A Full-Information Approach to Granular Instrumental Variables

June 30, 2023
Revised: November 10, 2023

Christiane Baumeister
University of Notre Dame
University of Pretoria
cjsbaumeister@gmail.com

James D. Hamilton
University of California at San Diego
jhamilton@ucsd.edu

ABSTRACT

Modeling how individual units interact to determine aggregate outcomes can be a rich source of identifying information. We use this insight to develop a generalization of granular instrumental variables estimation and show how parameters of a dynamic structural model can be estimated using full-information maximum likelihood. We apply the method to a study of the world oil market. We conclude that the supply responses of Saudi Arabia and adjustments of inventories have historically played a key role in stabilizing the price of oil.

JEL codes: Q43, C32, E32

Key words: structural vector autoregression, heterogeneous elasticities, granular instrumental variables, Russian oil supply cut

1 Introduction

Aggregate economic outcomes result from the interactions of many individual units. Modeling what those individual units have in common and the ways in which they differ can help identify the local and aggregate effects of structural shocks. A popular example is Bartik-type instruments, which use a weighted average of aggregate conditions with weights given by local shares as an instrument to estimate local elasticities; see Bartik (1991), Blanchard et al. (1992), Goldsmith-Pinkham, Sorkin and Swift (2020), and Carlino and Drautzburg (2020). Gabaix and Koijen (2020) demonstrated that in some situations one can use the difference between the arithmetic average and a share-weighted average of observations across units as a valid instrument, which they described as "granular instrumental variables." Qian (2023) extended granular instrumental variables to allow for heterogeneous responses of individual units. Banafti and Lee (2022) considered granular instrumental variables in large panels and Sarto (2022) in panel vector autoregressions. Another nice illustration of the underlying idea is Caldara, Cavallo and Iacoviello (2019), who used known exogenous shortfalls in oil production in certain countries as an instrument to estimate the response of producers in other countries to an exogenous increase in the price of oil.

In all these applications, the focus has been on developing a valid instrument for purposes of estimating a particular elasticity of interest. In this paper, we adopt a broader systemwide approach based on modeling how the actions of individual units interact to produce the aggregate outcome. We show that while such an approach could be used to motivate instrumental-variable estimation used in other studies, the underlying assumptions in fact provide a full characterization of the joint determination of local and aggregate magnitudes. Full-information maximum likelihood estimation of the general system can then be used as a framework to develop optimal instruments for every structural magnitude. In the case of a dynamic model, this amounts to the familiar approach in structural vector autoregressions of interpreting the correlations between the errors in forecasting individual observed variables as arising from an underlying set of structural shocks. The dynamic effects of structural shocks on local and aggregate variables can then be consistently estimated. Our approach typically produces a rich set of overidentifying restrictions that can be tested against the data.

We illustrate our approach with an analysis of the world oil market. We model market dynamics using a small-scale vector autoregression that includes production of oil from the three largest producers (the United States, Saudi Arabia, and Russia), consumption of oil by the three largest historical consumers (the United States, Japan, and Europe), and aggregate magnitudes. Our estimates imply a global short-run price elasticity of oil supply of 0.06, consistent with the estimates of Caldara, Cavallo and Iacoviello (2019, Table 3) of 0.05 to 0.08 and Baumeister and Hamilton (2019) of 0.15. Unlike any previous studies, our estimates are calculated by aggregating the heterogeneous responses for individual producing countries. We estimate a short-run supply elasticity of 0.26 for Saudi Arabia and 0.02 to 0.04 for other countries. A few studies have estimated separate supply elasticities for different countries or regions, and where this has been done these earlier estimates are consistent with our findings. Alonso-Alvarez, Di Nino and Venditti (2022) arrived at separate supply elasticities of 0.20 for OPEC and 0.06 for non-OPEC. When Caldara, Cavallo and Iacoviello (2019) estimated elasticities separately for different groups of countries, their estimates were 0.21 for Saudi Arabia, 0.19 for other OPEC countries, and essentially zero for non-OPEC countries, again in line with our findings. A substantially larger price elasticity for OPEC versus non-OPEC countries is also supported by the analysis in Almutairi, Pierru and Smith (2023). Our estimates also imply a large short-run supply elasticity coming from a willingness to sell oil out of inventories, which we conclude has been a very important stabilizing factor in world oil markets.

We estimate the global short-run price elasticity of the demand for oil to be −0.14, which is again consistent with the conclusions that earlier studies arrived at using very different methods from ours. Caldara, Cavallo and Iacoviello (2019, Table 3) estimated the short-run price elasticity of world petroleum demand to be −0.03 to −0.08, similar to the −0.05 estimate of Pierru, Smith and Zamrik (2018). Slightly larger estimates were obtained by Alonso-Alvarez, Di Nino and Venditti (2022) (−0.28), and Baumeister and Hamilton (2019) (−0.35). Meta-analyses of hundreds of earlier studies estimated short-run gasoline demand elasticities from −0.25 to −0.34 and short-run elasticities for the demand for crude oil from −0.05 to −0.07 (Hamilton (2009, Table 3)). Again our global elasticity is calculated directly by aggregating the demand responses for individual countries, which range from −0.02 for Japan to −0.22 for Europe.

To our knowledge, ours is the first study to simultaneously estimate demand and supply elasticities that differ across all countries. We do this in a unified statistical framework motivated by the principle that the price of oil equilibrates global supply and demand. Estimates like ours could prove helpful in calibrating theoretical models of oil markets and their effects on the world economy such as Bornstein, Krusell and

Rebelo (2023) and Balke, Jin and Yücel (2020).

We use our model to analyze the effects of different local and global shocks to supply or demand. We find that Saudi Arabian production is an important factor in stabilizing the price of oil, consistent with the conclusions of Almutairi, Pierru and Smith (2023). We further conclude that changes in inventories play a critical role in smoothing out temporary price shocks, supporting the conclusions of Knittel and Pindyck (2016) and in contrast to the claim sometimes made that inventory speculation is a destabilizing factor in world oil markets. In the absence of an ability to draw down or accumulate inventories, most of the short-run adjustment to shocks would come in the form of changes in demand rather than changes in supply.

As a case study, we use our model to analyze what would happen in response to a 50% cut in Russian oil production arising from exogenous political factors unique to that country. Our model predicts that about 1.5 million barrels a day of the 5.3 mb/d shortfall would be met by increased production, primarily from Saudi Arabia, and the rest by lower consumption.

The plan of the paper is as follows. The data are described in Section 2. Section 3 presents the model of market equilibrium that underlies our structural analysis. Section 4 discusses identification, or how we can use heterogeneous observations on price, production, and consumption to estimate supply and demand elasticities. Section 5 presents empirical estimates. Section 6 develops case studies illustrating how the model can be used. Section 7 briefly concludes.

2 Data

The U.S. Energy Information Administration publishes monthly data on the production of crude oil in a number of different countries going back to 1973 and consumption of petroleum products for a different set of countries going back to 1982. For purposes of this study, we used historical published issues of the *Monthly Energy Review* to extend the consumption data back to 1973 for a small number of countries.

2.1 Measuring growth rates

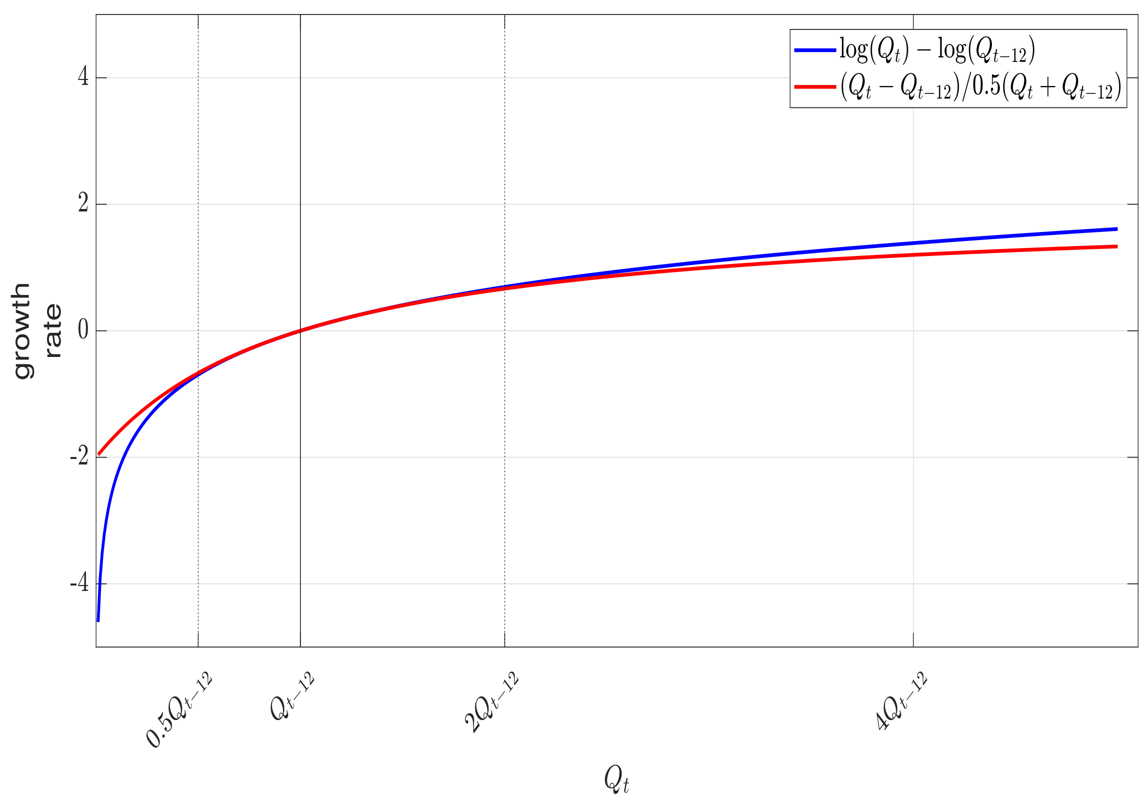
There is a strong seasonal component to petroleum consumption for some countries. For this reason we use year-over-year growth rates, which we measure as

$$q_{it} = \frac{Q_{it} - Q_{i,t-12}}{0.5(Q_{it} + Q_{i,t-12})} \tag{1}$$

$$c_{jt} = \frac{C_{jt} - C_{j,t-12}}{0.5(C_{jt} + C_{j,t-12})}. \tag{2}$$

Here Q_{it} is the quantity of oil produced in country i in month t and C_{jt} is the quantity of oil consumed in country j in month t. This is preferable to alternative measures of the growth rate such as $(Q_{it} - Q_{i,t-12})/Q_{i,t-12}$ or $\log(Q_{it}) - \log(Q_{i,t-12})$. The latter gets arbitrarily large when either Q_{it} or $Q_{i,t-12}$ get small, and would diverge to minus or plus infinity for example in the case of the complete cessation and subsequent resumption of production from Iraq and Kuwait in 1990. Expression (1) can be viewed as a first-order Taylor approximation to the function $\log Q_{it} - \log Q_{i,t-12}$ where the approximation is taken at a point halfway between Q_{it} and $Q_{i,t-12}$. The approximation is almost exact as long as Q_{it} is not less than half the size of $Q_{i,t-12}$ and not more than twice the size of $Q_{i,t-12}$; see Figure 1. For larger changes, expression (1) is less extreme than $\log(Q_{it}) - \log(Q_{i,t-12})$, and is bounded between ± 2 for all values of Q_{it} and $Q_{i,t-12}$.

Figure 1: Plot of $\log(Q_t) - \log(Q_{t-12})$ and $(Q_t - Q_{t-12})/0.5(Q_t + Q_{t-12})$ as a function of Q_t



Our production data describes countries $i = 1, 2, ..., n$ where country n is defined as "rest of world" so that $\sum_{i=1}^{n} Q_{it}$ exactly equals total measured world oil production in month t. Our baseline results use $n = 4$ where $i = 1, 2,$ or 3 correspond to the three largest producing countries over our historical sample, which were the U.S., Saudi Arabia, and Russia. A key magnitude is the average share of country i in total world production,

$$s_{qi} = T^{-1} \sum_{t=1}^{T} \frac{Q_{it}}{Q_t},$$

which we collect in an $(n \times 1)$ vector $\mathbf{s}_q$. In our data set, $\mathbf{s}_q = (0.12, 0.12, 0.15, 0.61)'$. We will approximate the annual growth in global production using the share-weighted average of individual country growth rates:

$$\frac{Q_t - Q_{t-12}}{0.5(Q_t + Q_{t-12})} = \frac{\sum_{i=1}^{n} Q_{it} - \sum_{i=1}^{n} Q_{i,t-12}}{0.5(Q_t + Q_{t-12})} = \sum_{i=1}^{n} \left[\frac{Q_{it} - Q_{i,t-12}}{0.5(Q_{it} + Q_{i,t-12})} \frac{(Q_{it} + Q_{i,t-12})}{(Q_t + Q_{t-12})} \right]$$
$$\simeq \sum_{i=1}^{n} \left[\frac{Q_{it} - Q_{i,t-12}}{0.5(Q_{it} + Q_{i,t-12})} s_{qi} \right] = \sum_{i=1}^{n} s_{qi} q_{it}. \quad (3)$$

Similarly, our consumption data describes countries $j = 1, 2, ..., m$ where country m is defined as "rest of world" so that $\sum_{j=1}^{m} C_{jt}$ exactly equals total measured world oil consumption in month t. Our procedure does not require n to equal m nor does it require the producers to be the same countries as the consumers. Our baseline results use the historically three largest consuming countries (the U.S., Japan, and Europe) so that $m = 4$. Average consumption shares are summarized by the $(m \times 1)$ vector $\mathbf{s}_c = (0.25, 0.07, 0.08, 0.60)'$. We approximate the year-over-year growth rate of global consumption as

$$\frac{C_t - C_{t-12}}{0.5(C_t + C_{t-12})} \simeq \sum_{j=1}^{m} s_{cj} c_{jt}. \quad (4)$$

It is not the case in EIA reported data that global oil consumption C_t is the same number as global oil production Q_t. There are three reasons for this. First, there are conceptual differences in definition. Production is measured in the number of barrels of oil taken out of the ground. One barrel of oil produces more than one barrel of refined product used by consumers, and additional consumable product comes from biofuels and processing of natural gas. For these reasons, measured global consumption exceeds measured global production. Second, consumption and production numbers are collected from different underlying data sources and there are acknowledged errors in measuring all of these variables. Third, true global production could be greater or less than true global consumption in a given month t if there is an increase or decrease in global oil inventories. We will take all these factors into account in the model developed below.

Although oil is produced and consumed in different locations around the world, it is a world market for oil in which the quality-adjusted product sells for essentially the same price everywhere in the world. We measure the global real price of oil in month t (denoted P_t) as the dollar price of a barrel of Brent crude oil deflated by the U.S. consumer price index. We convert this to monthly growth rates $p_t = \log(P_t) - \log(P_{t-1})$. The observed data for month t are summarized by the $[(n + m + 1) \times 1]$ vector $\mathbf{y}_t = (\mathbf{q}_t', \mathbf{c}_t', p_t)'$ consisting of the growth rates of production and consumption for each country in the world along with the world price of oil.

2.2 Data during the COVID-19 pandemic

The pandemic shut-downs in 2020 completely disrupted both oil supply and demand and had consequences that continue to affect world oil markets in 2023; for a description and analysis of these disruptions see Baumeister (2023). Events of 2020 also show up very dramatically in a broad range of other economic indicators. A number of approaches for dealing with this structural break have been proposed. Lenza and Primiceri (2022) suggested we could treat the underlying structural relations as unchanged but allow for a big increase in the magnitude of structural shocks. Ng (2021) argued that these disruptions were an entirely new shock that we could model using direct observations on measures of hospitalization, positive cases, or deaths. In our paper we adopt the more general view that potentially all the structural relations and structural shocks were different during the pandemic, implying that structural and reduced-form parameters during this episode should be estimated separately from the rest of the sample. Since there are not enough observations during the pandemic to estimate a full set of parameters over this short period, in practice this means dropping these observations from the sample and pooling post-COVID and pre-COVID observations into a single sample. Schorfheide and Song (2021), Lenza and Primiceri (2022), and Hamilton (2023) noted that this is what researchers might often want to do, and this is the approach followed in this paper.

Year-over-year growth rates of oil production and consumption are profoundly impacted for 2020:M3 through 2021:M2. Since we use twelve lags of these as explanatory variables, we therefore drop two years of data associated with the pandemic. The left-hand variable in our forecasting equations covers observations from 1975:M1 through 2020:M2 and 2022:M2 through 2023:M2, for a total of $T = 555$ observations. For notational convenience, we will write $\sum_{t \in \{1975:\text{M1-2020:M2} \,\cup\, 2022:\text{M2-2023:M2}\}}$ simply as $\sum_{t=1}^{T}$ where T is the total number of observations on the dependent variable in our pooled sample. We obtained very similar results if we just end the sample in 2019:M12.

3 Market equilibrium

The production of oil from country i is presumed to be governed by the structural equation

$$q_{it} = \phi_{qi} p_t + \mathbf{b}'_{qi}\mathbf{x}_{t-1} + u_{qit} + u_{\chi it} \tag{5}$$

for $i = 1, ..., n$. Here $\mathbf{x}_{t-1} = (1, \mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-r})'$ is a $(k \times 1)$ vector consisting of a constant term and r lags of the production and consumption of every country in the world along with the world price; thus $k = 1 + r(n + m + 1)$. The term u_{qit} represents factors other than the lags $\mathbf{x}_{t-1}$ and the current price p_t that determine the production in country i while $u_{\chi it}$ is the error in measuring the production of country i. The strategy for distinguishing a true supply shock u_{qit} from measurement error $u_{\chi it}$ will be the assumption that u_{qit} affects the equilibrium price p_t whereas $u_{\chi it}$ does not. The term $\mathbf{b}'_{qi}\mathbf{x}_{t-1}$ governs the dynamic behavior of oil supply in country i. We assume that structural dynamics are incorporated in the definition of $\mathbf{b}_{qi}$ so that $u_{qit} + u_{\chi it}$ can be regarded as serially uncorrelated. In our empirical analysis we take $r = 12$. Note that although we are measuring q_{it} in year-over-year growth rates, the inclusion of lags means that ϕ_{qi} represents the response of supply to an unanticipated change in price. Thus ϕ_{qi} should be interpreted as the within-month elasticity of supply for country i. The longer-run elasticity of supply is a function of $\mathbf{b}_{qi}$.

Likewise the structural demand equation for country j takes the form

$$c_{jt} = \phi_{cj} p_t + \mathbf{b}'_{cj}\mathbf{x}_{t-1} + u_{cjt} + u_{\psi jt} \tag{6}$$

for $j = 1, ..., m$. Here ϕ_{cj} is the short-run demand elasticity in country j, u_{cjt} is a shock to country j demand, and $u_{\psi jt}$ is measurement error.

If correctly measured global production is greater than consumption, the excess must have gone into inventories. If we knew the values of the measurement errors $u_{\chi it}$ and $u_{\psi jt}$, we could infer a growth rate of global inventories v_t from

$$v_t = \textstyle\sum_{i=1}^{n} s_{qi}(q_{it} - u_{\chi it}) - \sum_{j=1}^{m} s_{cj}(c_{jt} - u_{\psi jt}). \tag{7}$$

We assume that inventories also respond to global conditions according to the structural equation

$$v_t = \phi_v p_t + \mathbf{b}_v' \mathbf{x}_{t-1} + u_{vt} \tag{8}$$

where ϕ_v is the short-run price elasticity of inventory demand and u_{vt} a shock to inventory demand. Note that inclusion of a constant term in (8) allows for systematic average differences between measured production and consumption that have no implications for the price of oil.

The equilibrium price is determined by equations (5)-(8). It is helpful to rewrite these in vector form as[1]

$$\underset{(n\times 1)}{\mathbf{q}_t} = \underset{(n\times 1)}{\boldsymbol{\phi}_q} p_t + \underset{(n\times k)}{\mathbf{B}_q} \mathbf{x}_{t-1} + \underset{(n\times 1)}{\mathbf{u}_{qt}} + \underset{(n\times 1)}{\mathbf{u}_{\chi t}} \tag{9}$$

$$\underset{(m\times 1)}{\mathbf{c}_t} = \underset{(m\times 1)}{\boldsymbol{\phi}_c} p_t + \underset{(m\times k)}{\mathbf{B}_c} \mathbf{x}_{t-1} + \underset{(m\times 1)}{\mathbf{u}_{ct}} + \underset{(m\times 1)}{\mathbf{u}_{\psi t}} . \tag{10}$$

The equilibrium condition (7) can then be written

$$\mathbf{s}_q'(\mathbf{q}_t - \mathbf{u}_{\chi t}) = \mathbf{s}_c'(\mathbf{c}_t - \mathbf{u}_{\psi t}) + \phi_v p_t + \mathbf{b}_v' \mathbf{x}_{t-1} + u_{vt}$$

$$\mathbf{s}_q' \boldsymbol{\phi}_q p_t + \mathbf{s}_q' \mathbf{B}_q \mathbf{x}_{t-1} + \mathbf{s}_q' \mathbf{u}_{qt} = \mathbf{s}_c' \boldsymbol{\phi}_c p_t + \mathbf{s}_c' \mathbf{B}_c \mathbf{x}_{t-1} + \mathbf{s}_c' \mathbf{u}_{ct} + \phi_v p_t + \mathbf{b}_v' \mathbf{x}_{t-1} + u_{vt}$$

$$(\mathbf{s}_q' \boldsymbol{\phi}_q - \mathbf{s}_c' \boldsymbol{\phi}_c - \phi_v) p_t = (\mathbf{s}_c' \mathbf{B}_c - \mathbf{s}_q' \mathbf{B}_q + \mathbf{b}_v') \mathbf{x}_{t-1} + \mathbf{s}_c' \mathbf{u}_{ct} - \mathbf{s}_q' \mathbf{u}_{qt} + u_{vt}. \tag{11}$$

The structural model thus consists of equations (9), (10), and (11).

4 Identification

In this section we discuss how the structural elasticities can be estimated from the data under certain assumptions about the correlations between the structural disturbances.

4.1 A motivation for granular instrumental variables

To start with a simple example, suppose we were willing to assume that the shock to country j's demand consisted of a global demand shock f_{ct} that affects all consumers in the same way and a purely idiosyncratic shock η_{cjt}. Suppose further that there is a common demand elasticity ϕ_c across all consumers. Under these assumptions, (6) becomes

$$c_{jt} = \phi_c p_t + \mathbf{b}_{cj}' \mathbf{x}_{t-1} + f_{ct} + \eta_{cjt} + u_{\psi jt} \tag{12}$$

[1] Equation (9) is obtained by stacking the n equations in (5) defining $\mathbf{q}_t = (q_{1t}, ..., q_{nt})'$, $\boldsymbol{\phi}_q = (\phi_{q1}, .., \phi_{qn})'$, $\mathbf{B}_q = \begin{bmatrix} \mathbf{b}_{q1} & \mathbf{b}_{q2} & \cdots & \mathbf{b}_{qn} \end{bmatrix}'$, $\mathbf{u}_{qt} = (u_{q1t}, ..., u_{qnt})'$, and $\mathbf{u}_{\chi t} = (u_{\chi 1t}, ..., u_{\chi nt})'$. Equation (10) is likewise obtained by stacking the m equations in (6).

for $j = 1, ..., m$. Taking a simple arithmetic average of (12) across consumers results in

$$\bar{c}_t = \phi_c p_t + \bar{\mathbf{b}}_c' \mathbf{x}_{t-1} + f_{ct} + \bar{\eta}_{ct} + \bar{u}_{\psi t} \tag{13}$$

where $\bar{c}_t = m^{-1}\sum_{j=1}^{m} c_{jt}$, $\bar{\mathbf{b}}_c = m^{-1}\sum_{j=1}^{m} \mathbf{b}_{cj}$, $\bar{\eta}_{ct} = m^{-1}\sum_{j=1}^{m} \eta_{cjt}$, and $\bar{u}_{\psi t} = m^{-1}\sum_{j=1}^{m} u_{\psi jt}$. Subtracting (13) from (12) results in

$$c_{jt} - \bar{c}_t = (\mathbf{b}_{cj}' - \bar{\mathbf{b}}_c)' \mathbf{x}_{t-1} + (\eta_{cjt} - \bar{\eta}_{ct}) + (u_{\psi jt} - \bar{u}_{\psi t}). \tag{14}$$

From this equation we see that the difference between country j's consumption and average world consumption depends only on the idiosyncratic components of demand shocks and measurement errors. If these are uncorrelated with the supply shock for producer i, the value of (14) for any consumer j could serve as a valid instrument for estimation of the supply equation (5) for any producer i. If consumer j is large enough that its idiosyncratic demand shock has an effect on the world price, then (14) is a relevant instrument because it is correlated with the world price p_t. Since any linear combination of (14) across different consumers j would also be a valid instrument, we might expect to obtain a more powerful instrument by weighting each consumer by its share in the world total. Multiplying (14) by s_{cj} and summing over j,

$$c_t - \bar{c}_t = (\mathbf{b}_j' - \bar{\mathbf{b}}_j)' \mathbf{x}_{t-1} + (\eta_{ct} - \bar{\eta}_{ct}) + (u_{\psi t} - \bar{u}_{\psi t})$$

where $c_t = \sum_{j=1}^{m} s_{cj} c_{jt}$, $\mathbf{b}_j = \sum_{j=1}^{m} s_{cj} \mathbf{b}_{cj}$, $\eta_{ct} = \sum_{j=1}^{m} s_{cj} \eta_{cjt}$, and $u_{\psi t} = \sum_{j=1}^{m} s_{cj} u_{\psi jt}$. In other words, the difference between the growth rate of total world consumption c_t and the arithmetic average of the growth rate of each consuming country $\bar{c}_t$ can be used as a valid instrument for estimation of supply elasticities. This is the basic idea behind the granular instrumental variables proposed by Gabaix and Koijen (2020).

But under the above assumptions, even better estimates could be obtained by using *all* of the values of (14) for $j = 1, ..., m$ as instruments for *all* of the producing countries $i = 1, ..., n$. Moreover, the above assumptions imply a host of overidentifying assumptions that are informative about the demand parameters as well. One can characterize the first-order conditions for maximum likelihood estimation as providing the ideal instruments that should be used for estimation of all the parameters of the system; see for example Baumeister and Hamilton (2023, Section 2.2). Maximum likelihood estimation makes optimal use of all the available instruments in the sense of obtaining consistent estimates of the structural parameters with the smallest possible asymptotic variance (Rothenberg (1973)). We now develop a procedure for implementing this under a generalization of the above assumptions.

4.2 A general framework for examining identification

Our first step in generalizing the above approach is to characterize the reduced-form representation of the structural model. Define

$$\alpha = (\mathbf{s}_q'\boldsymbol{\phi}_q - \mathbf{s}_c'\boldsymbol{\phi}_c - \phi_v)^{-1}. \tag{15}$$

Equation (11) can then be written

$$\begin{aligned} p_t &= \alpha(\mathbf{s}_c'\mathbf{B}_c - \mathbf{s}_q'\mathbf{B}_q + \mathbf{b}_v')\mathbf{x}_{t-1} + \alpha\left(\mathbf{s}_c'\mathbf{u}_{ct} - \mathbf{s}_q'\mathbf{u}_{qt} + u_{vt}\right) \\ &= \boldsymbol{\pi}_p'\mathbf{x}_{t-1} + \varepsilon_{pt} \end{aligned} \tag{16}$$

for $\boldsymbol{\pi}_p' = \alpha(\mathbf{s}_c'\mathbf{B}_c - \mathbf{s}_q'\mathbf{B}_q + \mathbf{b}_v')$ and $\varepsilon_{pt} = \alpha\left(\mathbf{s}_c'\mathbf{u}_{ct} - \mathbf{s}_q'\mathbf{u}_{qt} + u_{vt}\right)$. If the structural shocks $\mathbf{u}_{qt}$, $\mathbf{u}_{ct}$ and u_{vt} are white noise, then ε_{pt} is the error one would make trying to forecast p_t using the lagged variables in $\mathbf{x}_{t-1}$, so that $\boldsymbol{\pi}_p$ could be estimated by OLS. The forecast error ε_{pt} results from unforecastable structural shocks to supply, demand, and inventories. Substituting (16) into (9) and (10) gives

$$\begin{aligned} \mathbf{q}_t &= (\boldsymbol{\phi}_q\boldsymbol{\pi}_p' + \mathbf{B}_q)\mathbf{x}_{t-1} + \boldsymbol{\phi}_q\varepsilon_{pt} + \mathbf{u}_{qt} + \mathbf{u}_{\chi t} \\ &= \boldsymbol{\Pi}_q\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_{qt} \end{aligned}$$

$$\begin{aligned} \mathbf{c}_t &= (\boldsymbol{\phi}_c\boldsymbol{\pi}_p' + \mathbf{B}_c)\mathbf{x}_{t-1} + \boldsymbol{\phi}_c\varepsilon_{pt} + \mathbf{u}_{ct} + \mathbf{u}_{\psi t} \\ &= \boldsymbol{\Pi}_c\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_{ct} \end{aligned}$$

for $\boldsymbol{\epsilon}_{qt} = \boldsymbol{\phi}_q\varepsilon_{pt} + \mathbf{u}_{qt} + \mathbf{u}_{\chi t}$ and $\boldsymbol{\epsilon}_{ct} = \boldsymbol{\phi}_c\varepsilon_{pt} + \mathbf{u}_{ct} + \mathbf{u}_{\psi t}$. Note $\boldsymbol{\epsilon}_{qt}$ and $\boldsymbol{\epsilon}_{ct}$ are the errors one would make forecasting production or consumption from lagged observables and $\boldsymbol{\Pi}_q$ and $\boldsymbol{\Pi}_c$ can again be estimated by OLS.

The reduced-form VAR is thus

$$\mathbf{y}_t = \boldsymbol{\Pi}\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t$$

for $\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}_{qt}', \boldsymbol{\epsilon}_{ct}', \varepsilon_{pt})'$. Note that $\boldsymbol{\epsilon}_t$ is characterized by

$$\mathbf{A}\boldsymbol{\epsilon}_t = \mathbf{u}_t \tag{17}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{nm} & -\boldsymbol{\phi}_q \\ \mathbf{0}_{mn} & \mathbf{I}_m & -\boldsymbol{\phi}_c \\ \mathbf{0}_{1n} & \mathbf{0}_{1m} & 1 \end{bmatrix} \tag{18}$$

$$\mathbf{u}_t = \begin{bmatrix} \mathbf{u}_{qt} + \mathbf{u}_{\chi t} \\ \mathbf{u}_{ct} + \mathbf{u}_{\psi t} \\ \alpha\left(\mathbf{s}_c'\mathbf{u}_{ct} - \mathbf{s}_q'\mathbf{u}_{qt} + u_{vt}\right) \end{bmatrix}. \tag{19}$$

The reduced-form parameters $\mathbf{\Pi}$ and $\mathbf{\Omega} = E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')$ can be estimated by OLS:

$$\hat{\mathbf{\Pi}} = \left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_{t-1}'\right)\left(\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}$$

$$\hat{\mathbf{\Omega}} = T^{-1}\sum_{t=1}^{T}(\mathbf{y}_t - \hat{\mathbf{\Pi}}\mathbf{x}_{t-1})(\mathbf{y}_t - \hat{\mathbf{\Pi}}\mathbf{x}_{t-1})'.$$

If $\mathbf{u}_t \sim$ i.i.d. $N(\mathbf{0}, \mathbf{D})$, these are also the maximum likelihood estimates.

Following Rothenberg (1971), a model that restricts $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{B}$ to a particular class is identified if any value for $\{\mathbf{\Pi}, \mathbf{\Omega}\}$ is associated with at most one value for $\{\mathbf{A}, \mathbf{D}, \mathbf{B}\}$ within the allowable class. Since we treat $\mathbf{B}$ as unrestricted, the question is whether there is at most one value for $\mathbf{A}$ and $\mathbf{D}$ for which $\mathbf{A\Omega A}' = \mathbf{D}$. We will maintain throughout that the inventory shock u_{vt} and measurement errors $\mathbf{u}_{\chi t}$ and $\mathbf{u}_{\psi t}$ are uncorrelated with all the other structural shocks, so that $\mathbf{D}$ takes the form

$$\mathbf{D} = \begin{bmatrix} \mathbf{K}_{qq} + \mathbf{\Sigma}_\chi & \mathbf{K}_{qc} & -\alpha\mathbf{K}_{qq}\mathbf{s}_q + \alpha\mathbf{K}_{qc}\mathbf{s}_c \\ \mathbf{K}_{cq} & \mathbf{K}_{cc} + \mathbf{\Sigma}_\psi & -\alpha\mathbf{K}_{cq}\mathbf{s}_q + \alpha\mathbf{K}_{cc}\mathbf{s}_c \\ -\alpha\mathbf{s}_q'\mathbf{K}_{qq} + \alpha\mathbf{s}_c'\mathbf{K}_{cq} & -\alpha\mathbf{s}_q'\mathbf{K}_{qc} + \alpha\mathbf{s}_c'\mathbf{K}_{cc} & \alpha^2(\mathbf{s}_q'\mathbf{K}_{qq}\mathbf{s}_q - 2\mathbf{s}_c'\mathbf{K}_{cq}\mathbf{s}_q + \mathbf{s}_c'\mathbf{K}_{cc}\mathbf{s}_c + \sigma_v^2) \end{bmatrix}. \tag{20}$$

Here $\mathbf{K}_{qq} = E(\mathbf{u}_{qt}\mathbf{u}_{qt}')$, $\mathbf{K}_{cc} = E(\mathbf{u}_{ct}\mathbf{u}_{ct}')$, $\mathbf{K}_{qc} = E(\mathbf{u}_{qt}\mathbf{u}_{ct}') = \mathbf{K}_{cq}'$, $\mathbf{\Sigma}_\chi = E(\mathbf{u}_{\chi t}\mathbf{u}_{\chi t}')$, $\mathbf{\Sigma}_\psi = E(\mathbf{u}_{\psi t}\mathbf{u}_{\psi t}')$, and $\sigma_v^2 = E(u_{vt}^2)$.

4.3 Identification when supply and demand shocks are uncorrelated with each other

A common assumption in structural VARs is that the underlying structural shocks are uncorrelated with each other. If demand shocks are uncorrelated with supply shocks, $\mathbf{K}_{cq} = \mathbf{0}$ and $\mathbf{D}$ simplifies to

$$\mathbf{D} = \begin{bmatrix} \mathbf{K}_{qq} + \boldsymbol{\Sigma}_{\chi} & \mathbf{0}_{nm} & -\alpha \mathbf{K}_{qq}\mathbf{s}_q \\ \mathbf{0}_{mn} & \mathbf{K}_{cc} + \boldsymbol{\Sigma}_{\psi} & \alpha \mathbf{K}_{cc}\mathbf{s}_c \\ -\alpha \mathbf{s}_q' \mathbf{K}_{qq} & \alpha \mathbf{s}_c' \mathbf{K}_{cc} & \alpha^2(\mathbf{s}_q'\mathbf{K}_{qq}\mathbf{s}_q + \mathbf{s}_c'\mathbf{K}_{cc}\mathbf{s}_c + \sigma_v^2) \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{qq} & \mathbf{0}_{nm} & \mathbf{D}_{qp} \\ \mathbf{0}_{mn} & \mathbf{D}_{cc} & \mathbf{D}_{cp} \\ \mathbf{D}_{pq} & \mathbf{D}_{pc} & D_{pp} \end{bmatrix}. \tag{21}$$

Recalling (17), the (2,1) block of (21) implies that

$$E(\boldsymbol{\epsilon}_{qt} - \boldsymbol{\phi}_q \varepsilon_{pt})(\boldsymbol{\epsilon}_{ct} - \boldsymbol{\phi}_c \varepsilon_{pt})' = \mathbf{0}_{nm}.$$

When $n = m = 1$, this is a single equation to determine the two unknowns ϕ_q and ϕ_c, and there would be an infinite number of combinations of ϕ_q and ϕ_c that would imply uncorrelated supply and demand shocks consistent with the data. For example, with no inventory changes or measurement error, $q_t = c_t$ and there is a continuum of values of (ϕ_q, ϕ_c) satisfying

$$\omega_{qq} - \phi_c \omega_{qp} - \phi_q \omega_{qp} + \phi_q \phi_c \omega_{pp} = 0$$

for any specified $\omega_{qq}, \omega_{pp}, \omega_{qp}$. This is a very familiar identification problem, as discussed for example in Baumeister and Hamilton (2015). However, when n and m are greater than one, there is additional information in the correlations across regions, and we might hope to estimate the $n + m$ elements of $\boldsymbol{\phi} = (\boldsymbol{\phi}_q', \boldsymbol{\phi}_c')'$ using the nm equations

$$T^{-1}\textstyle\sum_{t=1}^{T}(\hat{\boldsymbol{\epsilon}}_{qt} - \boldsymbol{\phi}_q \hat{\varepsilon}_{pt})(\hat{\boldsymbol{\epsilon}}_{ct} - \boldsymbol{\phi}_c \hat{\varepsilon}_{pt})' = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{nm} & -\boldsymbol{\phi}_q \end{bmatrix} \hat{\boldsymbol{\Omega}} \begin{bmatrix} \mathbf{0}_{nm} \\ \mathbf{I}_m \\ -\boldsymbol{\phi}_c' \end{bmatrix} = \mathbf{0}_{nm}. \tag{22}$$

For example, if we impose that all producers have the same supply elasticity ($\phi_{qi} = \phi_q$) and all consumers have the same demand elasticity ($\phi_{cj} = \phi_c$), this gives us a system of nm equations in the two unknowns ϕ_q and ϕ_c. More generally, with $\boldsymbol{\phi}_q$ and $\boldsymbol{\phi}_c$ unrestricted $(n \times 1)$ and $(m \times 1)$ vectors, counting equations and unknowns, one might suppose that the model is just identified when $n = m = 2$ and overidentified for larger n, m. However, it turns out that even when $nm \geq n + m$, equations (22) by themselves would allow us to estimate at most $n + m - 1$ of the elements in $(\boldsymbol{\phi}_q', \boldsymbol{\phi}_c')'$, as shown in the appendix. We could use (22) along with restrictions such as assuming that some of the elasticities are common across countries to estimate the elasticities. In the remainder of this subsection we show how in such cases we could use other elements of $\hat{\mathbf{\Omega}}$ to estimate the rest of the structural parameters. In the next subsection we present an alternative approach that does not rely on the assumption that demand and supply shocks are uncorrelated with each other.

Given $\boldsymbol{\phi}_q$ and $\boldsymbol{\phi}_c$, we would know the value of $\mathbf{A}$ and would be able to construct $\hat{\mathbf{D}} = \mathbf{A}\hat{\mathbf{\Omega}}\mathbf{A}'$. Suppose we assume that measurement errors have the same variance and are uncorrelated across countries: $\mathbf{\Sigma}_\chi = \sigma^2_\chi \mathbf{I}_n$ and $\mathbf{\Sigma}_\psi = \sigma^2_\psi \mathbf{I}_m$. From (21), we can then estimate the off-diagonal elements of $\mathbf{K}_{qq}$ and $\mathbf{K}_{cc}$ from the off-diagonal elements of $\hat{\mathbf{D}}_{qq}$ and $\hat{\mathbf{D}}_{cc}$, respectively. The diagonal elements of $\hat{\mathbf{D}}_{qq}$ can tell us the sum of $\sigma^2_\chi \mathbf{I}_n$ plus the diagonal elements of $\mathbf{K}_{qq}$. We can then use the observed covariances in $\hat{\mathbf{D}}_{qp}$ to obtain a separate estimate of diagonal elements of $\mathbf{K}_{qq}$ up to a constant of proportionality α, where the single remaining unknown element in α is the inventory elasticity ϕ_v. Combining these two sources of information, we can thus use the n diagonal elements in $\hat{\mathbf{D}}_{qq}$ along with the n elements of $\hat{\mathbf{D}}_{qp}$ to estimate the n diagonal elements of $\mathbf{K}_{qq}$ along with the two scalars σ^2_χ and ϕ_v. For example, suppose that the value of $\boldsymbol{\phi}'_q, \boldsymbol{\phi}'_c$ that makes $\hat{\mathbf{D}}_{qc}$ small turns out to predict a bigger correlation between supply shocks and prices than we see in the data, that is, a bigger value for $\mathbf{D}_{pq}$ in (21) than we observe in $\hat{\mathbf{D}}_{pq}$. This would be interpreted either as evidence of significant errors in measuring production ($\sigma^2_\chi > 0$) or as a negative value for ϕ_v. A negative value for ϕ_v would mean that inventory adjustments help mitigate the effects of supply shocks on the price of oil, accounting for why we observe such a small correlation between $\hat{\boldsymbol{\epsilon}}_{qt} - \boldsymbol{\phi}_q \hat{\varepsilon}_{pt}$ and $\hat{\varepsilon}_{pt}$. On the other hand, if the observed correlation between supply shocks and price is stronger than predicted by the structural model, that would be interpreted as evidence that $\phi_v > 0$, meaning that inventory changes magnify the effects of supply shocks. Similarly we can use the m diagonal elements of $\hat{\mathbf{D}}_{cc}$ along with the m covariances $\hat{\mathbf{D}}_{cp}$ to estimate the m diagonal elements of $\mathbf{K}_{cc}$ along with σ^2_ψ and to provide additional information about ϕ_v. We can then uncover the final unknown magnitude σ^2_v from $\hat{D}_{pp}$. The optimal way to use the information in the structural model is to find the value for the full vector of structural parameters that achieves the highest value for the likelihood of the observed data.

4.4 A specification with correlated supply and demand shocks

We find in our dataset that the overidentifying restrictions in (22) are rejected, leading us to conclude that, contrary to what is often assumed, there must be some correlation between shocks to oil demand and supply. Gabaix and Koijen (2020) raised the possible desirability of allowing the supply shock to producing region i to be correlated with the demand shock to consuming region j if i and j represent the same region. In our example that would mean allowing the U.S. supply shock u_{q1t} to be correlated with the U.S. demand shock u_{c1t} and the rest-of-world supply shock u_{qnt} to be correlated with u_{cmt}. We find our dataset is better described by allowing a single global factor f_t to potentially influence all the shocks. For example, a global economic downturn could affect all producers and consumers. In addition, we allow for a second factor f_{qt} contributing to the supply shocks across different countries, but which is presumed to influence demand only through its effect on price, and a third factor f_{ct} common to all countries' demand shocks but mattering for supply only through its effects on price. The specification is then

$$\mathbf{u}_{qt} = \mathbf{h}_q f_t + \boldsymbol{\gamma}_q f_{qt} + \boldsymbol{\eta}_{qt} \tag{23}$$

$$\mathbf{u}_{ct} = \mathbf{h}_c f_t + \boldsymbol{\gamma}_c f_{ct} + \boldsymbol{\eta}_{ct}. \tag{24}$$

Here $\boldsymbol{\gamma}_c$ is an $(m \times 1)$ vector summarizing how the common global demand shock f_{ct} shows up in each individual consuming country. Note that this is a strict generalization of (12) where it was assumed that $\gamma_{cj} = 1$ for all j, meaning the global demand shock affected each country the same way. Here we normalize the variance of f_{ct} to be one and allow for the possibility that the global demand shock affects each country differently. We assume that the factors f_t, f_{qt}, and f_{ct} are uncorrelated with each other and each have unit variance. Thus, for example, a one-standard-deviation increase in the global demand shock f_{ct} raises the consumption of country j by γ_{cj}. We normalize these weights by imposing that the loadings on the supply or demand factors are orthogonal to the loadings on the global factor: $\mathbf{h}_q'\boldsymbol{\gamma}_q = \mathbf{h}_c'\boldsymbol{\gamma}_c = 0$. The appendix describes how we implemented this mechanically. The $(n \times 1)$ vector $\boldsymbol{\eta}_{qt}$ represents purely idiosyncratic shocks to supply that are uncorrelated with any of the three global factors. We assume that these are uncorrelated across countries but allow the variances of purely idiosyncratic shocks to differ across countries. That is, we assume $E(\boldsymbol{\eta}_{qt}\boldsymbol{\eta}_{qt}') = \boldsymbol{\Sigma}_q$, where $\boldsymbol{\Sigma}_q$ is a diagonal $(n \times n)$ matrix whose ith diagonal element measures the variance of purely idiosyncratic shocks to supply. Similarly $E(\boldsymbol{\eta}_{ct}\boldsymbol{\eta}_{ct}') = \boldsymbol{\Sigma}_c$, a diagonal $(m \times m)$ matrix.

For this model, the matrix in (20) becomes

$$\mathbf{D} = \begin{bmatrix} \mathbf{h}_q\mathbf{h}_q' + \boldsymbol{\gamma}_q\boldsymbol{\gamma}_q' + \boldsymbol{\Sigma}_q + \sigma_\chi^2\mathbf{I}_n & \mathbf{h}_q\mathbf{h}_c' & -\alpha\left(\mathbf{h}_q\mathbf{h}_q' + \boldsymbol{\gamma}_q\boldsymbol{\gamma}_q' + \boldsymbol{\Sigma}_q\right)\mathbf{s}_q + \alpha\mathbf{h}_q\mathbf{h}_c'\mathbf{s}_c \\ \mathbf{h}_c\mathbf{h}_q' & \mathbf{h}_c\mathbf{h}_c' + \boldsymbol{\gamma}_c\boldsymbol{\gamma}_c' + \boldsymbol{\Sigma}_c + \sigma_\psi^2\mathbf{I}_m & -\alpha\mathbf{h}_c\mathbf{h}_q'\mathbf{s}_q + \alpha\left(\mathbf{h}_c\mathbf{h}_c' + \boldsymbol{\gamma}_c\boldsymbol{\gamma}_c' + \boldsymbol{\Sigma}_c\right)\mathbf{s}_c \\ -\alpha\mathbf{s}_q'\left(\mathbf{h}_q\mathbf{h}_q' + \boldsymbol{\gamma}_q\boldsymbol{\gamma}_q' + \boldsymbol{\Sigma}_q\right) + \alpha\mathbf{s}_c'\mathbf{h}_c\mathbf{h}_q' & -\alpha\mathbf{s}_q'\mathbf{h}_q\mathbf{h}_c' + \alpha\mathbf{s}_c'\left(\mathbf{h}_c\mathbf{h}_c' + \boldsymbol{\gamma}_c\boldsymbol{\gamma}_c' + \boldsymbol{\Sigma}_c\right) & \alpha^2\left[\mathbf{s}_q'\left(\mathbf{h}_q\mathbf{h}_q' + \boldsymbol{\gamma}_q\boldsymbol{\gamma}_q' + \boldsymbol{\Sigma}_q\right)\mathbf{s}_q - 2\mathbf{s}_c'\mathbf{h}_c\mathbf{h}_q'\mathbf{s}_q + \mathbf{s}_c'\left(\mathbf{h}_c\mathbf{h}_c' + \boldsymbol{\gamma}_c\boldsymbol{\gamma}_c' + \boldsymbol{\Sigma}_c\right)\mathbf{s}_c + \sigma_v^2\right] \end{bmatrix}. \tag{25}$$

In the model in the previous subsection, the elasticities $\boldsymbol{\phi}_q$ and $\boldsymbol{\phi}_c$ were identified from (22), which required that all linear combinations of $\boldsymbol{\epsilon}_{qt} - \boldsymbol{\phi}_q\varepsilon_{pt}$ be uncorrelated with all linear combinations of $\boldsymbol{\epsilon}_{ct} - \boldsymbol{\phi}_c\varepsilon_{pt}$ where $\boldsymbol{\epsilon}_{qt}$, $\boldsymbol{\epsilon}_{ct}$, and ε_{pt} are the observable errors one would make in forecasting $\mathbf{q}_t$, $\mathbf{c}_t$, and p_t one month in advance. The model here generalizes this assumption, looking for values of $\boldsymbol{\phi}_q$ and $\boldsymbol{\phi}_c$ for which all but one linear combination of the implied supply and demand shocks are uncorrelated with each other. The elasticities are also chosen to imply that off-diagonal elements of $\mathbf{D}_{qq}$ and $\mathbf{D}_{cc}$ can be summarized by single supply and demand factors beyond the common factor in the $\mathbf{D}_{qc}$ covariance. The values of $\boldsymbol{\Sigma}_q$, σ_χ^2, and ϕ_v are then inferred from the diagonal elements of $\mathbf{D}_{qq}$ and the covariances between $\boldsymbol{\epsilon}_{qt} - \boldsymbol{\phi}_q\varepsilon_{pt}$ and ε_{pt}. The latter correlation is nonzero because a shock to country i supply has an effect on the world price based on its share s_{qi} in global oil production. The values of $\boldsymbol{\Sigma}_c$ and σ_ψ^2 (and additional information about all the other parameters) are informed by the diagonal elements of $\mathbf{D}_{cc}$ and the covariances between $\boldsymbol{\epsilon}_{ct} - \boldsymbol{\phi}_c\varepsilon_{pt}$ and ε_{pt}.

4.5 Maximum likelihood estimation

Up to this point we have been discussing identification in terms of valid instruments and whether structural magnitudes could be inferred from observed correlations, following most of the earlier literature. However, it is well known that instrumental variables is not the optimal way to make use of such information. As a general estimation method, maximum likelihood uses all the identifying information in the optimal way to produce consistent parameter estimates with the smallest asymptotic variance. This is the method of estimation used in our paper.

For Gaussian structural shocks, the model in the previous subsection implies a value for the log likelihood of

$$\ell(\boldsymbol{\theta}) = -[T(n+m+1)/2]\log(2\pi) - (T/2)\log|\boldsymbol{\Omega}(\boldsymbol{\theta})| - (T/2)\text{trace}\left(\boldsymbol{\Omega}(\boldsymbol{\theta})^{-1}\hat{\boldsymbol{\Omega}}\right). \tag{26}$$

Here $\hat{\mathbf{\Omega}}$ is the OLS estimate $T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\epsilon}}_t\hat{\boldsymbol{\epsilon}}_t'$, $\hat{\boldsymbol{\epsilon}}_t = \mathbf{y}_t - \hat{\mathbf{\Pi}}\mathbf{x}_{t-1}$, and

$$\hat{\mathbf{\Pi}} = \left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_{t-1}'\right)\left(\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}.$$

The matrix $\mathbf{\Omega}(\boldsymbol{\theta})$ is the forecast-error variance matrix that is implied by the vector of structural parameters,

$$\mathbf{\Omega}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})^{-1}\mathbf{D}(\boldsymbol{\theta})\left[\mathbf{A}(\boldsymbol{\theta})^{-1}\right]'$$

$$\mathbf{A}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{nm} & \boldsymbol{\phi}_q \\ \mathbf{0}_{mn} & \mathbf{I}_m & \boldsymbol{\phi}_c \\ \mathbf{0}_{1n} & \mathbf{0}_{1m} & 1 \end{bmatrix},$$

where $\mathbf{D}(\boldsymbol{\theta})$ is the matrix in (25). The vector $\boldsymbol{\theta}$ contains $4(m+n)+2$ unknown structural parameters. The $4n$ supply parameters consist of the n supply elasticities $\boldsymbol{\phi}_q$, the n idiosyncratic supply variances $\mathbf{\Sigma}_q$, the n loadings of supply $\mathbf{h}_q$ on the global factor f_t, the $(n-1)$ loadings of supply $\boldsymbol{\omega}_q$ on the supply factor f_{qt} (which imposes the normalization condition $\mathbf{h}_q'\boldsymbol{\gamma}_q = 0$ as described in the appendix), and the variance σ_χ^2 of the error in measuring production for each country. The $4m$ demand parameters consist analogously of $\boldsymbol{\phi}_c, \mathbf{\Sigma}_c, \mathbf{h}_c, \boldsymbol{\omega}_c, \sigma_\psi^2$. In addition there are the two inventory demand parameters, which consist of the elasticity ϕ_v and variance of the inventory shock σ_v^2. Our proposal is to estimate $\boldsymbol{\theta}$ by maximizing the log likelihood (26) subject to the restrictions $\phi_{qi} \geq 0$ and $\phi_{cj} \leq 0$.

Another benefit of maximum likelihood estimation is that it gives an immediate test of the overidentifying assumptions. The structural model is a restricted version of an unconstrained VAR, which sets $\mathbf{\Omega}(\boldsymbol{\theta}) = \hat{\mathbf{\Omega}}$ achieving a value for the log likelihood of

$$\ell(\hat{\boldsymbol{\theta}}_{unrestricted}) = -[T(n+m+1)/2][1+\log(2\pi)] - (T/2)\log\left|\hat{\mathbf{\Omega}}\right|. \quad (27)$$

The unconstrained model estimates $(n+m+1)(n+m+2)/2$ parameters in the covariance matrix $\mathbf{\Omega}$. For our baseline model with $n=m=4$, there are 34 elements in the structural parameter vector $\boldsymbol{\theta}$ compared with 45 parameters for the reduced-form VAR. We can test the model assumptions by comparing the value of (26) with the value of (27).

4.6 Impulse-response functions

We can use equations (18), (19), (23), and (24) to rewrite (17) as

$$\boldsymbol{\epsilon}_t = \mathbf{A}^{-1}\mathbf{u}_t$$

$$\underset{[(n+m+1)\times(n+m+1)]}{\mathbf{A}^{-1}} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{nm} & \boldsymbol{\phi}_q \\ \mathbf{0}_{mn} & \mathbf{I}_m & \boldsymbol{\phi}_c \\ \mathbf{0}_{1n} & \mathbf{0}_{1m} & 1 \end{bmatrix}$$

$$\mathbf{u}_t = \begin{bmatrix} \mathbf{h}_q f_t + \boldsymbol{\gamma}_q f_{qt} + \boldsymbol{\eta}_{qt} + \mathbf{u}_{\chi t} \\ \mathbf{h}_c f_t + \boldsymbol{\gamma}_c f_{ct} + \boldsymbol{\eta}_{ct} + \mathbf{u}_{\psi t} \\ \alpha \mathbf{s}_c'(\mathbf{h}_c f_t + \boldsymbol{\gamma}_c f_{ct} + \boldsymbol{\eta}_{ct}) - \alpha \mathbf{s}_q'(\mathbf{h}_q f_t + \boldsymbol{\gamma}_q f_{qt} + \boldsymbol{\eta}_{qt}) + \alpha u_{vt} \end{bmatrix}.$$

We can use these equations to calculate the impact effect of any structural shock u_{kt} on the $(n+m+1)\times 1$ vector of observed variables $\mathbf{y}_t$ at time t using $\partial\boldsymbol{\epsilon}_t/\partial u_{kt} = \mathbf{A}^{-1}\partial\mathbf{u}_t/\partial u_{kt}$. For example, the effect of a one-percent increase in the supply from country i alone resulting from a one-unit increase in η_{qit} would be

$$\frac{\partial \boldsymbol{\epsilon}_t}{\partial \eta_{qit}} = \mathbf{A}^{-1}\begin{bmatrix} \mathbf{e}_i^{(n)} \\ \mathbf{0} \\ -\alpha \mathbf{s}_q' \mathbf{e}_i^{(n)} \end{bmatrix}$$

where $\mathbf{e}_i^{(n)}$ is the ith column of $\mathbf{I}_n$. Likewise, the effect of a one-percent increase in demand from region j alone is

$$\frac{\partial \boldsymbol{\epsilon}_t}{\partial \eta_{cjt}} = \mathbf{A}^{-1}\begin{bmatrix} \mathbf{0} \\ \mathbf{e}_j^{(m)} \\ \alpha \mathbf{s}_c' \mathbf{e}_j^{(m)} \end{bmatrix}$$

for $\mathbf{e}_j^{(m)}$ the jth column of $\mathbf{I}_m$.The effect of a one-standard-deviation increase in the global demand factor is

$$\frac{\partial \boldsymbol{\epsilon}_t}{\partial f_{ct}} = \mathbf{A}^{-1}\begin{bmatrix} \mathbf{0} \\ \boldsymbol{\gamma}_c \\ \alpha \mathbf{s}_c' \boldsymbol{\gamma}_c \end{bmatrix}. \tag{28}$$

The effect of a one-percent increase in inventory demand is

$$\frac{\partial \boldsymbol{\epsilon}_t}{\partial u_{vt}} = \mathbf{A}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \alpha \end{bmatrix}.$$

Let $\boldsymbol{\Psi}_s$ denote the usual reduced-form dynamic multiplier $\boldsymbol{\Psi}_s = \partial \mathbf{y}_{t+s} / \partial \boldsymbol{\epsilon}_t'$. We can calculate structural dynamic multipliers defined as the answers to the following question: if there is a structural shock to u_{kt}, how does this cause us to change our forecast of $\mathbf{y}_{t+s}$,

$$\frac{\partial E(\mathbf{y}_{t+s} | \mathbf{y}_t, \mathbf{y}_{t-1}, \ldots \mathbf{y}_{t-r+1})}{\partial u_{kt}} = \boldsymbol{\Psi}_s \frac{\partial \boldsymbol{\epsilon}_t}{\partial u_{kt}}.$$

To get this, we just plug in one of the above expressions for $\partial \boldsymbol{\epsilon}_t / \partial u_{kt}$. For example,

$$\frac{\partial E(\mathbf{y}_{t+s} | \mathbf{y}_t, \mathbf{y}_{t-1}, \ldots \mathbf{y}_{t-r+1})}{\partial f_{ct}} = \boldsymbol{\Psi}_s \mathbf{A}^{-1} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\gamma}_c \\ \alpha \mathbf{s}_c' \boldsymbol{\gamma}_c \end{bmatrix}. \tag{29}$$

We can calculate confidence bands for these estimates as described in the appendix.

5 Empirical results

We assembled year-over-year production growth rates for the U.S., Saudi Arabia, Russia and the rest of the world, year-over-year consumption growth rates for the U.S., Japan, Europe, and the rest of the world, and monthly growth rates for the real price of oil. The data are monthly with the dependent variable including observations from $t =$ 1975:M1 through 2020:M2 and 2022:M2 through 2023:M2.

5.1 Parameter estimates

The parameter values that maximize the likelihood (26) are reported in Table 1. We estimate a short-run price elasticity of oil supply of 0.26 for Saudi Arabia and 0.02 to 0.04 for other countries. These are all estimated to be positive without the need to impose sign constraints. Using equation (3), these estimates imply an overall world oil short-run supply elasticity of

$$\phi_q = \textstyle\sum_{i=1}^{n} s_{qi} \phi_{qi} = 0.064. \tag{30}$$

This is very similar to the estimates of Caldara, Cavallo and Iacoviello (2019, Table 3) of 0.05 to 0.08 and a little below the estimate in Baumeister and Hamilton (2019) of 0.15. Our detailed estimates also support the conclusions of previous researchers including Pierru, Smith and Zamrik (2018), Caldara, Cavallo and Iacoviello (2019), Alonso-Alvarez, Di Nino and Venditti (2022), and Almutairi, Pierru and Smith (2023) that OPEC production is much more responsive to price than production in other countries.

Our estimates of the short-run price elasticity of demand range from −0.02 for Japan to −0.22 for Europe. Again these are all estimated to be negative without imposing any sign constraints. From equation (4), our estimates imply a global short-run price elasticity of demand of

$$\phi_c = \sum_{j=1}^{m} s_{cj}\phi_{cj} = -0.139. \tag{31}$$

This is a little more responsive than the estimates of −0.03 to −0.08 in Cooper (2003), Caldara, Cavallo and Iacoviello (2019, Table 3), and Pierru, Smith and Zamrik (2018) and a little less responsive than the estimates obtained by Alonso-Alvarez, Di Nino and Venditti (2022) (−0.28), and Baumeister and Hamilton (2019) (−0.35).

Shocks to Saudi Arabian production (whose standard deviation is $\sigma_{q,Saudi} = 6.3$) are significantly larger than those to other producing countries and also larger than any shocks to consumption demand. The estimated variance of the measurement error in production data is smaller than the variance of idiosyncratic shocks to true production for any country. We find no evidence of measurement error in consumption data. The variance of shocks to inventory demand is comparable in magnitude to the variances of demand shocks for individual countries, and inventory demand is more responsive to price than is the product demand from any individual country. The feature of the data leading to the estimate of $\hat{\phi}_v < 0$ is the observation that the correlations between supply shocks and price and between demand shocks and price are observed to be smaller than they would be in the absence of adjustment of inventories. We interpret these estimates as consistent with the view that inventory investment responds to price changes in a stabilizing way. A shortfall in supply from any producing country is partially met by selling out of inventories.

The global factor f_t shows up primarily as an increase in demand from the rest of the world and an increase in production from Saudi Arabia and the rest of the world. One possibility is that this represents a direct response of OPEC to strong global demand that is not mediated through price changes. We found that estimates of a global supply factor f_{qt} did not contribute much to the fit to the data, and

Table 1: Maximum likelihood estimates

Parameter	MLE	Std err
$\phi_{q,US}$	0.019	(0.017)
$\phi_{q,Saudi}$	0.259	(0.056)
$\phi_{q,Russia}$	0.029	(0.011)
$\phi_{q,ROW}$	0.043	(0.029)
$\phi_{c,US}$	-0.094	(0.031)
$\phi_{c,Japan}$	-0.018	(0.037)
$\phi_{c,Europe}$	-0.225	(0.045)
$\phi_{c,ROW}$	-0.161	(0.045)
$\sigma_{q,US}$	2.335	(0.178)
$\sigma_{q,Saudi}$	6.348	(0.260)
$\sigma_{q,Russia}$	1.224	(0.308)
$\sigma_{q,ROW}$	0.825	(0.786)
$\sigma_{c,US}$	1.937	(0.106)
$\sigma_{c,Japan}$	3.065	(0.120)
$\sigma_{c,Europe}$	3.490	(0.152)
$\sigma_{c,ROW}$	2.472	(0.183)
σ_{χ}	0.915	(0.404)
σ_{ψ}	0	—
$h_{q,US}$	-0.036	(0.139)
$h_{q,Saudi}$	2.236	(0.493)
$h_{q,Russia}$	0.103	(0.084)
$h_{q,ROW}$	1.661	(0.297)
$h_{c,US}$	-0.126	(0.119)
$h_{c,Japan}$	-0.313	(0.151)
$h_{c,Europe}$	-0.190	(0.141)
$h_{c,ROW}$	1.075	(0.265)
ϕ_v	-0.314	(0.060)
σ_v	2.555	(0.337)
$\gamma_{c,US}$	1.415	(0.444)
$\gamma_{c,Japan}$	1.548	(0.525)
$\gamma_{c,Europe}$	2.044	(0.564)
$\gamma_{c,ROW}$	0.967	(0.364)
α	1.932	(0.138)
ϕ_q	0.064	(0.021)
ϕ_c	-0.139	(0.037)

Notes to Table 1. The four elements of $\boldsymbol{\gamma}_c$ were not estimated directly but were calculated from the three elements of $\boldsymbol{\omega}_c$ (not reported in the table) along with the four elements of $\mathbf{h}_c$ reported in the table using equations (A3) and (A2). Standard errors for $\boldsymbol{\gamma}_c$ were calculated by simulating draws from the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ as a byproduct of the algorithm used to calculate confidence bands for impulse-response functions. The values of α, ϕ_q and ϕ_c were not estimated directly but were inferred from $\hat{\boldsymbol{\theta}}$ using equations (15), (30), and (31) with standard errors for $\hat{\alpha}$, $\hat{\phi}_q$, and $\hat{\phi}_c$ obtained by simulation.

the estimates reported in Table 1 impose $\boldsymbol{\gamma}_q = \mathbf{0}$.[2] The coefficients $\boldsymbol{\gamma}_c$ on the global demand factor are quite similar across countries. A one-standard-deviation increase in f_{ct} leads to a 1-2% increase in oil demand everywhere in the world. These parameter estimates are consistent with the simpler specification of the role of global demand factors in equation (12).

5.2 Standard errors and hypothesis tests

The estimates reported in Table 1 do not satisfy the regularity conditions that are used to derive the usual asymptotic results for maximum likelihood estimation because the MLE of σ_ψ^2 is at the boundary zero of the allowable parameter space. We can deal with this issue by treating this parameter as fixed rather than estimated and calculating the matrix of second derivatives of (26) with respect to the 30 free parameters. The resulting standard errors are reported in Table 1. The estimated Saudi and Russian supply elasticities are statistically significantly different from zero, but we could not reject the hypothesis that the U.S. or rest-of-world supply elasticity is zero. All demand elasticities other than Japan are statistically significantly different from zero, as is the inventory demand elasticity ϕ_v. We fail to reject the null hypothesis that the aggregate factor f_t has no effect apart from that on $u_{q,Saudi}$, $u_{q,ROW}$, and $u_{c,ROW}$.

Two times the difference between the log likelihood achieved by the unrestricted reduced-form model (27) and the constrained structural model (26) is 23.39. The unconstrained model has $45 - 30 = 15$ additional parameters, which yields a p-value of 0.076 using an asymptotic $\chi^2(15)$ approximation to the likelihood ratio test. We therefore would not reject the null hypothesis that the overidentifying restrictions are all valid.[3] In addition to the 45 parameters used to fit the variance matrix $\boldsymbol{\Omega}$, there are another $k = 1 + r(n + m + 1) = 109$ parameters estimated for each row of $\boldsymbol{\Pi}$ in the unconstrained VAR. Adjusting to correct for small-sample bias[4] would result in a test statistic of

$$\frac{2(T-k)}{T}\left\{\ell(\hat{\boldsymbol{\theta}}_{unrestricted}) - \ell(\hat{\boldsymbol{\theta}}_{MLE})\right\} = 18.80$$

for which the $\chi^2(15)$ p-value is 0.22. We conclude that the overidentifying assumptions

[2]Maximum likelihood estimates when $\boldsymbol{\gamma}_q$ is unconstrained are reported in Table A.1 in the appendix, and are very similar to the estimates in Table 1.

[3]If we were to count $\hat{\sigma}_\psi^2 = 0$ as another estimated parameter it would bring the degrees of freedom to 14. Using a $\chi^2(14)$ approximation to the distribution of the likelihood ratio test can not be justified by the formal asymptotic theory, but gives a tighter distribution for evaluating statistical significance. Using critical values from a $\chi^2(14)$ distribution results in a p-value of 0.054, which just fails to reject.

[4]See Sims (1980, p. 17) and Hamilton (1994, p. 297).

Table 2: Impact effects of a global demand shock

Variable	with estimated ϕ_v				with $\phi_v = 0$			
	as % of country			% of world	as % of country			% of world
	direct effect (1)	response to price (2)	net effect (3)	net effect (4)	direct effect (5)	response to price (6)	net effect (7)	net effect (8)
p	2.330				5.935			
q_{US}	0	0.043	0.043	0.005	0	0.110	0.110	0.013
q_{Saudi}	0	0.604	0.604	0.072	0	1.538	1.538	0.185
q_{Russia}	0	0.068	0.068	0.010	0	0.173	0.173	0.026
q_{ROW}	0	0.101	0.101	0.061	0	0.256	0.256	0.156
q				0.149				0.380
c_{US}	1.415	-0.218	1.197	0.299	1.415	-0.555	0.860	0.215
c_{Japan}	1.548	-0.043	1.505	0.105	1.548	-0.108	1.440	0.101
c_{Europe}	2.044	-0.524	1.520	0.122	2.044	-1.334	0.710	0.057
c_{ROW}	0.967	-0.375	0.592	0.355	0.967	-0.954	0.013	0.008
c				0.882				0.380
v				0.733				0.000

Notes to Table 2. Impact effects of a one-standard-deviation increase in the global demand factor f_{ct} both given the historical average response of inventories (columns 1-4) and under the counterfactual of no adjustment of inventories (columns 5-8).

appear to be consistent with the correlations that are observed in these data.

6 Applications

Now we use our model to analyze the effects of certain structural shocks.

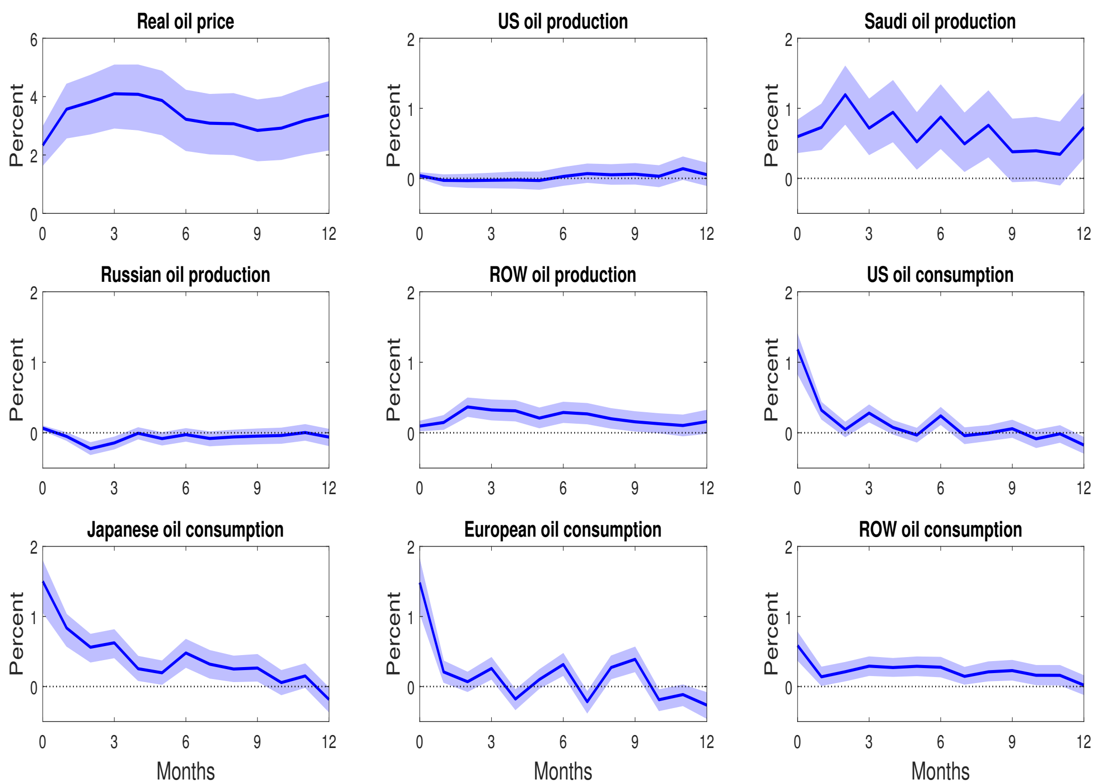
6.1 Example 1: The effects of a global demand shock

We first examine the effects of a one-standard-deviation shock to the global demand factor f_{ct}. This raises demand for country j by γ_{cj}, which is around 1-2% for every country. From the last row of equation (28), this leads to an immediate increase in the price of oil of $\alpha \mathbf{s}_c' \boldsymbol{\gamma}_c$, which equals 2.3%. These immediate impact effects are summarized in column 1 of Table 2.

The change in price in turn induces responses of quantities produced and consumed. Column 2 of Table 2 calculates the size of these responses by multiplying the price change 2.330 by the respective elasticities ϕ_{qi} or ϕ_{cj}. Saudi oil production increases by about 0.6% in response to the higher world demand. The price increase also substantially reduces the effect of the demand increase on realized consumption. The net effect (column 3) is the sum of columns 1 and 2.

Figure 2 plots the dynamic effects of the shock calculated using expression (29). Production of oil from Saudi Arabia and the rest of the world continue to climb in the first few months following the shock. These estimates support the conclusion of Almutairi, Pierru and Smith (2023) that Saudi Arabia and OPEC play a major role in stabilizing the world oil market. The effects of the shock on consumption of individual countries dies off relatively quickly.

Figure 2: Dynamic effects of a global demand shock



Notes to Figure 2. Dynamic effects of a one-standard-deviation increase in the global demand factor f_{ct} assuming the historical average response of inventories. First panel plots the cumulative effect on 100 times the log of the real price of oil. Other panels plot year-over-year changes of quantities as a percent of that country's production or consumption. Shaded regions indicate 68% confidence intervals.

Column 4 of Table 2 restates the magnitudes as a percent of the world total by multiplying the entries in column 3 by s_{qi} or s_{cj}. The total initial gains in production (the sum of the first four rows of column 4) only amount to a 0.15% increase in global production, compared with a 0.88% increase in consumption. Thus sales out of inventory play a major role in meeting the temporarily strong demand. Columns

5-8 of Table 2 report what the response to the demand shock would be if there were no changes in inventories, which can be calculated by setting $\phi_v = 0$. The immediate impact on price in that scenario would be $\tilde{\alpha}\mathbf{s}_c'\boldsymbol{\gamma}_c$ where $\tilde{\alpha} = 1/(\boldsymbol{\phi}_q'\mathbf{s}_q - \boldsymbol{\phi}_c'\mathbf{s}_c) = 4.92$. In this counterfactual, the demand increase would lead to a 5.9% increase in prices, almost three times as large as in column 1. If there is no inventory response, the increase in production (0.38% of world supply in column 8 of Table 2) would of necessity exactly equal the increase in world consumption. Comparing column 8 with column 4, most of the balancing in this case comes on the demand side, with the effect of price increases undoing much of the original stimulus to demand.

6.2 Example 2: The effects of a 50% decrease in Russian production

As a second example we examine the consequences if exogenous political events were to lead to a 50% decline in $u_{q,Russia}$. This would represent a loss of over 5 million barrels per day. For this example, we use production and consumption shares as of the end of our sample (February 2023).[5] Table 3 summarizes the effect on impact. We first highlight the calculations in columns 5-8 which assume that none of the shock is offset by use of inventory drawdowns. The model estimates imply that the price of oil would increase by about a third.[6] For convenience we summarize effects on production and consumption in column 8 in units of million barrels per day. This was calculated by multiplying the number reported in column 7 by $s_{qiT}Q_T$ or $s_{cjT}Q_T$ where $Q_T = 82.3$ mb/d is total world oil production in February 2023. Increased production from Saudi Arabia and the rest of the world makes up about 1.5 mb/d of the 5.25 mb/d shortfall. A much bigger part of the adjustment comes from the demand side, with a 500,000 b/d drop in U.S. consumption and a 3 mb/d drop in rest-of-world oil consumption.

Columns 1-4 of Table 3 report the impact response if instead inventories responded to this shock the same way they did to typical historical shocks. This would require drawing down inventories by 3.2 mb/d, or nearly a hundred million barrels in the first month, which clearly is not sustainable.[7] For this reason we emphasize the analysis that assumes that 100% of the shortfall must be met through a combination of decreased consumption and increased production from other countries.

We plot the predicted dynamic effects under the assumption of no inventory

[5] These were $\mathbf{s}_{qT} = (0.15,0.12,0.13,0.60)'$ and $\mathbf{s}_{cT} = (0.20,0.04,0.05,0.71)'$.

[6] This was calculated as $-0.5 s_{q,Russia,T}/(\mathbf{s}_{qT}'\boldsymbol{\phi}_q - \mathbf{s}_{cT}'\boldsymbol{\phi}_c)$.

[7] In January 2022, the U.S. had 415 million barrels of crude oil in commercial inventories and an additional 589 million barrels in the Strategic Petroleum Reserve. Over the course of the next year, 225 million barrels were released from the SPR.

Table 3: Impact effects of a shock to Russia supply

Variable	with estimated ϕ_v as % of country direct effect (1)	response to price (2)	net effect (3)	in mb/d net effect (4)	with $\phi_v = 0$ as % of country direct effect (5)	response to price (6)	net effect (7)	in mb/d net effect (8)
p	12.430				31.185			
q_{US}	0	0.231	0.231	0.029	0	0.580	0.580	0.072
q_{Saudi}	0	3.222	3.222	0.318	0	8.084	8.084	0.798
q_{Russia}	-50	0.362	-49.638	-5.311	-50	0.909	-49.091	-5.252
q_{ROW}	0	0.536	0.536	0.265	0	1.345	1.345	0.664
q				-4.699				-3.718
c_{US}	0.000	-1.163	-1.163	-0.191	0.000	-2.919	-2.919	-0.480
c_{Japan}	0.000	-0.227	-0.227	-0.007	0.000	-0.569	-0.569	-0.019
c_{Europe}	0.000	-2.795	-2.795	-0.115	0.000	-7.011	-7.011	-0.289
c_{ROW}	0.000	-1.999	-1.999	-1.168	0.000	-5.015	-5.015	-2.930
c				-1.482				-3.718
v				3.217				0.000

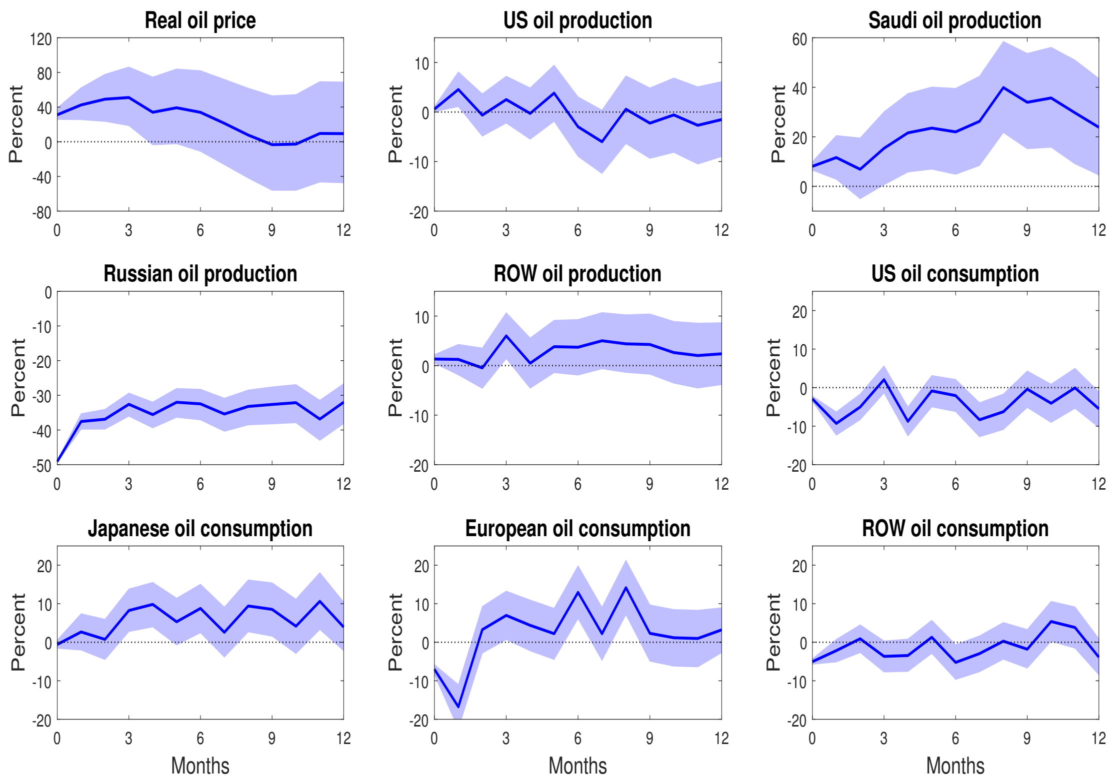
Notes to Table 3. Impact effects of a 50% cut in Russian oil production both given the historical average response of inventories (columns 1-4) and under the counterfactual of no adjustment of inventories (columns 5-8).

changes on impact in Figure 3. Saudi oil production continues to climb for many months following the shock, and is again likely to play a key role in the ability of the world to adapt to a shock like this.

7 Conclusion

The key assumption behind our approach is that correlations between country-specific supply and demand shocks can be summarized with a low-order factor structure. We showed that this assumption allows us to jointly estimate supply and demand elasticities for individual producers and consumers using maximum likelihood estimation of a structural vector autoregression, generalizing the method of granular instrumental variables developed by Gabaix and Koijen (2020). Our method could be applied in any context in which different units interact to determine a market equilibrium. We used this approach to analyze the world oil market. Our estimates of aggregate elasticities are similar to those obtained by earlier researchers who have used a variety of methods very different from ours. Our approach provides for the first time a characterization of heterogeneity in those elasticities across countries using a unified statistical model of the world oil market. Our estimates imply that variation in Saudi

Figure 3: Dynamic effects of a Russian supply shock



Notes to Figure 3. Dynamic effects of a 50% decrease in Russian oil prodution assuming no adjustment of inventories. First panel plots the cumulative effect on 100 times the log of the real price of oil. Other panels plot year-over-year changes of quantities as a percent of that country's production or consumption. Shaded regions indicate 68% confidence intervals.

Arabian production and the endogenous adjustment of inventories have historically played a key role in stabilizing the world price of oil.

References

Almutairi, Hossa, Axel Pierru, and James L Smith. 2023. "Oil Market Stabilization: The Performance of OPEC and Its Allies." *Energy Journal*, 44(6).

Alonso-Alvarez, Irma, Virginia Di Nino, and Fabrizio Venditti. 2022. "Strategic Interactions and Price Dynamics in the Global Oil Market." *Energy Economics*, 107: 105739.

Balke, Nathan S, Xin Jin, and Mine Yücel. 2020. "The Shale Revolution and the Dynamics of the Oil Market." *FRB of Dallas Working Paper.*

Banafti, Saman, and Tae-Hwy Lee. 2022. "Inferential Theory for Granular Instrumental Variables in High Dimensions." *arXiv preprint arXiv:2201.06605.*

Bartik, Timothy J. 1991. *Who Benefits From State and Local Economic Development Policies?* WE Upjohn Institute for Employment Research.

Baumeister, Christiane. 2023. "Pandemic, War, Inflation: Oil Markets at a Crossroads?" *NBER Working Paper*, 31496.

Baumeister, Christiane, and James D Hamilton. 2015. "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information." *Econometrica*, 83(5): 1963–1999.

Baumeister, Christiane, and James D Hamilton. 2019. "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks." *American Economic Review*, 109(5): 1873–1910.

Baumeister, Christiane, and James D. Hamilton. 2023. "Advances in Using Vector Autoregressions to Estimate Structural Magnitudes." *Econometric Theory.*

Blanchard, Olivier Jean, Lawrence F Katz, Robert E Hall, and Barry Eichengreen. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*, 1992(1): 1–75.

Bornstein, Gideon, Per Krusell, and Sergio Rebelo. 2023. "A World Equilibrium Model of the Oil Market." *Review of Economic Studies*, 90(1): 132–164.

Caldara, Dario, Michele Cavallo, and Matteo Iacoviello. 2019. "Oil Price Elasticities and Oil Price Fluctuations." *Journal of Monetary Economics*, 103: 1–20.

Carlino, Gerald, and Thorsten Drautzburg. 2020. "The Role of Startups for Local Labor Markets." *Journal of Applied Econometrics*, 35(6): 751–775.

Cooper, John CB. 2003. "Price Elasticity of Demand for Crude Oil: Estimates for 23 Countries." *OPEC Review*, 27(1): 1–8.

Gabaix, Xavier, and Ralph SJ Koijen. 2020. "Granular Instrumental Variables." *NBER Working Paper*, 28204.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review*, 110(8): 2586–2624.

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

Hamilton, James D. 2009. "Understanding Crude Oil Prices." *The Energy Journal*, 30(2): 179–206.

Hamilton, James D. 2023. *Vector Autoregressions.* Cambridge University Press.

Knittel, Christopher R, and Robert S Pindyck. 2016. "The Simple Economics of Commodity Price Speculation." *American Economic Journal: Macroeconomics*, 8(2): 85–110.

Lenza, Michele, and Giorgio E Primiceri. 2022. "How to Estimate a VAR After March 2020." *Journal of Applied Econometrics*, 37: 688–689.

Ng, Serena. 2021. "Modeling Macroeconomic Variations after Covid-19." *NBER Working Paper*, 29060.

Pierru, Axel, James L Smith, and Tamim Zamrik. 2018. "OPEC's Impact on Oil Price Volatility: The Role of Spare Capacity." *The Energy Journal*, 39(2): 103–122.

Qian, Eric. 2023. "Heterogeneity-Robust Granular Instruments." Princeton University.

Rothenberg, Thomas J. 1971. "Identification in Parametric Models." *Econometrica*, 39(3): 577–591.

Rothenberg, Thomas J. 1973. *Efficient Estimation with A Priori Information.* Yale University Press.

Sarto, Andres. 2022. "Recovering Macro Elasticities from Regional Data." New York University.

Schorfheide, Frank, and Dongho Song. 2021. "Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic." *NBER Working Paper*, 29535.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica*, 48: 1–48.

A Appendix

A.1 Demonstration that condition (22) alone allows estimation of at most $n+m-1$ parameters

Here we explore in more detail the information about elements of the $[(n+m)\times 1]$ vector $(\boldsymbol{\phi}_q', \boldsymbol{\phi}_c')'$ that is contained in the elements of the $(n\times m)$ matrix

$$g(\boldsymbol{\phi}_q, \boldsymbol{\phi}_c) = E(\boldsymbol{\epsilon}_{qt} - \boldsymbol{\phi}_q \varepsilon_{pt})(\boldsymbol{\epsilon}_{ct} - \boldsymbol{\phi}_c \varepsilon_{pt})' = \boldsymbol{\Omega}_{qc} - \boldsymbol{\phi}_q \boldsymbol{\Omega}_{pc} - \boldsymbol{\Omega}_{qp}\boldsymbol{\phi}_c' + \boldsymbol{\phi}_q \Omega_{pp} \boldsymbol{\phi}_c' \tag{A1}$$

where for example $\boldsymbol{\Omega}_{qc} = E(\boldsymbol{\epsilon}_{qt}\boldsymbol{\epsilon}_{ct}')$. Ignoring for the moment the final cross-product term in (A1), consider the terms in (A1) that depend linearly on $(\boldsymbol{\phi}_q', \boldsymbol{\phi}_c')'$, namely $-\boldsymbol{\phi}_q\boldsymbol{\Omega}_{pc} - \boldsymbol{\Omega}_{qp}\boldsymbol{\phi}_c'$. Observe that if we were to change the value of $\boldsymbol{\phi}_q$ from $\boldsymbol{\phi}_q^0$ to $\boldsymbol{\phi}_q^0 - \lambda\boldsymbol{\Omega}_{qp}$ and change $\boldsymbol{\phi}_c$ from $\boldsymbol{\phi}_c^0$ to $\boldsymbol{\phi}_c^0 + \lambda\boldsymbol{\Omega}_{cp}$, the change in the linear term would be $\lambda\boldsymbol{\Omega}_{qp}\boldsymbol{\Omega}_{pc} - \lambda\boldsymbol{\Omega}_{qp}\boldsymbol{\Omega}_{pc} = \mathbf{0}_{nm}$ for any scalar λ. This means for example that if we were to search within a neighborhood around $\boldsymbol{\phi}_q^0 = \mathbf{0}$ and $\boldsymbol{\phi}_c^0 = \mathbf{0}$, for λ small there are at most $n+m-1$ search directions that could be locally useful to try to get elements of the $(n\times m)$ matrix in (A1) closer to zero.

More generally, consider searching within a neighborhood of any arbitrary starting values $\boldsymbol{\phi}_q^0$ and $\boldsymbol{\phi}_c^0$. If we were to change $\boldsymbol{\phi}_q$ to $\boldsymbol{\phi}_q^0 - \lambda\boldsymbol{\Omega}_{qp} + \lambda\boldsymbol{\phi}_q^0\Omega_{pp}$ and change $\boldsymbol{\phi}_c$ to $\boldsymbol{\phi}_c^0 + \lambda\boldsymbol{\Omega}_{cp} - \lambda\boldsymbol{\phi}_c^0\Omega_{pp}$, the change in (A1) would be

$$\begin{aligned}
&g(\boldsymbol{\phi}_q^0 - \lambda\boldsymbol{\Omega}_{qp} + \lambda\boldsymbol{\phi}_q^0\Omega_{pp}, \boldsymbol{\phi}_c^0 + \lambda\boldsymbol{\Omega}_{cp} - \lambda\boldsymbol{\phi}_c^0\Omega_{pp}) - g(\boldsymbol{\phi}_q^0, \boldsymbol{\phi}_c^0) \\
&= -(-\lambda\boldsymbol{\Omega}_{qp} + \lambda\boldsymbol{\phi}_q^0\Omega_{pp})\boldsymbol{\Omega}_{pc} - \boldsymbol{\Omega}_{qp}(\lambda\boldsymbol{\Omega}_{cp} - \lambda\boldsymbol{\phi}_c^0\Omega_{pp})' \\
&+ (-\lambda\boldsymbol{\Omega}_{qp} + \lambda\boldsymbol{\phi}_q^0\Omega_{pp})\Omega_{pp}\boldsymbol{\phi}_c^{0\prime} + \boldsymbol{\phi}_q^0\Omega_{pp}(\lambda\boldsymbol{\Omega}_{cp} - \lambda\boldsymbol{\phi}_c^0\Omega_{pp})' \\
&+ (-\lambda\boldsymbol{\Omega}_{qp} + \lambda\boldsymbol{\phi}_q^0\Omega_{pp})\Omega_{pp}(\lambda\boldsymbol{\Omega}_{cp} - \lambda\boldsymbol{\phi}_c^0\Omega_{pp})' \\
&= \lambda^2(-\boldsymbol{\Omega}_{qp} + \boldsymbol{\phi}_q^0\Omega_{pp})\Omega_{pp}(\boldsymbol{\Omega}_{cp} - \boldsymbol{\phi}_c^0\Omega_{pp})'.
\end{aligned}$$

For λ small this again is arbitrarily close to zero. Thus from any point $(\boldsymbol{\phi}_q^{0\prime}, \boldsymbol{\phi}_c^{0\prime})'$ there are at most $(n+m-1)$ local search directions that could be used to try to get elements of (A1) closer to zero.

A.2 Imposing orthogonality of factor loadings

Typical applications of a factor structure such as principal component analysis use a normalization in which the factor loadings are orthogonal to each other. Here we describe how to implement the conditions $\mathbf{h}_q'\boldsymbol{\gamma}_q = \mathbf{h}_c'\boldsymbol{\gamma}_q = 0$ in our algorithm for maximum likelihood estimation.

Let $\mathbf{G}_{q\perp}$ denote the matrix consisting of the first $n-1$ columns of $(\mathbf{h}_q'\mathbf{h}_q)\mathbf{I}_n - \mathbf{h}_q\mathbf{h}_q'$:

$$\underset{n\times(n-1)}{\mathbf{G}_{q\perp}} = \left[(\mathbf{h}_q'\mathbf{h}_q)\mathbf{I}_n - \mathbf{h}_q\mathbf{h}_q'\right]\begin{bmatrix}\mathbf{I}_{n-1}\\ \mathbf{0}_{1,n-1}\end{bmatrix}.$$

Note that $\mathbf{G}_{q\perp}$ is constructed such that each column is orthogonal to $\mathbf{h}_q$:

$$\mathbf{h}_q'\mathbf{G}_{q\perp} = (\mathbf{h}_q'\mathbf{h}_q)(\mathbf{h}_q' - \mathbf{h}_q')\begin{bmatrix}\mathbf{I}_{n-1}\\ \mathbf{0}_{1,n-1}\end{bmatrix} = \mathbf{0}_{1,n-1}.$$

We can then parameterize $\boldsymbol{\gamma}_q = \mathbf{G}_{q\perp}\boldsymbol{\omega}_q$ where $\boldsymbol{\omega}_q$ is an $(n-1)\times 1$ vector of parameters to be estimated. Similarly, we define

$$\underset{m\times(m-1)}{\mathbf{G}_{c\perp}} = \left[(\mathbf{h}_c'\mathbf{h}_c)\mathbf{I}_m - \mathbf{h}_c\mathbf{h}_c'\right]\begin{bmatrix}\mathbf{I}_{m-1}\\ \mathbf{0}_{1,m-1}\end{bmatrix} \tag{A2}$$

and parameterize

$$\boldsymbol{\gamma}_c = \mathbf{G}_{c\perp}\boldsymbol{\omega}_c. \tag{A3}$$

Thus the specification becomes

$$\begin{bmatrix}\mathbf{K}_{qq} & \mathbf{K}_{qc}\\ \mathbf{K}_{cq} & \mathbf{K}_{cc}\end{bmatrix} = \begin{bmatrix}\mathbf{h}_q\mathbf{h}_q' + \mathbf{G}_{q\perp}\boldsymbol{\omega}_q\boldsymbol{\omega}_q'\mathbf{G}_{q\perp}' + \boldsymbol{\Sigma}_q & \mathbf{h}_q\mathbf{h}_c'\\ \mathbf{h}_c\mathbf{h}_q' & \mathbf{h}_c\mathbf{h}_c' + \mathbf{G}_{c\perp}\boldsymbol{\omega}_c\boldsymbol{\omega}_c'\mathbf{G}_{c\perp}' + \boldsymbol{\Sigma}_c\end{bmatrix}.$$

The likelihood function is then maximized with respect to the n elements of $\mathbf{h}_q$, the $(n-1)$ elements of $\boldsymbol{\omega}_q$, the n diagonal elements of the diagonal matrix $\boldsymbol{\Sigma}_q$, the m elements of $\mathbf{h}_c$, the $(m-1)$ elements of $\boldsymbol{\omega}_c$, and the m diagonal elements of the diagonal matrix $\boldsymbol{\Sigma}_c$.

A.3 Confidence bands for impulse-response functions

Let the $[(n+m+1)\times k]$ matrix $\hat{\boldsymbol{\Pi}}$ be the OLS estimate of the reduced-form coefficient matrices and $\hat{\boldsymbol{\Omega}}$ the OLS estimate of the reduced-form residual variance matrix. We know that the distribution of $\hat{\boldsymbol{\Pi}}$ is approximately given by

$$\text{vec}(\hat{\boldsymbol{\Pi}}') \sim N\left(\text{vec}(\boldsymbol{\Pi}'), (\boldsymbol{\Omega}\otimes\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}\right) \tag{A4}$$

and that this distribution is asymptotically independent of that of $\hat{\boldsymbol{\Omega}}$. Since the estimates of the structural parameters $\boldsymbol{\theta}$ are a function solely of $\hat{\boldsymbol{\Omega}}$ we can calculate confidence bands as follows. We draw a value for $\boldsymbol{\theta}^{(d)}$ from the distribution

$\boldsymbol{\theta} \sim N(\hat{\boldsymbol{\theta}}_{MLE}, \hat{\mathbf{V}})$ where $\hat{\boldsymbol{\theta}}_{MLE}$ is the maximum likelihood estimate and $\hat{\mathbf{V}}$ is its estimated variance-covariance matrix calculated from second derivatives of the log likelihood function. With this draw for $\boldsymbol{\theta}^{(d)}$ we calculate the implied value for $\boldsymbol{\Omega}^{(d)}$:

$$\boldsymbol{\Omega}^{(d)} = [\mathbf{A}(\boldsymbol{\theta}^{(d)})]^{-1} \left[\mathbf{D}(\boldsymbol{\theta}^{(d)}\right] [\mathbf{A}(\boldsymbol{\theta}^{(d)})^{-1}]'$$

$$\mathbf{D}(\boldsymbol{\theta}^{(d)}) = \left[\begin{array}{ccc} \mathbf{h}_q\mathbf{h}_q' + \boldsymbol{\gamma}_q\boldsymbol{\gamma}_q' + \boldsymbol{\Sigma}_q + \boldsymbol{\Sigma}_\chi & \mathbf{h}_q\mathbf{h}_c' & -\alpha\left(\mathbf{h}_q\mathbf{h}_q' + \boldsymbol{\gamma}_q\boldsymbol{\gamma}_q' + \boldsymbol{\Sigma}_q\right)\mathbf{s}_q + \alpha\mathbf{h}_q\mathbf{h}_c'\mathbf{s}_c \\ \mathbf{h}_c\mathbf{h}_q' & \mathbf{h}_c\mathbf{h}_c' + \boldsymbol{\gamma}_c\boldsymbol{\gamma}_c' + \boldsymbol{\Sigma}_c + \boldsymbol{\Sigma}_\psi & -\alpha\mathbf{h}_c\mathbf{h}_q'\mathbf{s}_q + \alpha(\mathbf{h}_c\mathbf{h}_c' + \boldsymbol{\gamma}_c\boldsymbol{\gamma}_c' + \boldsymbol{\Sigma}_c)\mathbf{s}_c \\ -\alpha\mathbf{s}_q'(\mathbf{h}_q\mathbf{h}_q' + \boldsymbol{\gamma}_q\boldsymbol{\gamma}_q' + \boldsymbol{\Sigma}_q) + \alpha\mathbf{s}_c'\mathbf{h}_c\mathbf{h}_q' & -\alpha\mathbf{s}_q'\mathbf{h}_q\mathbf{h}_c' + \alpha\mathbf{s}_c'(\mathbf{h}_c\mathbf{h}_c' + \boldsymbol{\gamma}_c\boldsymbol{\gamma}_c' + \boldsymbol{\Sigma}_c) & \alpha^2\left\{\mathbf{s}_q'\left(\mathbf{h}_q\mathbf{h}_q' + \boldsymbol{\gamma}_q\boldsymbol{\gamma}_q' + \boldsymbol{\Sigma}_q\right)\mathbf{s}_q - 2\mathbf{s}_q'\mathbf{h}_q\mathbf{h}_c'\mathbf{s}_c + \mathbf{s}_c'\left(\mathbf{h}_c\mathbf{h}_c' + \boldsymbol{\gamma}_c\boldsymbol{\gamma}_c' + \boldsymbol{\Sigma}_c\right)\mathbf{s}_c + \sigma_v^2\right\} \end{array}\right]$$

where for example we have simplified notation by writing $\mathbf{h}_q^{(d)}$ as $\mathbf{h}_q$. We use this value for $\boldsymbol{\Omega}^{(d)}$ to generate a draw for $\boldsymbol{\Pi}^{(d)}$ from

$$\text{vec}(\boldsymbol{\Pi}^{(d)\prime}) \sim N\left(\text{vec}(\hat{\boldsymbol{\Pi}}'), (\boldsymbol{\Omega}^{(d)} \otimes \left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}\right).$$

With this pair of $\boldsymbol{\theta}^{(d)}$ and $\boldsymbol{\Omega}^{(d)}$ we calculate the value of some structural magnitude of interest such as $\partial E(\mathbf{y}_{t+s}|\mathbf{y}_t, \mathbf{y}_{t-1}, ...\mathbf{y}_{t-r+1})/\partial u_{kt}$. We repeat this for draws $d = 1, ..., D$, and calculate the 68% bands for each object of interest.

A.4 Results for 3-factor model

Table A.1 reports maximum likelihood estimates when 3 factors (an overall global factor, a global supply factor, and a global demand factor) are used. The global supply factor is dominated by U.S. production. When this factor is included, there is no role for a separate idiosyncratic U.S. supply shock. All elasticity estimates and impulse-response functions are similar to those for the 2-factor model reported in Table 1.

Table A.1: Maximum likelihood estimates for 3-factor model

Parameter	MLE	Std err
$\phi_{q,US}$	0.028	(0.018)
$\phi_{q,Saudi}$	0.261	(0.051)
$\phi_{q,Russia}$	0.030	(0.011)
$\phi_{q,ROW}$	0.048	(0.028)
$\phi_{c,US}$	-0.104	(0.031)
$\phi_{c,Japan}$	-0.023	(0.036)
$\phi_{c,Europe}$	-0.236	(0.046)
$\phi_{c,ROW}$	-0.181	(0.048)
$\sigma_{q,US}$	0	—
$\sigma_{q,Saudi}$	6.362	(0.251)
$\sigma_{q,Russia}$	1.244	(0.289)
$\sigma_{q,ROW}$	0.838	(0.740)
$\sigma_{c,US}$	1.925	(0.108)
$\sigma_{c,Japan}$	3.082	(0.118)
$\sigma_{c,Europe}$	3.493	(0.150)
$\sigma_{c,ROW}$	2.508	(0.173)
σ_{χ}	0.888	(0.395)
σ_{ψ}	0	—
$h_{q,US}$	-0.468	(0.325)
$h_{q,Saudi}$	2.164	(0.500)
$h_{q,Russia}$	0.101	(0.083)
$h_{q,ROW}$	1.640	(0.274)
$h_{c,US}$	-0.203	(0.125)
$h_{c,Japan}$	-0.273	(0.154)
$h_{c,Europe}$	-0.208	(0.141)
$h_{c,ROW}$	1.029	(0.239)
ϕ_v	-0.294	(0.053)
σ_v	2.462	(0.295)
$\gamma_{q,US}$	2.234	(1.555)
$\gamma_{q,Saudi}$	0.227	(0.175)
$\gamma_{q,Russia}$	0.014	(0.026)
$\gamma_{q,ROW}$	0.314	(0.288)
$\gamma_{c,US}$	1.460	(0.455)
$\gamma_{c,Japan}$	1.567	(0.495)
$\gamma_{c,Europe}$	2.090	(0.548)
$\gamma_{c,ROW}$	1.113	(0.388)
α	1.930	(0.122)
ϕ_q	0.069	(0.020)
ϕ_c	-0.155	(0.039)