set of parameters at any

given point in time:

$$\theta_t = \theta_{t-1} + K_{1t}\nu_t \qquad \nu_t \sim N(0, Q) \tag{16}$$

$$\alpha_t = \alpha_{t-1} + K_{2t}\zeta_t \qquad \zeta_t \sim N(0, S) \tag{17}$$

$$\ln h_{i,t} = \ln h_{i,t-1} + K_{3t}\sigma_i\eta_{i,t} \qquad \eta_{i,t} \sim N(0, 1) \tag{18}$$

Following Koop et al. (2009), we use Beta priors for the transition probabilities. For details on the inclusion of the mixture innovation model into the MCMC algorithm, the reader is referred to the Technical Appendix of Koop et al. (2009).

The metric for judging whether our smoothly-evolving model receives support from the data is the posterior mean of the transition probabilities for each set of time-varying parameters. Table 1A summarizes the results for alternative choices of the prior hyperparameters α and β which reflect different a priori beliefs about the probability of change.[6] Starting with a prior that assigns a 50% probability that a break occurs at time t, we gradually reduce the likelihood of a break every period by tightening the prior, i.e. we move towards models with infrequent parameter changes (once a year and every 3 years) and more informative priors (smaller standard deviations).

Table 1A: Transition probabilities

Prior assumptions for Beta $B(\alpha, \beta)$	$E(p_1 \mid Y^T)$	$E(p_2 \mid Y^T)$	$E(p_3 \mid Y^T)$
$\alpha = 1; \beta = 1$ $(E(p) = 0.5;\ SD = 0.29)$	0.93 (0.06)	0.97 (0.02)	0.96 (0.03)
$\alpha = 1; \beta = 3$ $(E(p) = 0.25;\ SD = 0.19)$	0.73 (0.11)	0.93 (0.04)	0.94 (0.03)
$\alpha = 1; \beta = 10$ $(E(p) = 0.09;\ SD = 0.08)$	0.34 (0.09)	0.81 (0.06)	0.86 (0.04)

The results indicate that there is a very high posterior probability for changes in all three sets of model parameters at each point in time even in cases where the prior information suggests infrequent breaks. This shows that the data clearly favor a model that allows for a gradual and continuous evolution of all parameters as opposed to a Markov-switching structure, for example.

5 Further analysis

5.1 Evidence for time variation in the estimated model

In assessing the relative importance of time variation over the sample, we follow the approach proposed by Cogley, Primiceri, and Sargent (2010) who examine the joint posterior distribution of the object of interest across selected pairs of time periods presented in a scatterplot. Shifts of this

[6] We do not rely on the marginal likelihood, which is a common metric for Bayesian model comparison, because this measure tends to be more sensitive to prior information than posteriors, especially in models with high-dimensional parameter spaces (see, e.g., Koop et al. 2009).

distribution away from the 45-degree line are indicative of a systematic change over time. Figure 4A reports the joint posterior distribution of the estimated standard deviations of the real price of crude oil and world oil production shown in Figure 1, panel B in the paper for pairs of oil market episodes. As can be seen from the graphs, a considerable fraction of the pairwise posterior draws of standard deviations of the real oil price are located above the threshold for most combinations of dates, which suggests a systematic increase in oil price volatility over time. There is strong evidence for a systematic decrease in the standard deviation of global oil production with the majority of draws from the joint distribution lying below the 45-degree line as time progresses. Figure 5A displays the joint posterior distribution of the estimated short-run price elasticities of oil supply derived using the aggregate demand shock and the other oil demand shock as well as the short-run price elasticities of oil demand for representative dates. When pairs of dates pertain to periods that are not too far apart (e.g. $1976Q1$ vs. $1979Q3$ and $1990Q3$ vs. $2009Q1$), the points of their joint distribution are almost equally spread out across the dividing line, which implies that the estimated elasticities are similar for those dates. Marked differences arise for combinations of episodes early in the sample with those later in the sample, in which case the joint outcomes of the posterior draws are clustered far below or far above the 45-degree line. This provides compelling evidence that the decline in short-run price elasticities is a significant feature of the data.

5.2 Robustness analysis

5.2.1 Alternative model specification

For the reasons outlined in the main text, we re-estimate the time-varying VAR model with the real refiners' acquisition cost of oil imports expressed in log deviations from its mean instead of first differences and with the quarterly averages of the monthly global real activity indicator constructed in Kilian (2009) expressed in deviations from a linear deterministic trend instead of world industrial production growth to assess whether our findings are sensitive to this change in specification. Given that the real activity measure only starts in January 1968, we only use 5 years of data as a training sample, while everything else remains the same.

Figure 6A displays the time-varying median responses of world oil production and the real price of imported crude oil to one-standard deviation structural shocks on impact over the period $1974Q1$ to $2010Q4$ together with 68% and 95% posterior credible sets and the full set of admissible models. For ease of comparison, the dotted lines depict the median estimates obtained with the baseline model. The results from this alternative specification paint much the same picture of a gradual decline in the responsiveness of oil production and an increase in the effect on the real oil price after all three identified shocks. Figure 7A presents the corresponding median estimates for the short-run price elasticities of oil supply and oil demand with the 68% and 95% posterior credible sets and the

entire range of elasticity estimates. Again, the dotted lines refer to the median impact elasticities derived from the benchmark specification. Apart from minor differences limited to specific episodes, the evolution of the elasticities exhibits a remarkably similar pattern. Consistent with our baseline results, we observe a substantial decline in both the oil supply and oil demand elasticities over the sample period.

5.2.2 Alternative identification scheme

As discussed in the main text, we investigate the robustness of our results to a sluggish response of world industrial production to the other oil demand shock by imposing a zero restriction on impact while keeping all other sign restrictions in place. This amounts to the identification assumptions presented in Table 2A.

Table 2A: Identification restrictions

	Q_{oil}	P_{oil}	Y_{world}
Negative oil supply shock	$-$	$+$	$-$
Positive aggregate demand shock	$+$	$+$	$+$
Other positive oil demand shock	$+$	$+$	**0**

To implement a single zero restriction, we perform a determinstic rotation of the time-varying contemporaneous impact matrix $B_{0,t}$ along the lines suggested by Baumeister and Benati (2010). More specifically, we define a rotation matrix RM as

$$RM = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c & -s \\ 0 & s & c \end{bmatrix}$$

with $RM \cdot RM' = I_3$ where I_3 is a 3×3 identity matrix, $c = B_{0,t}(3,3)/\sqrt{B_{0,t}(3,2)^2 + B_{0,t}(3,3)^2}$, and $s = -B_{0,t}(3,2)/\sqrt{B_{0,t}(3,2)^2 + B_{0,t}(3,3)^2}$, where $B_{0,t}(i,j)$ denotes the (i,j) entry in the candidate impact matrix $B_{0,t}$ at time t such that we obtain a new impact matrix $\overline{B}_{0,t} = B_{0,t} \cdot RM$ that has a zero in the $(3,2)$ position. The rest of the algorithm is executed as described in section 3, and only those draws that jointly satisfy the sign restrictions, the zero restriction and the boundary restrictions are kept.

Figure 8A shows the time profile of the median impact responses of oil production and the real oil price after all three structural shocks along with the 68% and 95% posterior credible sets and the minimum and maximum responses. Applying this alternative identification strategy to the time-varying VAR model in section 3.1 in the main paper yields the same evolutionary pattern as in the benchmark case. While the median price responses are essentially identical across

identification schemes, minor differences arise for oil production, with the responses becoming slightly stronger after an aggregate demand shock and weaker after the other oil demand shock. Figure 9A portrays the evolution of the range of elasticity estimates over time together with the median impact elasticities of oil supply and oil demand. As before, the median elasticities closely track the dotted lines that represent the baseline model, showing that our main conclusions are not sensitive to the modified identification assumptions regarding the other oil demand shock.

5.3 Additional results

Since the real price of crude oil is determined by the interaction of different driving forces, it is of interest to provide additional evidence on the relative contribution of different types of structural shocks to the variability of oil market variables by means of a variance decomposition. This decomposition has been computed in the frequency domain following Benati and Mumtaz (2007) and Canova and Gambetti (2009). Figure 10A presents the time-varying percentage of the median forecast error variance of the real price of crude oil and world oil production that is accounted for by the three identified structural shocks together with the 68% and 95% posterior credible sets. The decomposition reveals that other oil demand shocks account for approximately 20 percent of the variance of the real price of oil, a contribution that experiences only moderate variations over time. In contrast, other oil demand shocks explain an increasing fraction of movements in oil production in the early part of the sample, reaching a peak of 50 percent around 1990, after which their contribution declines steadily and is currently fluctuating around 20 percent. The opposite is true for shocks to global real activity. While the contribution of aggregate demand shocks to the variability in world oil production has remained relatively stable over time –with the exception of the most recent temporary spike as a result of the financial crisis– these shocks are responsible for a substantial share of the volatility in the real oil price, a share that has increased notably since the early 1990s to between 60 and 80 percent. Taken together the demand-side shocks explain more than half of the variability in the real price of oil during the latter sample period, indicating that oil supply disturbances no longer explain the bulk of oil price fluctuations. In fact, oil supply disruptions currently contribute roughly 20 percent to the volatility of the real oil price, while their share was considerably higher in earlier periods. The fraction of the overall variance of oil production explained by supply-side shocks fluctuated between 20 and 60 percent in the early part of the sample, but has stabilized since the late 1980s at around 45 percent. While the evolutions of the variance shares are suggestive of changes over time, the posterior intervals surrounding these estimates overlap for most periods.

References

[1] Baumeister C, Benati L. 2010. *Unconventional monetary policy and the Great Recession - estimating the impact of a compression in the yield spread at the zero lower bound*. ECB Working Paper 1258.

[2] Benati L, Mumtaz H. 2007. *U.S. evolving macroeconomic dynamics: a structural investigation*. ECB Working Paper 746.

[3] Carter CK, Kohn R. 1994. On Gibbs sampling for state space models. *Biometrika* **81**(3): 541-553.

[4] Canova F, Gambetti L. 2009. Structural changes in the U.S. economy: is there a role for monetary policy? *Journal of Economic Dynamics and Control* **33**(2): 477-490.

[5] Cogley T, Primiceri GE, Sargent TJ. 2010. Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics* **2**(1): 43-69.

[6] Cogley T, Sargent TJ. 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics* **8**(2): 262-302.

[7] Energy Information Administration. 2011 (April). *Monthly Energy Review*.

[8] Geweke J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics*, Bernardo JM, Berger J, David AP, Smith AFM (eds.). Oxford University Press: Oxford; 169-193.

[9] IMF. 2006 (August). *World Economic Outlook*.

[10] Jacquier E, Polson NG, Rossi PE. 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* **12**(4): 371-418.

[11] 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.

[12] Kilian L, Murphy DP. 2010. Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models. *Journal of the European Economic Association*, forthcoming.

[13] Kilian L, Murphy DP. 2011. *The role of inventories and speculative trading in the global market for crude oil*. Mimeo, http://www-personal.umich.edu/~lkilian/MS2010126R1.pdf [15 September 2011].

[14] Koop G, Leon-Gonzalez R, Strachan RW. 2009. On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control* **33**(4): 997-1017.

[15] Koop G, Pesaran MH, Potter SM. 1996. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* **74**(1): 119-147.

[16] Litterman RB. 1983. A random walk, Markov model for the distribution of time series. *Journal of Business and Economic Statistics* **1**(2): 169-173.

[17] *Oil & Gas Journal*, weekly, various issues for 1953-1972, Tulsa, Oklahoma: Pennwell Corporation.

[18] Primiceri GE. 2005. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* **72**(3): 821-852.

[19] Rubio-Ramírez JF, Waggoner DF, Zha T. 2010. Structural vector autoregressions: theory of identification and algorithms for inference. *Review of Economic Studies* **77**(2): 665-696.

[20] United Nations. 1947-2008. *Monthly Bulletin of Statistics*.

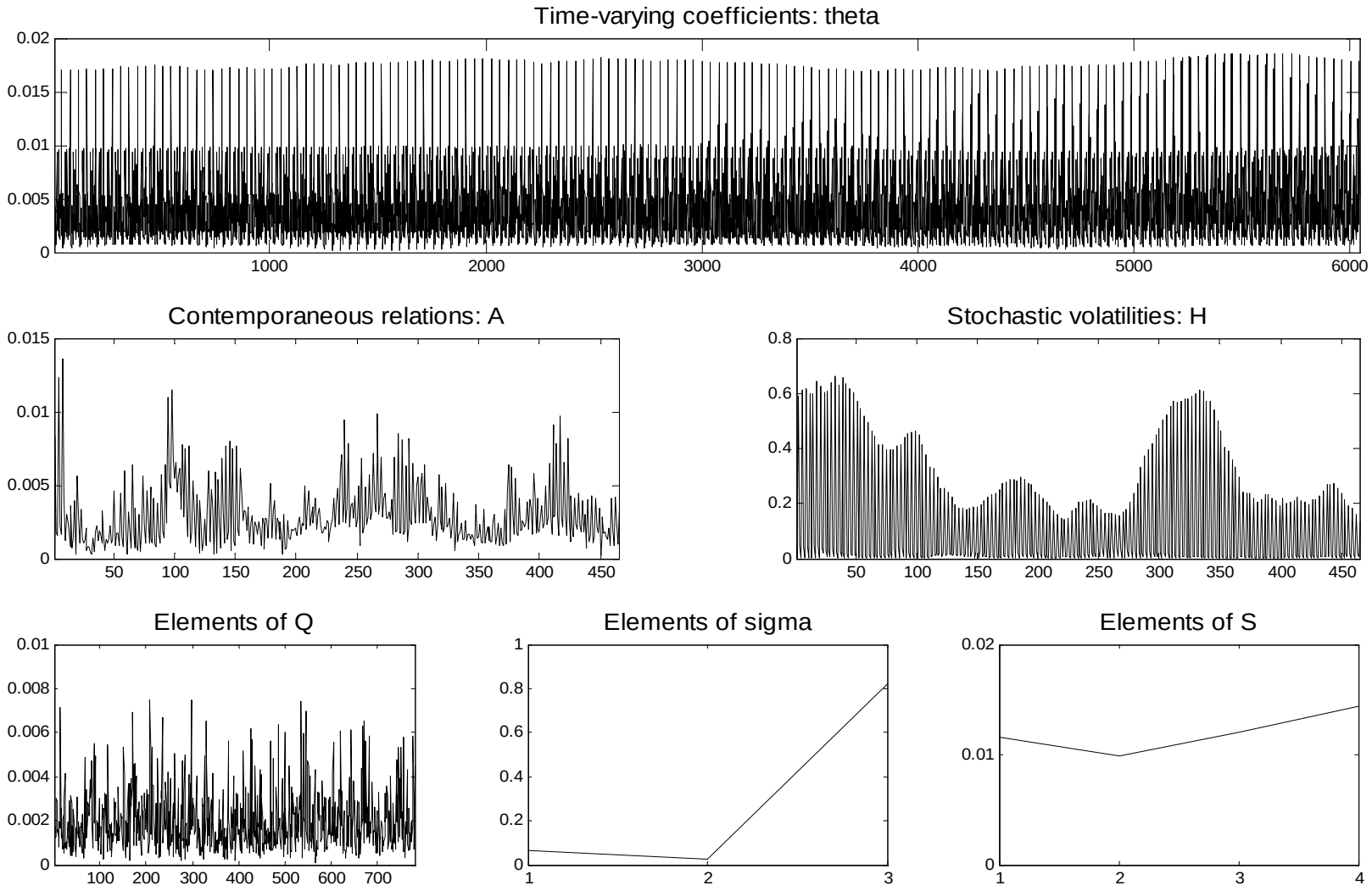
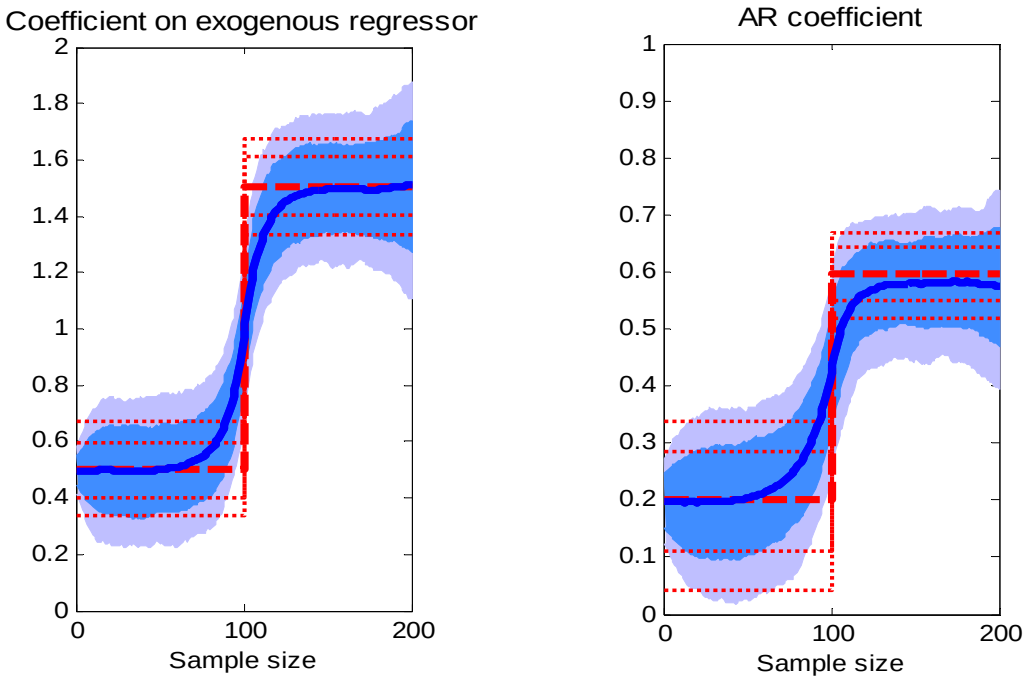


Figure 1A: Assessing the convergence of the Markov chain: inefficiency factors for the draws from the ergodic distribution for the states and the hyperparameters.

Panel A



Panel B

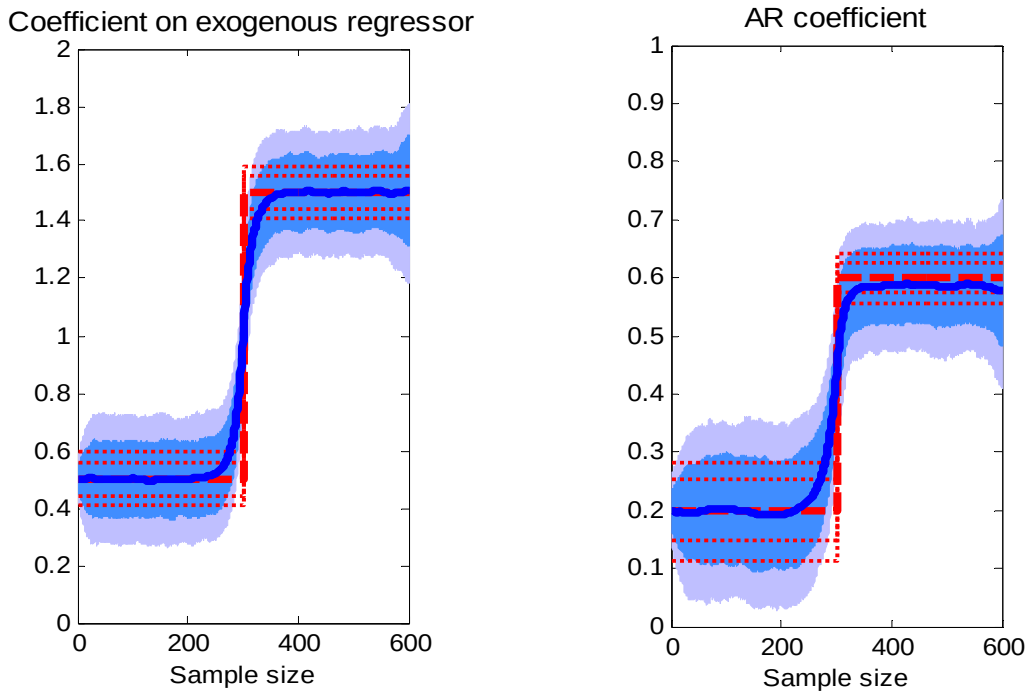
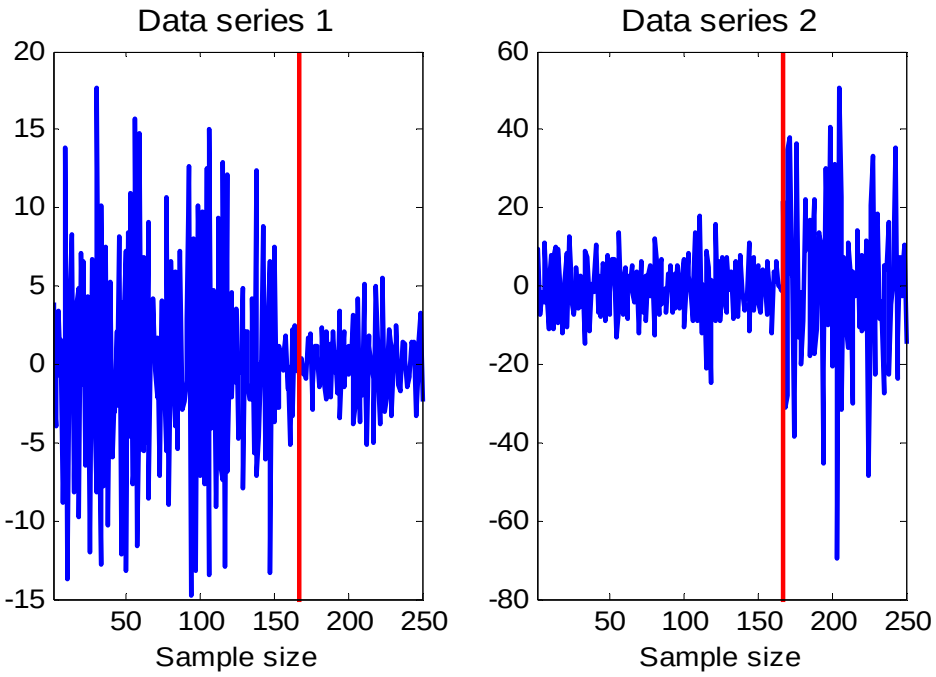


Figure 2A: Mean, 68% and 90% posterior credible sets of coefficient estimates from a smooth-transition model (solid line and shaded areas) and a discrete-break model (dashed and dotted lines) for 1,000 Monte Carlo replications.
Panel A: Sample size *T=200*.
Panel B: Sample size *T=600*.

Panel A



Panel B

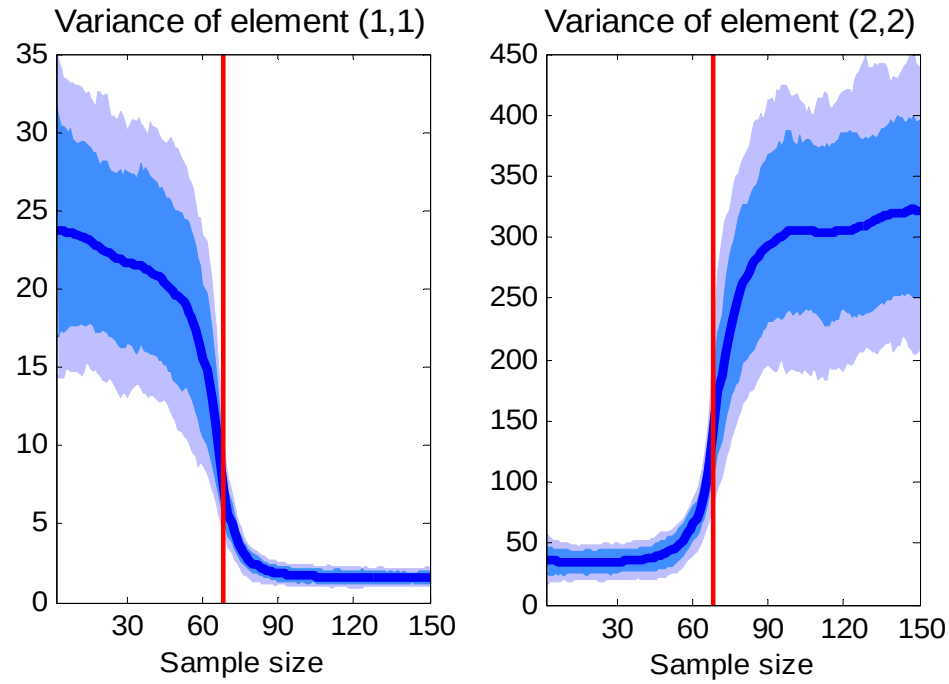


Figure 3A: Panel A: Random sample generated from bivariate VAR(4) model with break in variance at *t=168* (vertical line).
Panel B: Mean, 68% and 90% posterior credible sets of variance estimates for 250 Monte Carlo replications.

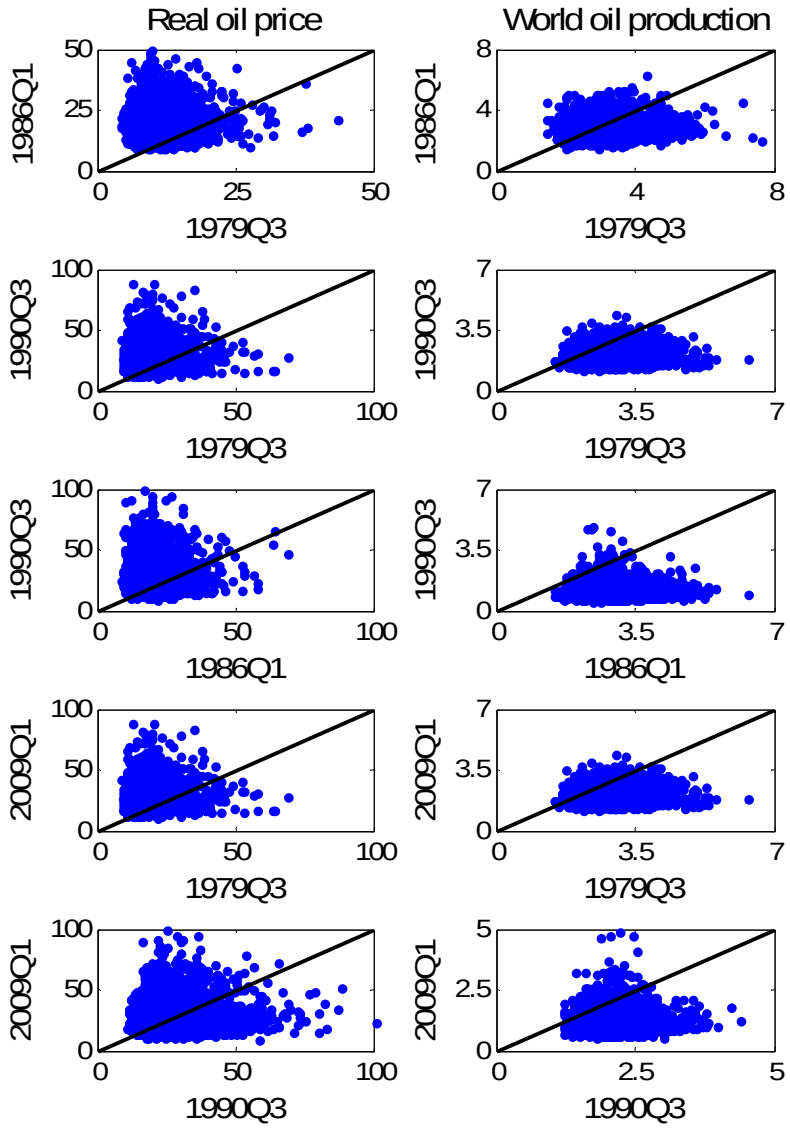


Figure 4A: Joint posterior distribution of the estimated volatilities of the real price of crude oil and world oil production for selected pairs of episodes.

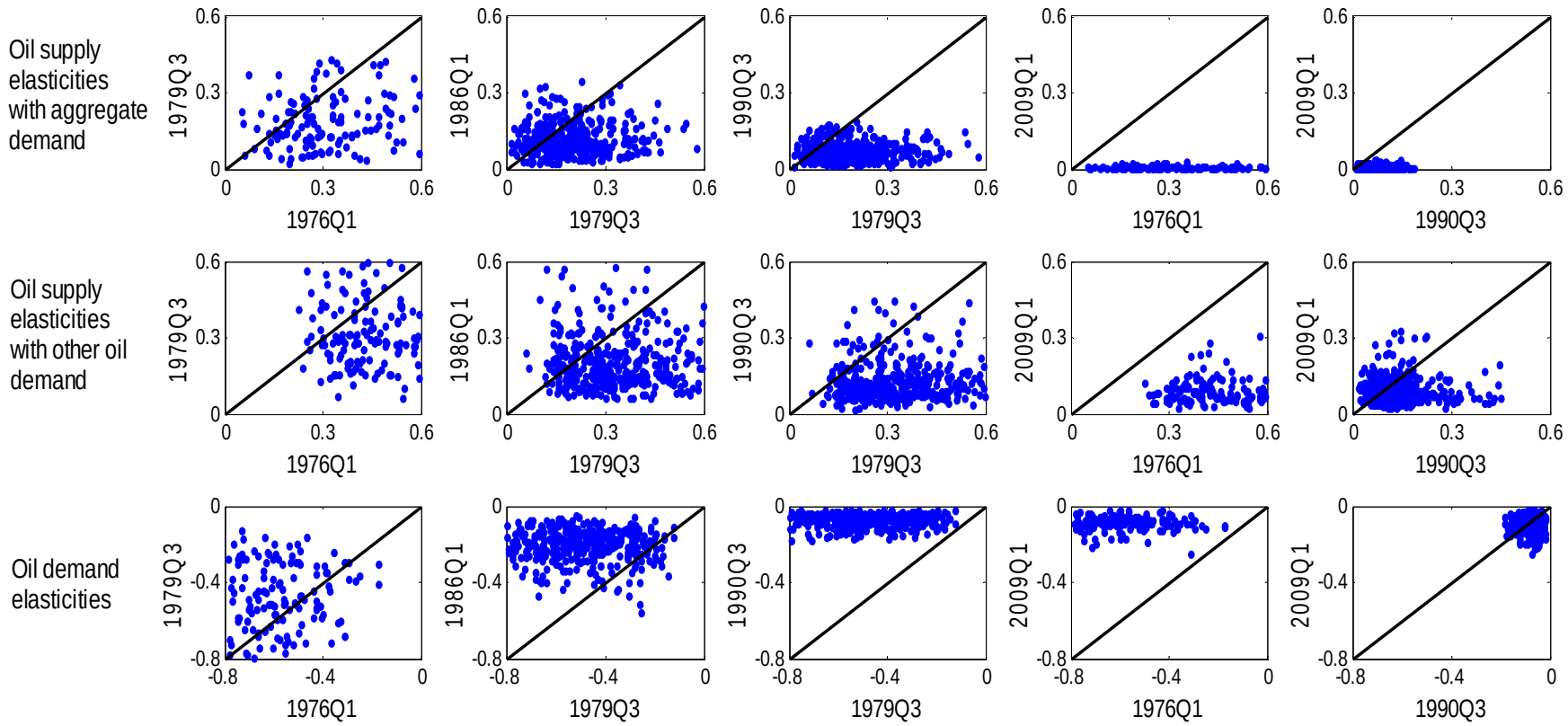


Figure 5A: Joint posterior distribution of estimated short-run price elasticities of oil supply derived with aggregate demand (1st row) and other oil demand (2nd row), and of oil demand (3rd row) for selected pairs of dates.

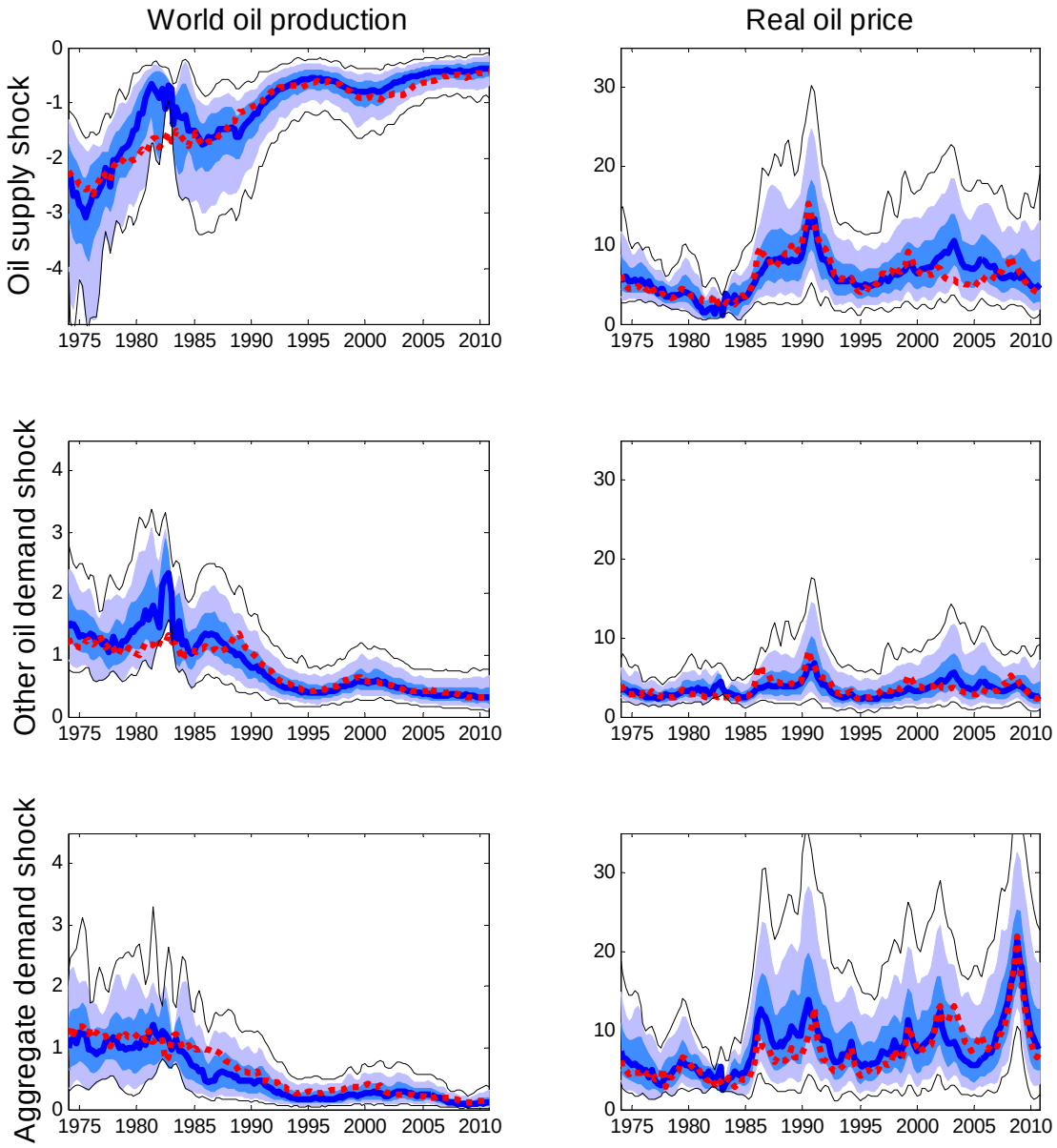


Figure 6A: Time-varying median impact impulse responses (thick solid lines) of world oil production and the real price of crude oil after oil supply shocks (1st row), other oil demand shocks (2nd row) and aggregate demand shocks (3rd row) for an alternative model specification (with real oil price in levels and Kilian's (2009) global real economic activity measure) where the dark and light shaded areas indicate respectively 68% and 95% posterior credible sets and the thin black lines indicate the full set of admissible models. The dotted red lines indicate the median responses obtained with the benchmark model.

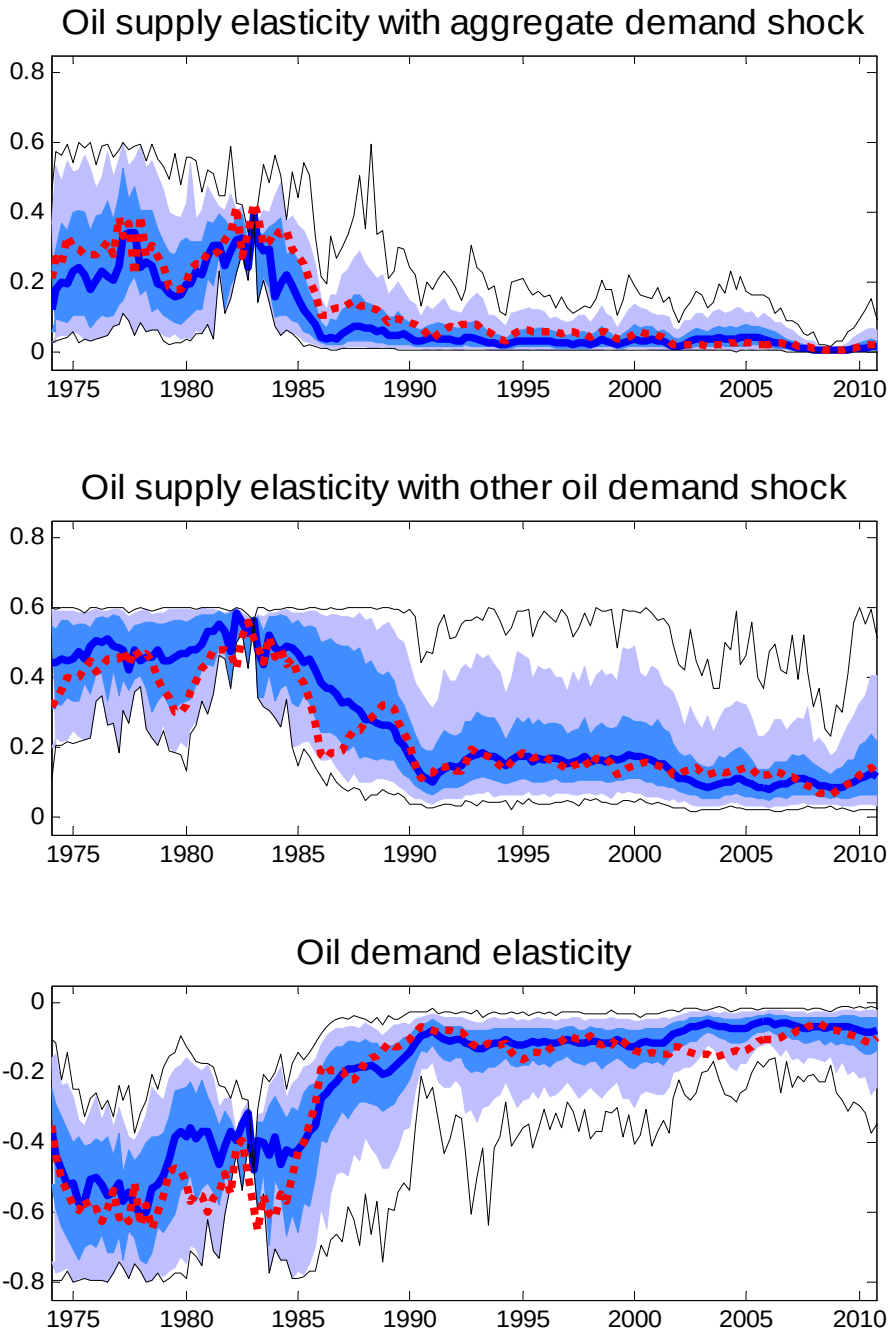


Figure 7A: Median short-run price elasticities of oil supply and oil demand (bold solid lines) for an alternative model specification (with real oil prices in levels and Kilian's (2009) global real economic activity measure) together with the 68% and 95% posterior credible sets (dark and light shaded areas) and the range of admissible models (thin solid lines). The dotted red lines indicate the median elasticities obtained with the benchmark model.

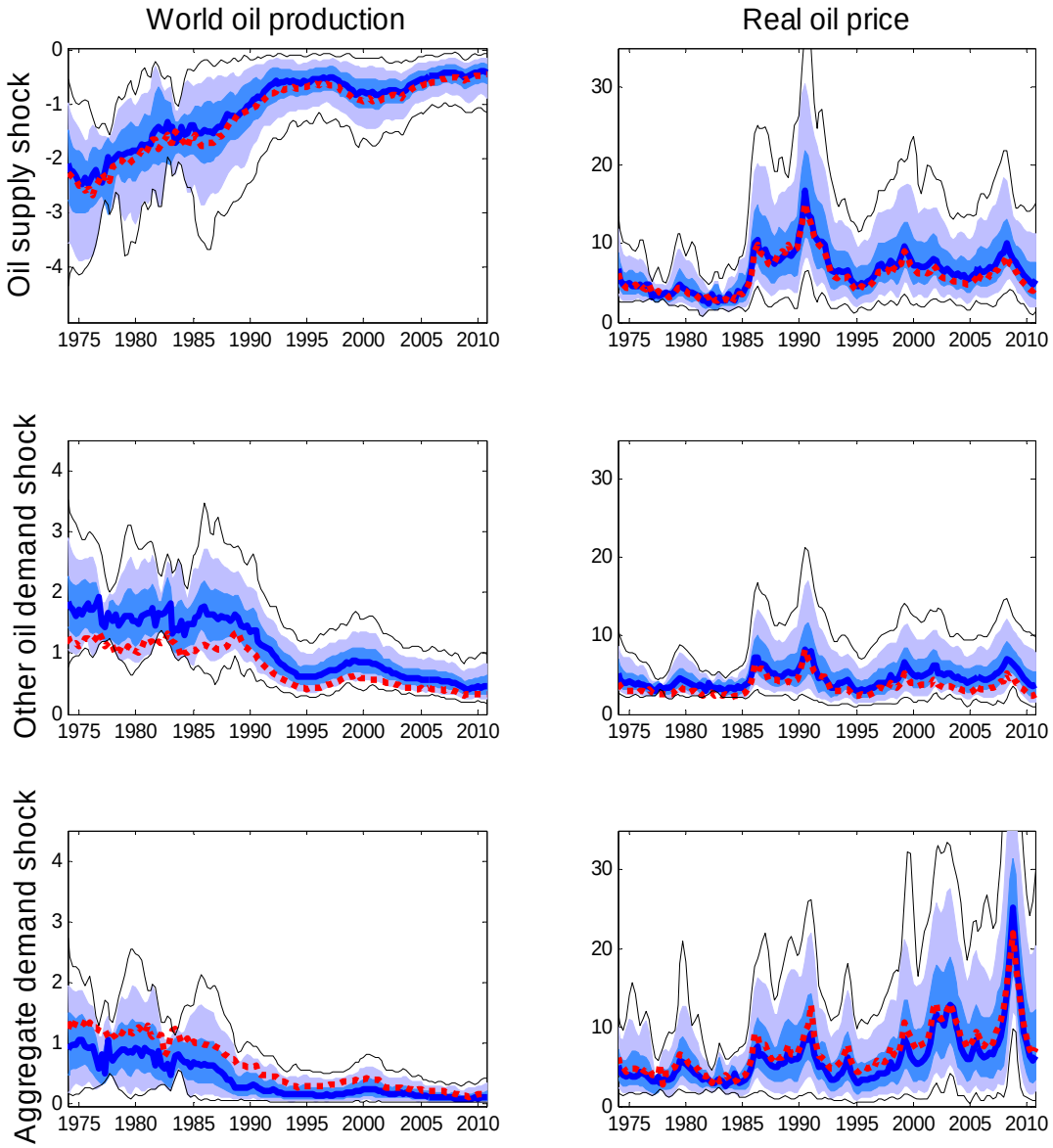


Figure 8A: Time-varying median impact impulse responses (thick solid lines) of world oil production and the real price of crude oil after oil supply shocks (1st row), other oil demand shocks (2nd row) and aggregate demand shocks (3rd row) for different identification assumptions (zero impact restriction on world industrial production after other oil demand shocks) where the dark and light shaded areas indicate respectively 68% and 95% posterior credible sets and the thin black lines indicate the full set of admissible models. The dotted red lines indicate the median responses obtained with the benchmark model.

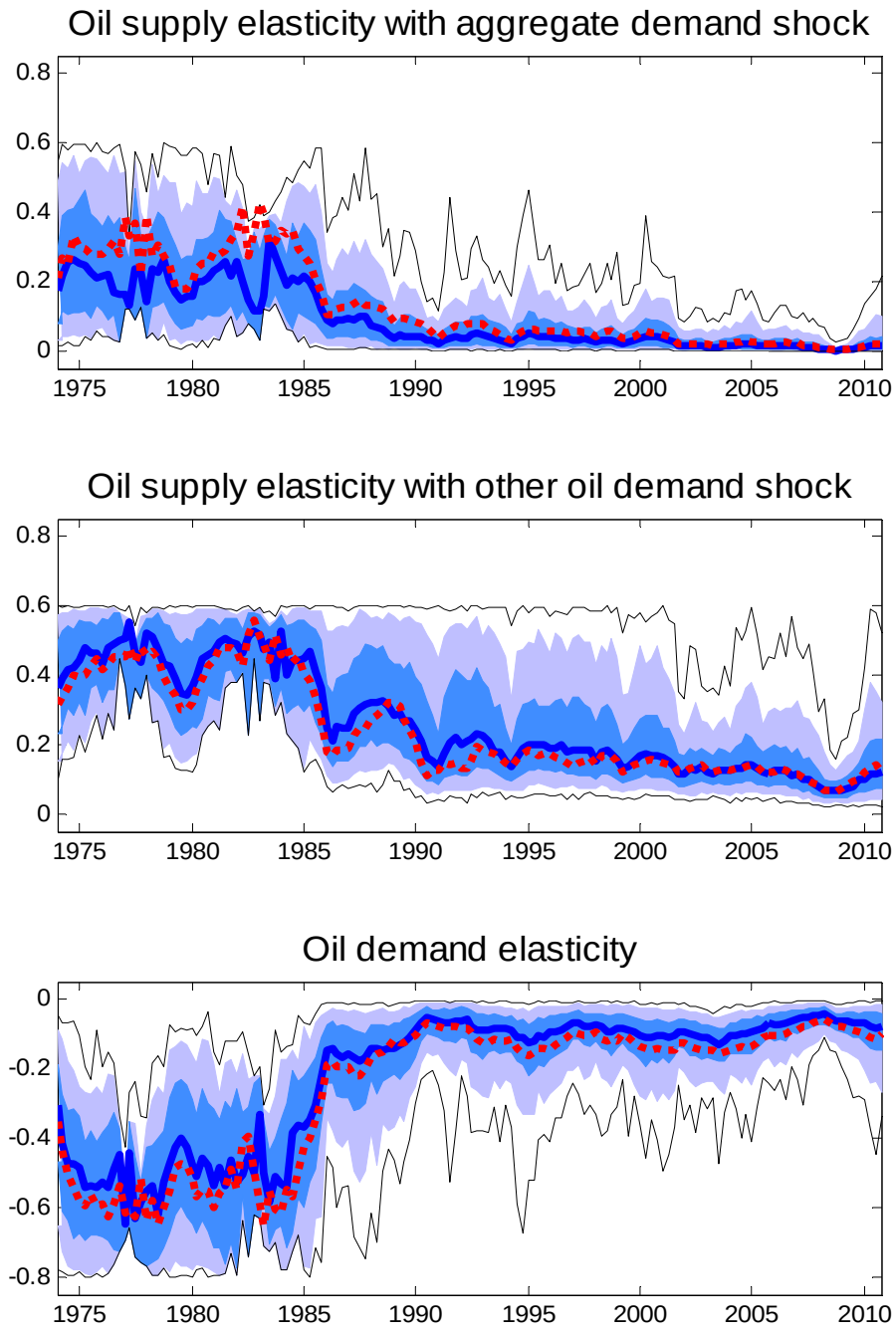


Figure 9A: Median short-run price elasticities of oil supply and oil demand (bold solid lines) for different identification assumptions (zero impact restriction on world industrial production after other oil demand shocks) together with the 68% and 95% posterior credible sets (dark and light shaded areas) and the range of admissible models (thin black lines). The dotted red lines indicate the median elasticities obtained with the benchmark model.

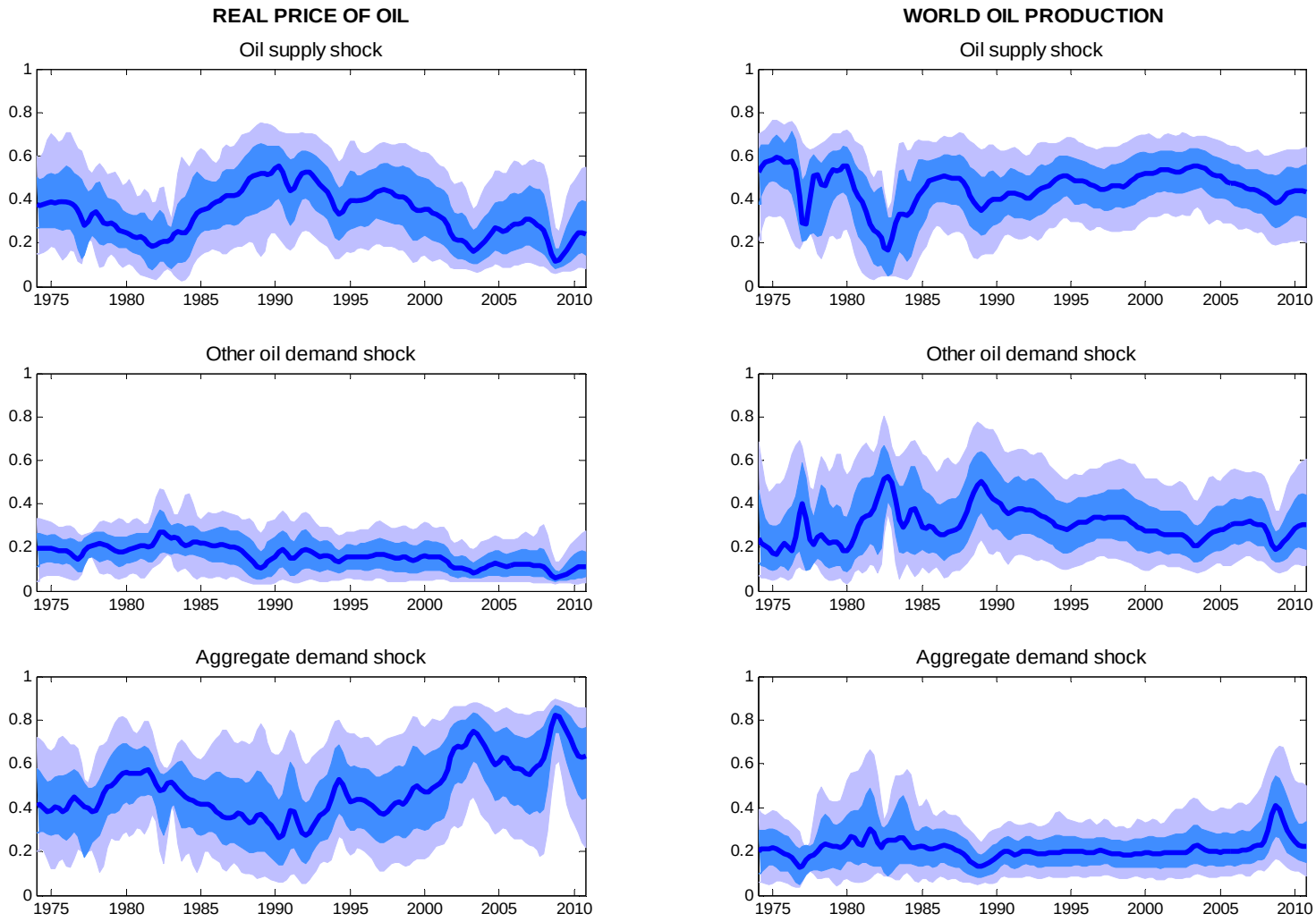


Figure 10A: Contribution of oil supply shocks, other oil demand shocks and aggregate demand shocks to the variance of the real price of crude oil and world oil production where the dark and light shaded areas indicate respectively 68% and 95% posterior credible sets and the solid line is the median.