A Full-Information Approach to Granular Instrumental Variables

June 30, 2023
Revised: July 6, 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, heterogenous 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 heterogenous 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.

2 Data

The U.S. Energy Information Administration publishes monthly data on the production of oil in a number of different countries going back to 1973 and consumption 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

\[ q_{it} = \frac{Q_{it} - Q_{i,t-12}}{0.5(Q_{it} + Q_{i,t-12})} \]

\[ c_{jt} = \frac{C_{jt} - C_{j,t-12}}{0.5(C_{jt} + C_{j,t-12})}. \]

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 (2) is less extreme than \( \log( Q_{it} ) - \log( Q_{i,t-12} ) \), and is bounded between \( \pm 2 \) for all values of \( Q_{it} \) and \( Q_{i,t-12} \).

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

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 \(s_q\). In our data set, \(s_q = (0.12, 0.12, 0.15, 0.61)'\). We will approximate the annual year-over-year growth in global production using the share-weighted average of individual country growth rates:

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

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 production shares are summarized by the \((m \times 1)\) vector \(s_c = (0.25, 0.07, 0.08, 0.60)'\). We approximate the year-over-year growth rate of global consumption as

\[
\frac{C_t - C_{t-12}}{0.5(C_t + C_{t-12})} \approx \sum_{j=1}^{m} s_{cj} c_{jt}. \tag{4}
\]

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 $\mathbf{y}_t = (q'_t, c'_t, p_t)'$ 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 $\sum_{t \in \{1975:M1-2020:M2 \cup 2022:M2-2023:M2\}}$ 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

$$q_{it} = \phi_{qi} p_t + b'_{qit} x_{t-1} + u_{qit} + u_{xit} \quad (5)$$
for \( i = 1, \ldots, n \). Here \( \mathbf{x}_{t-1} = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-r})' \) 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 \( \mathbf{b}'_{qit} \mathbf{x}_{t-1} \) governs the dynamic behavior of oil supply in country \( i \). We assume that structural dynamics are incorporated in the definition of \( \mathbf{b}_{qit} \) 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 \( \phi_{qit} \) represents the response of supply to an unanticipated change in price. Thus \( \phi_{qit} \) should be interpreted as the within-month elasticity of supply for country \( i \). The longer-run elasticity of supply is a function of \( \mathbf{b}_{qit} \).

Likewise the structural demand equation for country \( j \) takes the form

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

for \( j = 1, \ldots, m \). Here \( \phi_{cjt} \) 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

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

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 \( \phi_v \) 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\footnote{Equation (9) is obtained by stacking the $n$ equations in (5) defining $\mathbf{q}_t = (q_{1t}, ..., q_{nt})'$, $\phi_q = (\phi_{q1}, ..., \phi_{qn})'$, $\mathbf{B}_q = [\mathbf{b}_{q1}, \mathbf{b}_{q2}, ..., \mathbf{b}_{qn}]'$, $\mathbf{u}_{qt} = (u_{q1t}, ..., u_{qnt})'$, and $\mathbf{u}_{xt} = (u_{x1t}, ..., u_{xnt})'$. Equation (10) is likewise obtained by stacking the $m$ equations in (6).}

$$\mathbf{q}_t = \phi_q \mathbf{p}_t + \mathbf{B}_q \mathbf{x}_{t-1} + \mathbf{u}_{qt} + \mathbf{u}_{xt} \tag{9}$$

$$\mathbf{c}_t = \phi_c \mathbf{p}_t + \mathbf{B}_c \mathbf{x}_{t-1} + \mathbf{u}_{ct} + \mathbf{u}_{wt} \tag{10}$$

The equilibrium condition (7) can then be written

$$\mathbf{s}_t' \mathbf{q}_t - \mathbf{u}_\chi t = \mathbf{s}_t' \mathbf{c}_t - \mathbf{u}_\psi t + \phi_q^t \mathbf{p}_t + \mathbf{b}_q' \mathbf{x}_{t-1} + \mathbf{u}_{et}$$

$$\mathbf{s}_t' \phi_q \mathbf{p}_t + \mathbf{s}_t' \mathbf{B}_q \mathbf{x}_{t-1} + \mathbf{s}_t' \mathbf{u}_{qt} = \mathbf{s}_t' \phi_c \mathbf{p}_t + \mathbf{s}_t' \mathbf{B}_c \mathbf{x}_{t-1} + \mathbf{s}_t' \mathbf{u}_{ct} + \phi_q^t \mathbf{p}_t + \mathbf{b}_q' \mathbf{x}_{t-1} + \mathbf{u}_{et}$$

$$\mathbf{s}_t' \phi_q - \mathbf{s}_t' \phi_c - \mathbf{c}_t \mathbf{p}_t = (\mathbf{s}_t' \mathbf{B}_c - \mathbf{s}_t' \mathbf{B}_q + \mathbf{b}_q' \mathbf{x}_{t-1} + \mathbf{s}_t' \mathbf{u}_{ct} - \mathbf{s}_t' \mathbf{u}_{qt} + \mathbf{u}_{et}.$$ \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 $\eta_{ctj}$. Suppose further that there is a common demand elasticity $\phi_c$ across all consumers. Under these assumptions, (6) becomes

$$c_{jt} = \phi_c p_t + b_{cj}' x_{t-1} + f_{ct} + \eta_{ctj} + \psi_{jt} \tag{12}$$

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

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

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

$$c_{jt} - \bar{c}_t = (b_{cj}' - \bar{b}_c)' x_{t-1} + (\eta_{ctj} - \bar{\eta}_{ct}) + (\psi_{jt} - \bar{\psi}_{jt}). \tag{14}$$
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 $j$,

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

where $c_t = \sum_{j=1}^{m}s_{cj}c_{jt}$, $b_j = \sum_{j=1}^{m}s_{cj}b_{cj}$, $\eta_{ct} = \sum_{j=1}^{m}s_{cj}\eta_{cjt}$, 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 $j = 1, \ldots, m$ as instruments for all of the producing countries $i = 1, \ldots, n$. 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

$$\alpha = (s'_q\phi_q - s'_c\phi_c - \phi_w)^{-1}. \quad (15)$$
Equation (11) can then be written

\[ p_t = \alpha (s'_e B_c - s'_q B_q + b'_v) x_{t-1} + \alpha (s'_e u_{ct} - s'_q u_{qt} + u_{vt}) \]

\[ = \pi'_p x_{t-1} + \varepsilon_{pt} \tag{16} \]

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

\[ q_t = (\phi_q \pi'_p + B_q) x_{t-1} + \phi_q \varepsilon_{pt} + u_{qt} + u_{xt} \]

\[ = \Pi_q x_{t-1} + \epsilon_{qt} \]

\[ c_t = (\phi_c \pi'_p + B_c) x_{t-1} + \phi_c \varepsilon_{pt} + u_{ct} + u_{\psi t} \]

\[ = \Pi_c x_{t-1} + \epsilon_{ct} \]

for \( \epsilon_{qt} = \phi_q \varepsilon_{pt} + u_{qt} + u_{xt} \) and \( \epsilon_{ct} = \phi_c \varepsilon_{pt} + u_{ct} + u_{\psi t} \). Note \( \epsilon_{qt} \) and \( \epsilon_{ct} \) are the errors one would make forecasting production or consumption from lagged observables and \( \Pi_q \) and \( \Pi_c \) can again be estimated by OLS.

The reduced-form VAR is thus

\[ y_t = \Pi x_{t-1} + \epsilon_t \]

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

\[ A \epsilon_t = u_t \tag{17} \]

\[ A = \begin{bmatrix} I_n & 0_{nm} & -\phi_q \\ 0_{mn} & I_m & -\phi_c \\ 0_{1n} & 0_{1m} & 1 \end{bmatrix} \tag{18} \]

\[ u_t = \begin{bmatrix} u_{qt} + u_{xt} \\ u_{ct} + u_{\psi t} \\ \alpha (s'_e u_{ct} - s'_q u_{qt} + u_{vt}) \end{bmatrix} \tag{19} \]
Recalling (17), the (2,1) block of (21) implies that shocks, $K$, are uncorrelated with each other. If demand shocks are uncorrelated with supply $D$, a common assumption in structural VARs is that the underlying structural shocks $\Phi = \Sigma \phi$.

Here $D = u$ maintains throughout that the inventory shock $u_A$ is whether there is at most one value for $\{\Pi, \Omega\}$ within the allowable class. Since we treat $u_A$ and $u_{it}$ are uncorrelated with all the other structural shocks, so that $D$ takes the form

$$D = \begin{bmatrix} K_{qq} + \Sigma_{\chi} & K_{qc} & -\alpha K_{qq} s_q + \alpha K_{qc} s_c \\ K_{cq} & K_{cc} + \Sigma_{\psi} & -\alpha K_{cq} s_q + \alpha K_{cc} s_c \\ -\alpha s_q' K_{qq} + \alpha s_c' K_{cq} & -\alpha s_q' K_{qc} + \alpha s_c' K_{cc} & \alpha^2 (s_q' K_{qq} s_q - 2 s_c' K_{cq} s_q + s_c' K_{cc} s_c + \sigma_v^2) \end{bmatrix}.$$  

(20)

Here $K_{qq} = E(u_{it} u_{it}')$, $K_{cc} = E(u_{it} u_{it}')$, $K_{qc} = E(u_{it} u_{it}') = K'_{cq}$, $\Sigma_{\chi} = E(u_{it} u_{it}')$, $\Sigma_{\psi} = E(u_{it} u_{it}')$, and $\sigma_v^2 = E(u_{it}^2)$.  

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, $K_{cq} = 0$ and $D$ simplifies to

$$D = \begin{bmatrix} K_{qq} + \Sigma_{\chi} & 0_{nm} & -\alpha K_{qq} s_q \\ 0_{nm} & K_{cc} + \Sigma_{\psi} & \alpha K_{cc} s_c \\ -\alpha s_q' K_{qq} & \alpha s_c' K_{cc} & \alpha^2 (s_q' K_{qq} s_q + s_c' K_{cc} s_c + \sigma_v^2) \end{bmatrix} = \begin{bmatrix} D_{qq} & 0_{nm} & D_{qp} \\ 0_{mn} & D_{cc} & D_{cp} \\ D_{pq} & D_{pc} & D_{pp} \end{bmatrix}.$$  

(21)

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

$$E(\epsilon_{qt} - \phi_q \bar{\epsilon}_{pt})(\epsilon_{ct} - \phi_c \bar{\epsilon}_{pt})' = 0_{nm}.$$  

When $n = m = 1$, this is a single equation to determine the two unknowns $\phi_q$ and $\phi_c$, and there would be an infinite number of combinations of $\phi_q$ and $\phi_c$ 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 \((\phi_q, \phi_c)\) 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 \( \phi = (\phi'_q, \phi'_c)' \) using the \( nm \) equations

\[
T^{-1} \sum_{t=1}^{T} (\hat{e}_{qt} - \phi_q \hat{e}_{pt})(\hat{e}_{ct} - \phi_c \hat{e}_{pt})' = \begin{bmatrix} I_n & 0_{nm} & -\phi_q \\ I_m & -\phi'_c \end{bmatrix} \hat{\Omega} \begin{bmatrix} 0_{nm} \\ -\phi'_q \\ -\phi'_c \end{bmatrix} = 0_{nm}. \tag{22}
\]

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 \( \phi_q \) and \( \phi_c \). More generally, with \( \phi_q \) and \( \phi_c \) 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 \( (\phi'_q, \phi'_c)' \), 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 \( \hat{\Omega} \) 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 \( \phi_q \) and \( \phi_c \), we would know the value of \( A \) and would be able to construct \( \hat{D} = A \hat{\Omega} A' \). Suppose we assume that measurement errors have the same variance and are uncorrelated across countries: \( \Sigma_\chi = \sigma^2_\chi I_n \) and \( \Sigma_\psi = \sigma^2_\psi I_m \). From (21), we can then estimate the off-diagonal elements of \( K_{qq} \) and \( K_{cc} \) from the off-diagonal elements of \( \hat{D}_{qq} \) and \( \hat{D}_{cc} \), respectively. The diagonal elements of \( \hat{D}_{qq} \) can tell us the sum of \( \sigma^2_\chi I_n \) plus the diagonal elements of \( K_{qq} \). We can then use the observed covariances in \( \hat{D}_{qp} \) to obtain a separate estimate of diagonal elements of \( K_{qq} \) up to a constant of proportionality \( \alpha \), where the single remaining unknown element in \( \alpha \) is the inventory
elasticity $\phi_v$. Combining these two sources of information, we can thus use the $n$ diagonal elements in $\hat{D}_{qq}$ along with the $n$ elements of $\hat{D}_{qp}$ to estimate the $n$ diagonal elements of $K_{qq}$ along with the two scalars $\sigma^2_\chi$ and $\phi_v$. For example, suppose that the value of $\phi'_q, \phi'_c$ that makes $\hat{D}_{qc}$ 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 $D_{pq}$ in (21) than we observe in $\hat{D}_{pq}$. This would be interpreted either as evidence of significant errors in measuring production ($\sigma^2_\chi > 0$) or as a negative value for $\phi_v$. A negative value for $\phi_v$ 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 $\hat{\epsilon}_{qt} - \phi_q \hat{\epsilon}_{pt}$ and $\hat{\epsilon}_{pt}$. 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 $\hat{D}_{cc}$ along with the $m$ covariances $\hat{D}_{cp}$ to estimate the $m$ diagonal elements of $K_{cc}$ along with $\sigma^2_\psi$ and to provide additional information about $\phi_v$. We can then uncover the final unknown magnitude $\sigma^2_v$ 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

$$
\mathbf{u}_{qt} = \mathbf{h}_q f_t + \gamma_q f_{qt} + \eta_{qt} \quad (23)
$$
Here $\gamma_c$ 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 $\gamma_{cj}$. We normalize these weights by imposing that the loadings on the supply or demand factors are orthogonal to the loadings on the global factor: $h_q' \gamma = h_c' \gamma_c = 0$. The appendix describes how we implemented this mechanically. The $(n \times 1)$ vector $\eta_{qt}$ 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 $E(\eta_{qt} \eta_{qt}') = \Sigma_q$, where $\Sigma_q$ is a diagonal $(n \times n)$ matrix whose $i$th diagonal element measures the variance of purely idiosyncratic shocks to supply. Similarly $E(\eta_{ct} \eta_{ct}') = \Sigma_c$, a diagonal $(m \times m)$ matrix.

For this model, the matrix in (20) becomes

$$D = \begin{bmatrix}
    h_qh_q' + \gamma_q\gamma_q' + \Sigma_q + \sigma^2_{\epsilon}I_n & h_qh_c' \\
    h_qh_c' & h_c'h_c' + \gamma_c\gamma_c' + \Sigma_c + \sigma^2_{\epsilon}I_m \\
    -\alpha s_q'(h_qh_q' + \gamma_q\gamma_q' + \Sigma_q) + \alpha s_c'h_c'h_q' - \alpha s_q'h_qh_c' + \alpha s_c'(h_c'h_c' + \gamma_c\gamma_c' + \Sigma_c) \\
    -\alpha (h_qh_q' + \gamma_q\gamma_q' + \Sigma_q)s_q + \alpha h_qh_c's_c \\
    -\alpha h_qh_q's_q + \alpha (h_c'h_c' + \gamma_c\gamma_c' + \Sigma_c)s_c \\
    \alpha^2 [s_q'(h_qh_q' + \gamma_q\gamma_q' + \Sigma_q)s_q - 2s_q'h_qh_c's_q + s_c'(h_c'h_c' + \gamma_c\gamma_c' + \Sigma_c)s_c + \sigma^2_{\epsilon}] \\
\end{bmatrix}.$$  

In the model in the previous subsection, the elasticities $\phi_q$ and $\phi_c$ were identified from (22), which required that all linear combinations of $\epsilon_{qt} - \phi_q\epsilon_{pt}$ be uncorrelated with all linear combinations of $\epsilon_{ct} - \phi_c\epsilon_{pt}$, where $\epsilon_{qt}$, $\epsilon_{ct}$, and $\epsilon_{pt}$ are the observable errors one would make in forecasting $q_t$, $c_t$, and $p_t$ one month in advance. The model here generalizes this assumption, looking for values of $\phi_q$ and $\phi_c$ 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 $D_{qq}$ and $D_{cc}$ can be summarized by single supply and demand factors beyond the
common factor in the $D_{qc}$ covariance. The values of $\Sigma_q$, $\sigma^2$, and $\phi_v$ are then inferred from the diagonal elements of $D_{qq}$ and the covariances between $\epsilon_{qt} - \phi_q \epsilon_{pt}$ and $\epsilon_{pt}$. 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 $\Sigma_c$ and $\sigma^2$ (and additional information about all the other parameters) are informed by the diagonal elements of $D_{cc}$ and the covariances between $\epsilon_{cqt} - \phi_c \epsilon_{pt}$ and $\epsilon_{pt}$.

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

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

(26)

Here $\hat{\Omega}$ is the OLS estimate $T^{-1} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t'$, $\hat{\epsilon}_t = y_t - \hat{\Pi} x_{t-1}$, and

$$\hat{\Pi} = \left( \sum_{t=1}^T y_t' x_{t-1} \right) \left( \sum_{t=1}^T x_{t-1}' x_{t-1} \right)^{-1}.$$ 

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

$$\Omega(\theta) = A(\theta)^{-1} D(\theta) [A(\theta)^{-1}]',$$

$$A(\theta)^{-1} = \begin{bmatrix} I_n & 0_{nm} & \phi_q \\ 0_{mn} & I_m & \phi_c \\ 0_{1n} & 0_{1m} & 1 \end{bmatrix},$$

where $D(\theta)$ is the matrix in (25). The vector $\theta$ contains $4(m + n) + 2$ unknown structural parameters. The $4n$ supply parameters consist of the $n$ supply elasticities $\phi_q$, the $n$ idiosyncratic supply variances $\Sigma_n$, the $n$ loadings of supply $h_q$ on the global factor $f_t$, the $(n - 1)$ loadings of supply $\omega_q$ on the supply factor $f_{qt}$ (which imposes the normalization condition $h_q' \gamma_q = 0$ as described in the appendix), and the
variance $\sigma^2$ of the error in measuring production for each country. The $4m$ demand parameters consist analogously of $\phi_c, \Sigma_c, h_c, \omega_c, \sigma^2_p$. In addition there are the two inventory demand parameters, which consist of the elasticity $\phi_v$ and variance of the inventory shock $\sigma^2_v$. Our proposal is to estimate $\theta$ 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 $\Omega(\theta) = \hat{\Omega}$ achieving a value for the log likelihood of

$$\ell(\hat{\theta}_{unrestricted}) = -[T(n + m + 1)/2][1 + \log(2\pi)] - (T/2) \log \hat{\Omega}.$$  

The unconstrained model estimates $(n + m + 1)(n + m + 2)/2$ parameters in the covariance matrix $\Omega$. For our baseline model with $n = m = 4$, there are 34 elements in the structural parameter vector $\theta$ 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

$$\epsilon_t = A^{-1}u_t.$$ 

$$A^{-1}_{(n+m+1)\times(n+m+1)} = \begin{bmatrix} I_n & 0_{nm} & \phi_q \\ 0_{mn} & I_m & \phi_c \\ 0_{1n} & 0_{1m} & 1 \end{bmatrix}$$ 

$$u_t = \begin{bmatrix} h_q f_t + \gamma_q f_{qt} + \eta_{qt} + u_{xt} \\ h_c f_t + \gamma_c f_{ct} + \eta_{ct} + u_{ct} \\ \alpha s'_c(h_c f_t + \gamma_c f_{ct} + \eta_{ct}) - \alpha s'_q(h_q f_t + \gamma_q f_{qt} + \eta_{qt}) + \alpha u_{vt} \end{bmatrix}.$$ 

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 $y_t$ at time $t$ using $\partial \epsilon_t/\partial u_{kt} = A^{-1}\partial u_t/\partial u_{kt}$. For example, the effect of a one-percent increase in the supply from country $i$ alone resulting from a one-unit increase in $\eta_{qit}$ would be

$$\frac{\partial \epsilon_t}{\partial \eta_{qit}} = A^{-1} \begin{bmatrix} e_i^{(n)} \\ 0 \\ -\alpha s'_q e_i^{(n)} \end{bmatrix}.$$
where \( e_i^{(n)} \) is the \( i \)th column of \( I_n \). Likewise, the effect of a one-percent increase in demand from region \( j \) alone is

\[
\frac{\partial \epsilon_t}{\partial \eta_{cjt}} = A^{-1} \begin{bmatrix} 0 \\ e_j^{(m)} \\ \alpha s' e_j^{(m)} \end{bmatrix}
\]

for \( e_j^{(m)} \) the \( j \)th column of \( I_m \). The effect of a one-standard-deviation increase in the global demand factor is

\[
\frac{\partial \epsilon_t}{\partial f_{ct}} = A^{-1} \begin{bmatrix} 0 \\ \gamma_c \\ \alpha s' \gamma_c \end{bmatrix}.
\]

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

\[
\frac{\partial \epsilon_t}{\partial u_{vt}} = A^{-1} \begin{bmatrix} 0 \\ 0 \\ \alpha \end{bmatrix}.
\]

Let \( \Psi_s \) denote the usual reduced-form dynamic multiplier \( \Psi_s = \partial y_{t+s}/\partial \epsilon_t \). 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 \( y_{t+s} \),

\[
\frac{\partial E(y_{t+s}|y_t, y_{t-1}, \ldots y_{t-r+1})}{\partial u_{kt}} = \Psi_s \frac{\partial \epsilon_t}{\partial u_{kt}}.
\]

To get this, we just plug in one of the above expressions for \( \frac{\partial \epsilon_t}{\partial u_{kt}} \). For example,

\[
\frac{\partial E(y_{t+s}|y_t, y_{t-1}, \ldots y_{t-r+1})}{\partial f_{ct}} = \Psi_s A^{-1} \begin{bmatrix} 0 \\ \gamma_c \\ \alpha s' \gamma_c \end{bmatrix}.
\]

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

\[ \phi_q = \sum_{i=1}^{n} s_{qi} \phi_{qi} = 0.064. \]  

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 the demand of

\[ \phi_c = \sum_{j=1}^{m} s_{cj} \phi_{cj} = -0.139. \]  

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
Table 1: Maximum likelihood estimates

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

Notes to Table 1. The four elements of \( \gamma_{c} \) were not estimated directly but were calculated from the three elements of \( \omega_{c} \) (not reported in the table) along with the four elements of \( h_{c} \) reported in the table using equations (A3) and (A2). Standard errors for \( \gamma_{c} \) were calculated by simulating draws from the asymptotic distribution of \( \hat{\theta} \) as a byproduct of the algorithm used to calculate confidence bands for impulse-response functions. The values of \( \alpha, \phi_{q} \) and \( \phi_{c} \) were not estimated directly but were inferred from \( \hat{\theta} \) using equations (15), (30), and (31) with standard errors for \( \hat{\alpha}, \hat{\phi}_{q}, \) and \( \hat{\phi}_{c} \) obtained by simulation.
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 the estimates reported in Table 1 impose \( \gamma_q = 0 \). Estimates of other parameters when \( \gamma_q \) is unconstrained are similar to those reported in Table 1. The coefficients \( \gamma_c \) 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 \( \sigma^2_{\psi} \) 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 \( \phi_v \). 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.\(^2\) In addition to the 45 parameters used to fit the variance matrix \( \Omega \), there

\(^2\)If we were to count \( \hat{\sigma}_{\psi}^2 = 0 \) as another estimated parameter it would bring the degrees of
are another \( k = 1 + r(n + m + 1) = 109 \) parameters estimated for each row of \( \Pi \) in the unconstrained VAR. Adjusting to correct for small-sample bias\(^3\) would result in a test statistic of

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

for which the \( \chi^2(15) \) \( p \)-value is 0.17. We conclude that the overidentifying assumptions 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 \( \gamma_{cj} \), 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 \( \alpha s^t \gamma_c \), 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 \( \phi_{qi} \) or \( \phi_{cj} \). 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.

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 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.

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.
Table 2: Impact effects of a global demand shock

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

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).

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 $\tilde{\alpha}s_c\gamma_c$ where $\tilde{\alpha} = 1/(\phi_q s_q - \phi_c s_c) = 4.92$. 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 $a_{q,Russia}$. This would represent a loss of over 5 million barrels per day. For this example, we use production and consumption shares as of
Table 3: Impact effects of a shock to Russia supply

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \text{direct effect})</th>
<th>( \text{response to price})</th>
<th>( \text{net effect})</th>
<th>( \text{as } % \text{ of country} )</th>
<th>( \text{in } \text{mb/d} )</th>
<th>( \text{with } \phi_v = 0 )</th>
<th>( \text{as } % \text{ of country} )</th>
<th>( \text{in } \text{mb/d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>12.430</td>
<td>31.185</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( q_{US} )</td>
<td>0</td>
<td>0.231</td>
<td>0.231</td>
<td>0.029</td>
<td>0</td>
<td>0.580</td>
<td>0.580</td>
<td>0.072</td>
</tr>
<tr>
<td>( q_{Saudi} )</td>
<td>0</td>
<td>3.222</td>
<td>3.222</td>
<td>0.318</td>
<td>0</td>
<td>8.084</td>
<td>8.084</td>
<td>0.798</td>
</tr>
<tr>
<td>( q_{Russia} )</td>
<td>-50</td>
<td>0.362</td>
<td>-49.638</td>
<td>-5.311</td>
<td>-50</td>
<td>0.909</td>
<td>8.084</td>
<td>-5.252</td>
</tr>
<tr>
<td>( q_{ROW} )</td>
<td>0</td>
<td>0.536</td>
<td>0.536</td>
<td>0.265</td>
<td>0</td>
<td>1.345</td>
<td>1.345</td>
<td>0.664</td>
</tr>
<tr>
<td>( q )</td>
<td></td>
<td></td>
<td>-4.699</td>
<td>-4.699</td>
<td></td>
<td></td>
<td></td>
<td>-3.718</td>
</tr>
<tr>
<td>( c_{US} )</td>
<td>0.000</td>
<td>-1.163</td>
<td>-1.163</td>
<td>-0.191</td>
<td>0.000</td>
<td>-2.919</td>
<td>-2.919</td>
<td>-0.480</td>
</tr>
<tr>
<td>( c_{Japan} )</td>
<td>0.000</td>
<td>-0.227</td>
<td>-0.227</td>
<td>-0.007</td>
<td>0.000</td>
<td>-0.569</td>
<td>-0.569</td>
<td>-0.019</td>
</tr>
<tr>
<td>( c_{Europe} )</td>
<td>0.000</td>
<td>-2.795</td>
<td>-2.795</td>
<td>-0.115</td>
<td>0.000</td>
<td>-7.011</td>
<td>-7.011</td>
<td>-0.289</td>
</tr>
<tr>
<td>( c_{ROW} )</td>
<td>0.000</td>
<td>-1.999</td>
<td>-1.999</td>
<td>-1.168</td>
<td>0.000</td>
<td>-5.015</td>
<td>-5.015</td>
<td>-2.930</td>
</tr>
<tr>
<td>( c )</td>
<td></td>
<td></td>
<td>-1.482</td>
<td>-1.482</td>
<td></td>
<td></td>
<td></td>
<td>-3.718</td>
</tr>
<tr>
<td>( v )</td>
<td></td>
<td></td>
<td>3.217</td>
<td>3.217</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>

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).

The end of our sample (February 2023). 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. 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 \( s_{qT}Q_T \) or \( s_{cT}Q_T \) where \( Q_T = 82.3 \text{ mb/d} \) 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 on 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

---

4These were \( s_{qT} = (0.15,0.12,0.13,0.60)' \) and \( s_{cT} = (0.20,0.04,0.05,0.71)' \).

5This was calculated as \(-0.5\bar{s}_{q,\text{Russia},T}/(s'_{qT}\Phi_q - s'_{cT}\Phi_c)\).
first month, which clearly is not sustainable.\footnote{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.} 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.

Figure 3: Dynamic effects of a Russian supply shock

\textit{Notes to Figure 3.} Dynamic effects of a 50\% decrease in Russian oil production 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.

We plot the predicted dynamic effects under the assumption of no inventory 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 Arabian production and the endogenous adjustment of inventories have historically played a key role in stabilizing the world price of oil.
References


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 \((\phi_q', \phi_c')'\) that is contained in the elements of the \((n \times m)\) matrix

\[
g(\phi_q, \phi_c) = E(\epsilon_q t - \phi_q z_p t)(\epsilon c t - \phi c z_p t)' = \Omega_{qc} - \phi_q \Omega_{pc} - \Omega_{qp} \phi_c' + \phi_q \Omega_{pp} \phi_c' \tag{A1}
\]

where for example \(\Omega_{qc} = E(\epsilon_q t \epsilon_c t').\) Ignoring for the moment the final cross-product term in (A1), consider the terms in (A1) that depend linearly on \((\phi_q', \phi_c')'\), namely

\[-\phi_q \Omega_{pc} - \Omega_{qp} \phi_c'.\]

Observe that if we were to change the value of \(\phi_q\) from \(\phi_q^0\) to \(\phi_q^0 - \lambda \Omega_{qp}\) and change \(\phi_c\) from \(\phi_c^0\) to \(\phi_c^0 + \lambda \Omega_{cp}\), the change in the linear term would be \(\lambda \Omega_{qp} \Omega_{pc} - \lambda \Omega_{qp} \Omega_{pc} = 0_{nm}\) for any scalar \(\lambda\). This means for example that if we were to search within a neighborhood around \(\phi_q^0 = 0\) and \(\phi_c^0 = 0\), for \(\lambda\) 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 \(\phi_q^0\) and \(\phi_c^0\). If we were to change \(\phi_q\) to \(\phi_q^0 - \lambda \Omega_{qp} + \lambda \phi_q^0 \Omega_{pp}\) and change \(\phi_c\) to \(\phi_c^0 + \lambda \Omega_{cp} - \lambda \phi_c^0 \Omega_{pp}\), the change in (A1) would be

\[
g(\phi_q^0 - \lambda \Omega_{qp} + \lambda \phi_q^0 \Omega_{pp}, \phi_c^0 + \lambda \Omega_{cp} - \lambda \phi_c^0 \Omega_{pp}) - g(\phi_q^0, \phi_c^0) \]
\[
= -(-\lambda \Omega_{qp} + \lambda \phi_q^0 \Omega_{pp}) \Omega_{pc} - \Omega_{qp}(\lambda \Omega_{cp} - \lambda \phi_c^0 \Omega_{pp}) - g(\phi_q^0, \phi_c^0) \]
\[
= -(-\lambda \Omega_{qp} + \lambda \phi_q^0 \Omega_{pp}) \Omega_{pc} - \Omega_{qp}(\lambda \Omega_{cp} - \lambda \phi_c^0 \Omega_{pp}) - g(\phi_q^0, \phi_c^0) \]
\[
= \lambda^2 (-\Omega_{qp} + \phi_q^0 \Omega_{pp}) \Omega_{pp} (\Omega_{cp} - \phi_c^0 \Omega_{pp})'.
\]

For \(\lambda\) small this again is arbitrarily close to zero. Thus from any point \((\phi_q^0', \phi_c^0')'\) 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 \(h_q' \gamma_q = h_c' \gamma_q = 0\) in our algorithm for maximum likelihood estimation.
Let $G_{q\perp}$ denote the matrix consisting of the first $n-1$ columns of $(h_q'h_q)I_n - h_q h_q'$:

$$
G_{q\perp} = \begin{bmatrix}
(h_q'h_q)I_n - h_q h_q' \\
I_{n-1} \\
0_{1,n-1}
\end{bmatrix}.
$$

Note that $G_{q\perp}$ is constructed such that each column is orthogonal to $h_q$:

$$
h_q'G_{q\perp} = (h_q'h_q)(h_q' - h_q') = 0_{1,n-1}.
$$

We can then parameterize $\gamma_q = G_{q\perp}\omega_q$ where $\omega_q$ is an $(n-1) \times 1$ vector of parameters to be estimated. Similarly, we define

$$
G_{c\perp} = \begin{bmatrix}
(h_c'h_c)I_m - h_c h_c' \\
i_{m-1} \\
0_{1,m-1}
\end{bmatrix}
$$

and parameterize

$$
\gamma_c = G_{c\perp}\omega_c.
$$

Thus the specification becomes

$$
\begin{bmatrix}
K_{qq} & K_{qc} \\
K_{eq} & K_{cc}
\end{bmatrix} = \begin{bmatrix}
h_q h_q' + G_{q\perp}\omega_q \omega_q' G_{q\perp}' + \Sigma_q & h_q h_c' \\
h_c h_q' & h_c h_c' + G_{c\perp} \omega_c \omega_c' G_{c\perp}' + \Sigma_c
\end{bmatrix}.
$$

The likelihood function is then maximized with respect to the $n$ elements of $h_q$, the $(n-1)$ elements of $\omega_q$, the $n$ diagonal elements of the diagonal matrix $\Sigma_q$, the $m$ elements of $h_c$, the $(m-1)$ elements of $\omega_c$, and the $m$ diagonal elements of the diagonal matrix $\Sigma_c$.

### A.3 Confidence bands for impulse-response functions

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

$$
\text{vec}(\hat{\Pi})' \sim N\left(\text{vec}(\Pi)', (\Omega \otimes \left(\sum_{t=1}^{T}x_{t-1}x_{t-1}'\right)^{-1})\right)
$$

and that this distribution is asymptotically independent of that of $\hat{\Omega}$. Since the estimates of the structural parameters $\theta$ are a function solely of $\hat{\Omega}$ we can calculate confidence bands as follows. We draw a value for $\theta^{(d)}$ from the distribution
\( \theta \sim N(\hat{\theta}_{MLE}, \hat{V}) \) where \( \hat{\theta}_{MLE} \) is the maximum likelihood estimate and \( \hat{V} \) is its estimated variance-covariance matrix calculated from second derivatives of the log likelihood function. With this draw for \( \theta^{(d)} \) we calculate the implied value for \( \Omega^{(d)} \):

\[
\Omega^{(d)} = [A(\theta^{(d)})]^{-1} D(\theta^{(d)}) [A(\theta^{(d)})^{-1}]'
\]

\[
D(\theta^{(d)}) = \\
\begin{bmatrix}
  h_q h'_q + \gamma_q\gamma'_q + \Sigma_q + \Sigma_\chi & h_q h'_c \\
  h_h'_q & h_h'_c + \gamma_c\gamma'_c + \Sigma_c + \Sigma_\psi \\
  -\alpha s'_q(h_q h'_q + \gamma_q\gamma'_q + \Sigma_q) + \alpha s'_c h_h'_q & -\alpha s'_q h_h'_c + \alpha s'_c(h_h'_c + \gamma_c\gamma'_c + \Sigma_c)
\end{bmatrix}
\]

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

\[
\text{vec}(\Pi^{(d)'}) \sim N\left(\text{vec}(\hat{\Pi}'), (\Omega^{(d)} \otimes (\sum_{t=1}^{T} x_{t-1} x_{t-1}')^{-1})\right).
\]

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