CESifo WORKING PAPERS

8153
2020

March 2020

The Econometrics of Oil Market VAR Models

Lutz Kilian, Xiaoqing Zhou

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8153*

The Econometrics of Oil Market VAR Models

Abstract

Oil market VAR models have become the standard tool for understanding the evolution of the real price of oil and its impact in the macro economy. As this literature has expanded at a rapid pace, it has become increasingly difficult for mainstream economists to understand the differences between alternative oil market models, let alone the basis for the sometimes divergent conclusions reached in the literature. The purpose of this survey is to provide a guide to this literature. Our focus is on the econometric foundations of the analysis of oil market models with special attention to the identifying assumptions and methods of inference. We not only explain how the workhorse models in this literature have evolved, but also examine alternative oil market VAR models. We help the reader understand why the latter models sometimes generated unconventional, puzzling or erroneous conclusions. Finally, we discuss the construction of extraneous measures of oil demand and oil supply shocks that have been used as external or internal instruments for VAR models.

JEL-Codes: Q430, Q410, C360, C520.

Keywords: identification, model specification, elasticity, Bayesian estimation, structural VAR, instruments, textual analysis.

*Lutz Kilian**
Federal Reserve Bank of Dallas
Research Department
2200 N. Pearl St.
USA – Dallas, TX 75201
lkilian2019@gmail.com

Xiaoqing Zhou
Federal Reserve Bank of Dallas
Research Department
2200 N. Pearl St.
USA – Dallas, TX 75201
xqzhou3@gmail.com

*corresponding author

February 26, 2020
The views expressed in this paper are our own and should not be interpreted as reflecting the views of the Federal Reserve Bank of Dallas or any other member of the Federal Reserve System. We thank Ana Maria Herrera for helpful discussions.

1. Introduction

In the last decade, structural vector autoregressive (VAR) models of the global oil market have become the standard tool for understanding the evolution of the real price of oil and its effect on the macro economy. These models have helped create a consensus among researchers and policymakers that shifts in the global demand for oil are the primary determinant of the real price of oil. As the oil market literature has grown at a rapid pace, there has been a proliferation of alternative approaches to modeling the real price of oil, some refining existing models of the global oil market and others challenging this consensus. As a result, it has become increasingly difficult for mainstream economists to understand the differences between alternative oil market models, let alone the basis for the sometimes divergent conclusions reached in the literature. The purpose of this survey is to provide a guide to this literature.[1]

Our focus is on the econometric foundations of the analysis of oil market models with special attention to the identifying assumptions and methods of inference. We not only explain how the workhorse models in this literature have evolved, but also examine alternative oil market VAR models. Our review is intended to help the reader understand why the latter models sometimes have generated unconventional, puzzling or erroneous conclusions. Finally, we discuss the construction of extraneous measures of oil demand and oil supply shocks that have been used as external or internal instruments for VAR models. Our analysis is of interest not only to applied researchers interested in modeling oil markets and their relationship with the domestic economy, but also to applied econometricians interested in structural VAR modeling more generally.

The remainder of the paper is organized as follows. In section 2, we trace the evolution of

[1] We do not address in this survey the related VAR literature on the effects of global oil price shocks on the domestic economy, which abstracts from the fact that not all oil price shocks are alike.

the oil market literature from its origins to the workhorse model of Kilian and Murphy (2014). We also discuss extensions to larger-dimensional models. Section 3 discusses estimation and inference about oil market models, including the usefulness of time-varying coefficient VAR models. In section 4, we review some common mistakes in applied studies. Section 5 examines some recent controversies about the specification and estimation of oil market VAR models. Non-traditional approaches to identifying oil demand and oil supply shocks are discussed in section 6. The concluding remarks are in section 7.

2. Conventional Identification Strategies in Oil Market Models

Oil market VAR models evolved in three main stages, starting with the work of Kilian (2006, 2008a, 2009) which exploited short-run exclusion restrictions. This approach was followed by the introduction of static and dynamic inequality restrictions as an alternative to short-run exclusion restrictions, as exemplified by Lippi and Nobili (2012), Kilian and Murphy (2012) and Inoue and Kilian (2013). Finally, Kilian and Murphy (2014) extended this framework to allow oil price expectations to have an effect on the real price of oil through shifts in storage demand. This last extension is crucial because it addresses concerns about previous oil market models being informationally deficient. It also provides a direct link to the literature on modeling oil futures markets (see Alquist and Kilian 2010, Knittel and Pindyck 2016).[2] In this section, we review these benchmark models and their extensions.

2.1. Short-Run Exclusion Restrictions

The importance of disentangling demand and supply shocks in oil markets was first pointed out by Barsky and Kilian (2002, 2004). Kilian (2008a, 2009) provided a quantitative framework for

[2] Variations of the Kilian and Murphy (2014) framework have been used in a range of recent studies including Kilian and Lee (2014), Juvenal and Petrella (2015), Kilian (2017), Baumeister and Hamilton (2019a), Herrera and Rangaraju (2019),Zhou (2019), Kilian and Zhou (2019a,b), and Cross (2019).

this task based on a structural VAR model. These studies focused on the global oil market since 1973. The analysis is based on a model including the percent change in global oil production $(\Delta prod)$, an index of cyclical variation in global real economic activity derived from bulk dry cargo ocean freight rates (rea), and the log of the real price of oil $(rpoil)$. Variation in these data is explained based on three shocks: (1) a shock to the amount of oil pumped out of the ground ("oil supply shock"), (2) a shock to the demand for all industrial commodities including crude oil ("aggregate demand shock"), and (3) a residual oil demand shock ("oil-specific demand shock") designed to capture precautionary oil demand shocks (see Alquist and Kilian 2010). The latter shock may also be interpreted as a preference shock. For example, an increased preference for smaller, more fuel-efficient automobiles would result in lower demand for oil, given the same level of global real activity. Thus, there are two oil demand shocks and one oil supply shock in this model.

The identifying restrictions are imposed on the structural matrix impact multiplier matrix, B_0^{-1} with *ij*th element b_{ij}^0, that links the vector of reduced-form errors, u_t, to the vector of structural shocks, w_t:

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_{11}^0 & 0 & 0 \\ b_{21}^0 & b_{22}^0 & 0 \\ b_{31}^0 & b_{32}^0 & b_{33}^0 \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix} \tag{1}$$

The real price of oil is allowed to respond instantaneously to all three structural shocks. Kilian (2009) postulates that the impact price elasticity of oil supply is zero, which means that global oil production does not respond to either demand shock on impact ($b_{12}^0 = b_{13}^0 = 0$). He also postulates that an oil-specific demand shock that raises the real price of oil does not lower global real activity within the same month ($b_{23}^{0} = 0$).[3] There are no restrictions on the responses at longer horizons.[4]

Although Kilian (2009) imposes a recursive structure on the structural impact multiplier matrix, this structure is not atheoretical. Every exclusion restriction is motivated on economic grounds. For example, the assumption of a zero impact price elasticity of oil supply is consistent with evidence about how OPEC oil producers historically have set production levels (see Kilian 2009). It is also consistent with the theoretical result that oil producers in response to higher prices adjust their production levels only with a delay, because adjusting the level of oil production is costly (see Anderson, Kellogg and Salant 2018).[5]

2.2. Long-Run Exclusion Restrictions

Long-run identifying restrictions are rarely used in oil market models. Jacks and Stürmer (2019) analysis of commodity markets between 1870 and 2013 is an exception. Although Jacks and Stürmer's framework was designed for other commodities, it may be applied to the oil market as well (see Stürmer 2018). Their model relies on the same three variables as Kilian (2009), except that all variables are annual with global real GDP growth taking the place of the global real activity index. Since their data are annual, the short-run identifying restrictions employed by

[3] Kilian and Zhou (2018) establish the validity of this assumption for the measure of global real activity utilized by Kilian (2009). Whether this assumption holds for alternative measures of global real activity is less clear.

[4] As Kilian and Zhou (2018) point out, there is a strong case to be made for imposing the additional restriction $b_{21}^{0} = 0,$ because, based on the same reasoning that suggests that $b_{23}^{0} = 0,$ oil supply shocks should not affect global real activity on impact. Kilian and Lütkepohl (2017) show that this overidentifying restriction cannot be rejected at conventional significance levels. When the Kilian (2009) model is estimated by GMM subject to the overidentifying restriction, the impulse response estimates are indistinguishable from the original estimates in Kilian (2009). The reason is that even without the overidentifying restriction, the estimate of b_{21}^{0} is close to zero.

[5] While the identifying restrictions in the recursive Kilian (2009) model are not testable, they may be treated as overidentifying restrictions within an oil market VAR model identified by heteroskedasticity. Lütkepohl and Netšunajev (2014) report not being able to reject these overidentifying restrictions. Similarly, Herwartz and Plödt (2016) establish that the impulse response estimates obtained from a non-Gaussian oil market VAR model under the stronger assumption of independent (rather than mutually uncorrelated) structural shocks are similar to those reported in Kilian (2009).

Kilian (2009) are not appealing. Jacks and Stürmer therefore develop a different approach to identification.

They observe that persistent global economic expansions tend to be associated with gains in total factor productivity and hence permanently raise the level of global real output. These expansions are also associated with sustained increases in the real prices of industrial commodities because they raise the flow demand for industrial commodities. Persistently high real commodity prices in turn stimulate technological innovation in resource extraction and discoveries of new deposits of industrial raw materials. Thus, such commodity demand shocks also tend to have a permanent effect on the level of the production of industrial commodities. In contrast, commodity-specific supply shocks are associated with strikes, industrial accidents, natural disasters, and political unrest or wars. Such supply shocks typically have only short-lived effects on global real output. Hence, it makes sense to model the effect of such shocks on real output as transitory, while allowing commodity supply shocks to have permanent effects on the production of that commodity.[6] Finally, commodity-market specific demand shocks such as inventory demand shocks do not affect either global real output or commodity production in the long run. There are no restrictions on the short-run effects of the structural shocks on the model variables.

2.3. Sign Restrictions

One limitation of the Kilian (2009) model is that it suggests that the spike in the price of oil following the invasion of Kuwait in 1990 was entirely caused by the oil-specific demand shock, which does not seem credible. This prompted Kilian and Murphy (2012, 2014) to explore

[6] This contrasts with the traditional approach in macroeconomics of viewing domestic demand shocks as having no long-run effect on real output and identifying any domestic shock that affects real output in the long as a supply or productivity shock (see Blanchard and Quah 1989).

alternative identification schemes that restrict the sign of selected impulse responses rather than imposing zero restrictions on B_0^{-1}. The advantage of sign-identified structural VAR models is that they allow the impact price elasticity of oil supply to be close to zero without requiring it to be literally zero (see Kilian and Murphy 2012). It should be noted that sign-identified oil market models cannot be used to validate the exclusion restrictions in oil market models, because the latter models are not nested by sign-identified models (see Kilian and Lütkepohl 2017).

2.3.1. The Kilian and Murphy (2012) Model

Kilian and Murphy (2012) proposed a model that is analogous to Kilian (2009) except that the identification is based on sign restrictions rather than exclusion restrictions.

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} - & + & + \\ - & + & - \\ + & + & + \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix} \tag{2}$$

The rationale is that an unexpected decline in the flow supply of oil shifts the oil supply curve to the left along the oil demand curve, causing global oil production to fall, the real price to increase, and global real activity to fall. In contrast, a positive aggregate demand shock raises global real activity and the real price of oil and stimulates global oil production. Finally, an increase in oil-specific demand (such as an increase in inventory demand) causes the real price of oil to increase, while stimulating global oil production and causing a decline in global real activity.

The key difference from other early examples of sign-identified oil market models such as Peersman and van Robays (2009) and Lippi and Nobili (2012) is that Kilian and Murphy (2012) emphasized the importance of bounding the impact price elasticity of oil supply, which is defined as the ratio of the impact responses of global oil production and of the real price of oil to an oil demand shock. They observed that there is a consensus in the literature that the one-month price elasticity of oil supply is near zero. Not imposing a bound on this elasticity means that we treat models that imply impact price elasticities of oil supply as high as 2 as equally plausible as models that imply elasticities close to 0. That approach is clearly unreasonable. Kilian and Murphy (2012) proposed a bound of 0.0258 for the one-month price elasticity of oil supply based on historical evidence (for related discussion see Kilian (2019a)). Kilian and Murphy concluded that it is not possible to generate large responses of the real price of oil to oil supply shocks, once the oil supply elasticity is restricted in this manner.[7] Once a reasonable supply elasticity bound is imposed, the responses to oil supply shocks in the Kilian and Murphy (2012) model are broadly similar to those reported in Kilian (2009). As Kilian and Murphy showed, even with a supply elasticity as high as 0.08, oil supply shocks would explain only 10% of the variation in the real price of oil.

Inoue and Kilian (2013) extend the Kilian and Murphy (2012) model to include additional dynamic sign restrictions. They restrict the real price of oil to be positive for the first year in response to unanticipated oil supply disruptions and in response to positive oil demand shocks. The validity of the identifying assumptions in the Kilian and Murphy (2012) and Inoue and Kilian (2013) model has been empirically assessed and confirmed in a number of studies including Lütkepohl and Netšunajev (2014) and Herwartz and Plödt (2016).

Antolin-Diaz and Rubio-Ramirez (2018) show how more precise estimates of the Kilian and Murphy (2012) model may be obtained by imposing additional narrative sign restrictions. Narrative sign restrictions refer to restrictions on the signs or relative magnitudes of structural shocks or of historical decompositions during selected periods, for which extraneous

[7] This conclusion has recently been confirmed for other oil market models (see Herrera and Rangaraju 2019).

evidence exists. For example, we know from oil industry sources that there was a surge in storage demand for oil between May and December 1979. Thus, it makes sense to impose the restriction that this feature also holds in the estimated oil market model. Similarly, we know that in August 1990 a negative oil supply shock took place when Iraq invaded Kuwait. Oil production in both countries ceased and the real price of oil rose. Any model that does not reproduce this feature clearly would be incredible. We also know that the sharp increase in the real price of oil in mid-1990 was not caused by increased flow demand, and we know that there must have been a simultaneous increase in storage demand, raising the real price of oil further, because otherwise oil inventories would have fallen sharply in response to the oil supply disruption. These considerations provide further restrictions on the historical decomposition of the real price of oil.

2.3.2. The Kilian and Murphy (2014) Model

The Kilian and Murphy (2014) model generalizes the Kilian and Murphy (2012) model by explicitly identifying shocks to storage demand (also referred to as speculative demand or inventory demand shocks). This extension is made possible by the inclusion of a proxy for global crude oil inventories, constructed by scaling U.S. crude oil inventories, as reported by the EIA, by the ratio of OECD petroleum inventories over U.S. petroleum inventories.[8] Global oil inventories are best expressed in changes (Δinv) rather than growth rates.[9]

The importance of modeling storage demand is that it allows oil price expectations to

[8] This proxy excludes Chinese crude oil inventories, which became important sometime after 2006, when China started building its strategic oil reserve. The Kilian and Murphy proxy also excludes crude oil stored on oil tankers at sea. A simple way of addressing this problem of measurement is to combine this proxy with the proprietary global crude oil inventory series compiled by the Energy Intelligence Group (EIG), which includes these and other missing components of global crude oil inventories starting in the 2000s. Sensitivity analysis in Kilian and Lee (2014) using this EIG inventory series confirms the substantive conclusions in Kilian and Murphy (2014).

[9] This stabilizes the variance of the oil inventory series. It also facilitates the computation of the price elasticity of oil demand and the imposition of identifying restrictions based on economic theory. Alternative approaches such as expressing inventory changes as a share of last month's oil production make it more difficult to impose standard identifying restrictions (see Kilian and Murphy 2014).

affect the real price of oil, even when oil price expectations cannot be directly observed. This fact allows the structural VAR model to capture shifts in oil price expectations not captured by the information set of the VAR model.[10] For added clarity, the oil supply and aggregate demand shocks in this model are relabeled as flow supply shocks and flow demand shocks, respectively.

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \\ u_t^{\Delta inv} \end{pmatrix} = \begin{bmatrix} - & + & + & b_{14}^0 \\ - & + & - & b_{24}^0 \\ + & + & + & b_{34}^0 \\ - & - & + & b_{44}^0 \end{bmatrix} \begin{pmatrix} w_t^{\text{flow supply}} \\ w_t^{\text{flow demand}} \\ w_t^{\text{storage demand}} \\ w_t^{\text{other oil demand}} \end{pmatrix} \tag{3}$$

Effectively, this model decomposes the residual oil demand shock in Kilian and Murphy (2012) into a storage demand shock and another oil demand shock that represents, for example, shocks to inventory technology and preferences as well as idiosyncratic shocks to the U.S. Strategic Petroleum Reserve that are not otherwise accounted for. This other oil demand shock is defined as the complement to the first three structural shocks and has no explicit structural interpretation. Negative flow supply and positive flow demand shocks are associated with declines in oil inventories, as refiners smooth the production of refined products. A positive storage demand shock, in contrast, increases oil inventories, the real price of oil and global oil production on impact, while lowering global real activity.[11]

As in Kilian and Murphy (2012), the one-month price elasticity of oil supply is bounded by 0.0258. The estimates are robust to relaxing this bound to 0.04 (see, e.g., Zhou 2019). In addition, the one-month price elasticity of oil demand is bounded by the fact that this elasticity

---

[10] Abstracting from changes in the risk premium, expectations shifts would also be captured by oil futures prices, but for much of the estimation period in Kilian and Murphy (2014), the oil futures market did not exist. Kilian and Murphy (2014) show that the oil futures spread, when it does exist, does not contain more information than already captured by the information set of the Kilian and Murphy (2014) model. This amounts to testing whether the VAR model is fundamental in the sense described in Kilian and Lütkepohl (2017).

[11] Kilian and Murphy (2014) originally did not impose the restrictions $b_{41}^0 < 0$ and $b_{42}^0 < 0$, but adding these restrictions tends to sharpen the results when conducting inference.

cannot exceed the corresponding long-run oil demand elasticity. The latter elasticity is bounded by 0.8 based on extraneous microeconomic estimates (see Hausman and Newey 1995). One legacy of the analysis in Kilian and Murphy (2012, 2014) is that it drew attention to the importance of oil supply and oil demand elasticities for interpreting the relationship between prices and quantities in the global oil market. Indeed, much of the recent controversy about how to model oil markets evolves around this question (see, e.g., Baumeister and Hamilton 2019a; Caldara, Cavallo and Iacoviello 2019, Kilian 2019a).[12]

In addition, Kilian and Murphy (2014) restricted the dynamic responses of global oil production and global real activity to a negative flow supply shock to be negative, and the response of the real price of oil to be positive for the first 12 months. This joint dynamic restriction can be motivated on economic grounds. Finally, Kilian and Murphy (2014) impose what nowadays would be referred to as narrative sign restrictions on the historical decomposition of the real price of oil (see Antolin-Diaz and Rubio-Ramirez 2018). Such restrictions are especially important when conducting inference in this class of models (see, e.g., Kilian and Zhou 2019a,b).

In model (3), storage takes place above the ground and is driven by the demand side of the oil market. Most storage is held by refineries in oil importing countries. Although it is possible to view the stock of oil left below the ground as another form of oil inventories, the latter type of inventories is economically distinct because it cannot be used by refineries to smooth the production of refined products in the event of an unexpected shortfall of domestic oil production or oil imports. Below-ground inventories are important in their own right because oil

[12] It should also be noted that the rationale for adding elasticity bounds is not to make the impulse response error bands narrower, although this may be a side-effect, but to eliminate as inadmissible models that are economically implausible. Including these models in the admissible set would bias the estimates of the oil market model (see Kilian and Murphy 2012).

producers expecting a higher price for future deliveries may withhold oil from the market and accumulate inventories below the ground, resulting in a reduction in flow supplies, a higher spot price, lower oil consumption and hence lower global real activity. Such a speculative supply shock, however, is observationally equivalent to any other disruption of flow supplies, say due to civil strife or a strike in the oil industry. Thus, Kilian and Murphy stress that, for all practical purposes, speculative supply shocks and flow supply shocks cannot be separately identified.

2.3.3. Larger-Dimensional Extensions of the Kilian-Murphy Framework

As with any structural model, the validity of structural oil market VAR models hinges on the premise that the model does not omit any important determinants of the real price of oil. An obvious concern is that decomposing the structural shocks further within an extended VAR model may change the substantive conclusions obtained in the baseline model. In general, it is difficult to assess the importance of omitted structural shocks, short of extending the model to include additional variables and imposing the restrictions required to identify the additional structural shocks.

A case in point is the analysis in Kilian and Murphy (2014) who showed that the substantive conclusions in Kilian (2009) and Inoue and Kilian (2013) are largely robust to including the change in global oil inventories and explicitly modeling storage demand. They also found, however, that for specific episodes such as the 1990 oil price spike there are noticeable differences in the interpretation of the data. By the same token, we need to ask how robust the conclusions in Kilian and Murphy (2014) are to further extensions of their model.

This question could not be addressed at the time Kilian and Murphy (2014) was written, because the estimation period was too short to consider models with more than four variables. There are several more recent studies, however, that suggest that the substantive conclusions of

Kilian and Murphy (2014) are robust to extending their model further. For example, Kilian and Zhou (2019a) confirm these findings based on a model that explicitly allows for shocks to the U.S. Strategic Petroleum Reserve. Cross (2019) reaches the same conclusion when differentiating between storage demand shocks driven by changes in the expected price of oil and changes in oil price uncertainty. Finally, Kilian and Zhou (2019b) confirm the robustness of the conclusions of Kilian and Murphy (2014) to modeling exogenous variation in the U.S. real market rate of interest and in the trade-weighted U.S. real exchange rate.

2.3.4. Incomplete Oil Market Models

There are also examples of studies that seek to simplify the benchmark models (1), (2) and (3) in the interest of greater parsimony. Just as we compared the Kilian-Murphy framework to larger-dimensional extensions of that framework in section 2.3.3, we need to ask whether such lower-dimensional models are capable of recovering at least approximately the estimates from the benchmark models (1), (2), and (3). If they do not, these lower-dimensional models must be considered incomplete and misleading.

A case in point is Bjørnland and Zhulanova (2019) who rely on a global oil market model of the form

$$\begin{pmatrix} u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_{11}^0 & 0 \\ b_{21}^0 & b_{22}^0 \end{bmatrix} \begin{pmatrix} w_t^{\text{flow demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix}. \tag{5}$$

Bjørnland and Zhulanova's discussion makes it clear that they have in their mind a simplified version of model (1) in which there are only two shocks, namely the oil supply shock and an oil-specific demand shock, which is identified by a delay restriction.[13] They argue that the oil supply

[13] In discussing their model, we deliberately abstract from the possible inclusion of a second block of domestic variables and the additional identifying assumptions required to separately identify domestic and global shocks. We also abstract from the fact that the global real activity measure may be any of the indicators discussed in Kilian and Zhou (2018) or the leading principal component of a panel of real activity indicators.

shock in model (1) may be dropped, given that Kilian (2009) showed that this shock plays only a modest role. This argument is problematic. Although oil supply shocks play only a modest role on average, for specific episodes such as the shale oil boom after 2010 their effect is quantitatively important. Moreover, there is overwhelming evidence for the importance of including oil inventories in oil market models and of modeling storage demand explicitly (see Kilian and Murphy 2014; Kilian and Lee 2014; Kilian 2017; Herrera and Rangaraju 2019). Thus, model (5) is misspecified in that it conflates the structural shocks in the data generating process.

This type of problem also arises in sign-identified models. For example, the identification of the oil market block in Baumeister and Peersman (2013b) boils down to the sign-identified model.

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} - & + \\ + & + \end{bmatrix} \begin{pmatrix} w_t^{\text{flow supply}} \\ w_t^{\text{flow } demand} \end{pmatrix}, \tag{6}$$

which is easily recognized as a restricted version of the model in Kilian and Murphy (2014) with two of the four structural shocks suppressed. It is immediately clear that such a model conflates the three distinct demand shocks contained in the fully specified model. More importantly, even if we restrict attention to the flow supply shock and treat this model as partially identified, there is no reason for the underspecified model (6) to be able to recover the flow supply shock in the data generating process because the information set is different.

2.3.5. Understanding the Impact of Oil Demand and Oil Supply Shocks on Domestic Macroeconomic Aggregates

Oil market VAR models are typically estimated on monthly data, because the exclusion and inequality restrictions used for identification tend to be more credible at monthly frequency than at quarterly or annual frequency. If we are interested in the effect of global oil demand and oil supply shocks on a monthly U.S. macroeconomic aggregate such as U.S. industrial production, a natural approach is to specify a block recursive VAR model with the domestic variable ordered last. A good example is the model of Kilian and Park (2009), who extended model (1) to include U.S. real stock returns under the maintained assumption that global oil market variables are predetermined with respect to the U.S. stock market, which implies that $b_{14}^0 = b_{24}^0 = b_{34}^0 = 0.$ The latter assumption is supported by evidence in Kilian and Vega (2011).

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \\ u_t^{ret} \end{pmatrix} = \begin{bmatrix} b_{11}^0 & 0 & 0 & 0 \\ b_{21}^0 & b_{22}^0 & 0 & 0 \\ b_{31}^0 & b_{32}^0 & b_{33}^0 & 0 \\ b_{41}^0 & b_{42}^0 & b_{43}^0 & b_{44}^0 \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \\ w_t^{\text{other shocks to stock returns}} \end{pmatrix} \tag{7}$$

Here the upper left 3×3 matrix contained in B_0^{-1} represents the oil market block and the lower right 1×1 matrix is the domestic block. Obviously, the same approach would work, if the oil market block were identified based on sign restrictions, except that the conventional approach to Bayesian inference would have to be modified, as described in Arias, Rubio-Ramirez and Waggoner (2018).[14] One can include without loss of generality additional variables in the lower right block as long as one restricts attention to the responses of these variables to oil demand and oil supply shocks.[15]

Sometimes this block-recursive VAR approach is not feasible. One common situation is that we are interested in the effect of global oil demand and oil supply shocks recovered from monthly VAR models on quarterly or annual U.S. macroeconomic aggregates. One approach to

[14] Further discussion of Bayesian estimation and inference in block recursive oil market model can be found in Kilian and Zhou (2019a,b).

[15] This does not mean that the shocks in the lower-right block cannot be identified. One example is Kilian (2010) who jointly modeled the global oil market block and a U.S. gasoline market block. Another example is Kilian and Zhou's (2019b) model of oil prices, exchange rates and interest rates.

this problem could be to estimate a mixed-frequency VAR model (see Ghysels 2016; Chudik and Giorgiadis 2019). A simpler approach proposed by Kilian (2009) is to sum (or, equivalently, average) the monthly structural shocks obtained from model (1) by quarter and to run a quarterly distributed-lag model second-stage regression. For example, consider the second-stage model

$$\Delta gdp_t = \alpha_i + \beta_{0,i} w_t^i + \beta_{1,i} w_{t-1}^i + ... + \beta_{h,i} w_{t-h}^i + v_{t,i}, \quad i = 1, 2, 3,$$

where $v_{t,i}$ denotes the possibly serially correlated and heteroskedastic regression innovation in equation i, w_t^i denotes the sum of the three realizations of the structural VAR shock i in quarter t, and $\partial gdp_{t+j} \big/ w_t^i = \partial gdp_t \big/ w_{t-j}^i = \beta_{j,i}$. Inference on the impulse responses may be conducted using HAC standard errors or the block bootstrap. This approach works because the structural shocks are mutually uncorrelated at monthly frequency and approximately mutually uncorrelated at lower frequency. Similar temporal aggregation schemes are commonly used in the literature on aggregating high-frequency monetary policy shocks. Efficiency may be gained by including current and lagged values of all shocks:

$$\begin{aligned}\Delta gdp_t = \alpha + \beta_{0,1} w_t^1 + \beta_{1,1} w_{t-1}^1 + ... + \beta_{h,1} w_{t-h}^1 + \\ \beta_{0,2} w_t^2 + \beta_{1,2} w_{t-1}^2 + ... + \beta_{h,2} w_{t-h}^2 + \\ \beta_{0,3} w_t^3 + \beta_{1,3} w_{t-1}^3 + ... + \beta_{h,3} w_{t-h}^3 + v_t,\end{aligned}$$

and imposing a common intercept.[16]

This two-stage approach has been widely employed in the literature (e.g., Kilian, Rebucci and Spatafora 2009); Kilian and Hicks 2013; Bützer, Habib and Stracca 2016). The same two-

[16] Alternatively, one could also estimate these responses based on linear local projections, but that approach would be less parsimonious (see Plagborg-Møller and Wolf 2019). It should be noted that, as in the local projection literature, using residuals from first-stage regressions creates a generated regressor problem, which is typically ignored in practice.

stage approach could be applied to sign-identified models, except that in this case inference is complicated by the fact that we need to evaluate the second-stage regression for each admissible draw from the posterior of the oil market model (see Herrera and Rangaraju 2019).

Another common situation in which block-recursive VAR models cannot be applied occurs when the U.S. macroeconomic aggregate of interest is not available for the full estimation period, but only for a comparatively short time span. In that case, it makes sense to recover the oil demand and oil supply shocks from a structural VAR model estimated on the full sample, but to fit the second-stage distributed lag model on the shorter subsample, even when the U.S. variable of interest is available at monthly frequency.

Because crude oil is traded in U.S. dollars, modeling the propagation of global oil demand and oil supply shocks to the U.S. economy is comparatively straightforward. The impact of oil demand and oil supply shocks on other net oil-importing economies, in contrast, will in addition depend on the value of the real exchange rate. Although some studies such as Chen (2009) fail to control for the real exchange rate in estimating the impact of oil demand and oil supply shocks on the domestic economy, the fact that the real price of oil in domestic consumption units depends on the value of the real exchange rate makes it essential to include the bilateral U.S. dollar real exchange rate in the block recursive VAR models. It is not clear how to employ second-stage distributed lag models in that case.

3. Estimation and Inference

Oil market models identified by short-run or long-run exclusion restrictions are typically estimated by the least-squares method or the GMM method with inference based on bootstrap methods, possibly making allowance for conditional heteroskedasticity in the error term (see Kilian and Lütkepohl 2017). In contrast, all sign-identified oil market VAR models in the literature have been estimated by Bayesian methods. Because these models are identified by inequality restrictions, they do not generate unique point estimates of the impulse responses, but a potentially large set of admissible models that are consistent with the data and satisfy the identifying restrictions.

3.1. Standard Bayesian Inference in Sign-Identified Models

The standard approach to estimating sign-identified models is based on Rubio-Ramirez, Waggoner and Zha (2010), Arias, Rubio-Ramirez and Waggoner (2018) and Antolin-Diaz and Rubio-Ramirez (2018). Consider a set of reduced-form VAR parameters consisting of the slope parameters, A, and of the reduced-form error covariance matrix, Σ_u, from which we can compute the lower triangular Cholesky decomposition, P, with positive elements on the diagonal. Candidate solutions for sign-identified models are generated by generating at random draw for the orthogonal matrix Q such that PQ represents a candidate solution for the structural impact multiplier matrix B_0^{-1}. Algorithms for generating random draws for Q are discussed in Kilian and Lütkepohl (2017). This procedure is repeated for each of many random draws from the posterior distribution of (A, Σ). Any combination of A and PQ is a model solution and is associated with a set of structural impulse responses. Model solutions that generate structural impulse response functions that satisfy all identifying restrictions on the impulse responses are considered admissible and are retained. All other solutions are discarded.[17]

The traditional approach to summarizing this set of admissible models has been to report

[17] When sign restrictions are augmented by zero restrictions on elements of B_0^{-1} or by narrative restrictions, these posterior draws must, in addition, be reweighted based on the importance samplers proposed by Arias et al. (2018) and Antolin-Diaz and Rubio-Ramirez (2018).

so-called posterior median response functions. This response function is constructed by connecting the posterior medians obtained from the marginal posterior distribution of each impulse response to form a line across the horizons of a given response function. Inference is based on the quantiles of the marginal distribution of the impulse responses. As noted by Fry and Pagan (2011), Kilian and Murphy (2012), Inoue and Kilian (2013), Kilian and Lütkepohl (2017) and Kilian (2019a), among others, this approach is highly questionable because it conflates the dynamic responses of different structural models and ignores the co-movement across impulse response functions. Put differently, the shape of median response function may look substantially different from that of any of the impulse response function that could conceivably be produced by the underlying model. Inoue and Kilian (2020a) prove that this problem may persist even when considering an infinite number of posterior draws.

The use of pointwise medians is not only a problem for impulse response inference. For example, the same problem applies to the vector of pointwise medians of the oil supply shock series obtained from the Baumeister and Hamilton (2019a) model, as reported on Hamilton's homepage. Likewise, vectors of pointwise quantiles of forecast error variance decompositions, of linear combinations of impulse responses, and of historical decompositions violate the adding up constraint, calling into question their interpretation. Thus, caution is called for in interpreting the results of oil market VAR studies based on vectors of pointwise posterior medians and related statistics.

The same concerns apply equally to vectors of any other quantile constructed from the marginal impulse response posterior distributions and hence invalidates the pointwise upper and lower quantile error bands commonly reported in applied work. Moreover, as is well known, pointwise error bands tend to understate the uncertainty about the impulse response estimates and provide no indication of what the likely departures from the preferred estimate of the response function are (e.g., Sims and Zha 1999; Inoue and Kilian 2013, 2020a). Their only justification is that until recently there did not exist alternative methods of inference that address these concerns.

As shown in Inoue and Kilian (2020a), a more compelling approach to Bayesian inference is to evaluate the mean, median or mode of the posterior density value of the models contained in the set of admissible structural models. One approach is to report the set of structural impulse responses that maximizes the joint posterior density of the structural impulse responses. Closed-form solutions for this joint density are readily available for most structural VAR models. The corresponding joint credible sets may be constructed as highest posterior density regions. This approach is optimal under Dirac delta loss, preserves the dynamics of the structural impulse response functions, and is easy to implement.[18]

An alternative are the estimators of the impulse response function derived in Inoue and Kilian (2020a) under quadratic and under absolute loss. These estimators are optimal in the relevant sense, they respect the dynamics of the impulse response function, and they are equally easy to implement. Inoue and Kilian (2020a) also show how to construct joint credible sets under absolute loss and under quadratic loss that capture the full uncertainty about the impulse response estimates, that provide information about likely departures from the path of the estimated impulse response function, and that are guaranteed to be contained in the set of feasible impulse responses functions. Thus, the key difference between the methods discussed in

[18] This approach was originally introduced by Inoue and Kilian (2013, 2019) for models identified by recursive zero restrictions or by sign restrictions and has been applied by Herwartz and Plödt (2016), Herrera and Rangaraju (2019), Zhou (2019) and Cross (2019), for example. Inoue and Kilian (2020a) generalize this approach to to allow for any combination of short-run and long-run exclusion restrictions, sign restrictions and short-run exclusion restrictions, and narrative inequality restrictions.

Inoue and Kilian (2020a) and the traditional approach of reporting posterior median response functions is not the choice of the loss function, but the focus on joint rather than marginal Bayesian inference.[19]

3.2. Alternative Bayesian Approaches to Estimating Sign-Identified Oil Market Models

The standard approach to Bayesian inference described in section 3.1 involves postulating a Gaussian-inverse Wishart prior for the reduced form parameters and a uniform prior for Q. Identifying restrictions are imposed directly on the structural impulse responses or transformations of these responses. Baumeister and Hamilton (2019a) recently reiterated the well-known fact that this two-step approach may be unintentionally informative about the structural impulse response functions. It should be noted that much the same problem may arise from imposing priors on the reduced-form parameters, since impulse responses are nonlinear transformations of these parameters (see Kilian and Zha 2002). The reason to be particularly concerned about the prior for Q is that this prior is never updated based on the data. How practically relevant this concern is for oil market models such as the Kilian and Murphy (2014) model is unclear.

It is important to recognize that the alternative Bayesian approach proposed by Baumeister and Hamilton, while dispensing with the prior for Q, suffers from exactly the same problem of inadvertently informative priors as the existing approach (see Kilian 2019b). There is nothing to choose between their approach and the conventional approach in that regard. There are only two studies that address the problem of unintentionally informative priors for Q. One

[19] Another approach sometimes used in the literature is to report the admissible model that comes closest to an extraneous demand elasticity estimate and to construct credible sets based on the admissible models that come closest to this estimate (see, e.g., Kilian and Lee 2014, Kilian 2017). The drawback of this penalty function approach is that we must be confident in the extraneous demand elasticity estimate.

approach is to adapt the prior on the reduced-form parameters, given a uniform prior on Q, to ensure that the implied prior on the structural impulse responses is flat (see Arias, Rubio-Ramirez and Waggoner 2015). The other approach is to directly specify a possibly informative prior on the structural impulse responses, as proposed by Plagborg-Møller (2019). The latter approach dispenses with reduced-form priors. Neither of these two approaches has been employed in the oil market literature to date.

Given that Baumeister and Hamilton's approach has been applied to oil market models, it is useful to discuss the differences between their approach and the conventional approach. In short, their proposal is to specify priors on the parameters of the structural VAR process

$$B_0 y_t = B_1 y_{t-1} + ... + B_p y_{t-p} + w_t ,$$

where w_t is zero mean Gaussian white noise. Abstracting from the lagged coefficients, this involves specifying priors directly on the elements of the matrix B_0. In special cases, it is also possible to impose priors on selected elements of B_0 and B_0^{-1}. For example, in their preferred model, the authors consider restricting one element of B_0^{-1}. It is important to understand that Baumeister and Hamilton's approach is not designed to handle the restrictions on B_0^{-1} typical of conventional oil market models, except in the special case of a recursively identified model.[20] Their proposal instead is to replace these models by an alternative oil market model that is more

[20] For example, Baumeister and Hamilton are unable to estimate either the Kilian and Murphy (2012) or the Kilian and Murphy (2014) model using their method. Their "replication" of the Kilian and Murphy (2012) model using their Bayesian methodology actually involves estimating a different structural model with additional identifying restrictions not imposed in the original model that generates impulse responses that look very similar to those in the Kilian and Murphy (2012) model.

amenable to their econometric approach, but differs in many dimensions from the current state of the art.[21]

The seeming advantage of Baumeister and Hamilton's method is that we can impose nondegenerate priors on parameters of the model. Rather than imposing that some element of B_0 is positive, for example, we can impose an explicit prior distribution of how the probability mass is distributed in the positive region. This advantage is more hypothetical than real. The problem is that, in practice, the existing literature provides little guidance about the nature of this prior. Baumeister and Hamilton therefore rely on ad hoc priors chosen for their computational tractability such as truncated Student-t distributions without any economic rationale. In some cases, they postulate diffuse priors intended to be agnostic. We discuss the consequences of the latter approach in section 5.3.

As shown by Herrera and Rangaraju (2019), despite substantial differences in the model specification and identifying assumptions, the specific four-variable oil market model proposed by Baumeister and Hamilton (2019a) generates responses to flow supply shocks similar to those in the Kilian and Murphy (2014) model, controlling for the bound on the impact price elasticity of oil supply. Likewise, the alternative estimation results reported by Baumeister and Hamilton for the Kilian (2009) model match those in the original paper. Thus, there is no evidence that their econometric methodology makes any difference for the substance of the results.

3.3. How Credible Are Time-Varying Coefficient Models of the Oil Market?

---

[21] An additional drawback of imposing restrictions on B_0 is that it forces the user to define the impact price elasticity of oil demand in terms of the parameters of this matrix (see, e.g., Caldara et al. 2019; Bruns and Piffer 2019). This approach is problematic when dealing with storable commodities, as discussed in section 5.2. Moreover, Baumeister and Hamilton define the elasticity holding constant the responses of all other model variables to an exogenous demand or supply shift, whereas extraneous microeconomic elasticity estimates do not control for these responses, making it impossible to appeal to this micro evidence in motivating the elasticity priors (see Kilian 2019).

A long-standing question is whether to allow the coefficients of global oil market models to vary over time or not. There is a multitude of reasons to expect smooth structural change in the coefficients of models of the global oil market. Examples include temporary capacity constraints in storage and production, transportation bottlenecks, changes in market structure and contract structure over time, changes in the share of oil in value added reflecting conservation and diversification away from oil products, and possibly the development of oil futures markets in the 1980s. This does not necessarily mean that the linear model is a poor approximation, but it raises the question of whether we need to consider the possibility of smooth structural change.

Baumeister and Peerman (2013a,b), in particular, made the case for estimating global oil market models as time-varying coefficient (TVC)-VAR models, building on Primiceri (2005). The coefficients in TVC-VAR models are typically expressed as latent random walk processes and estimated using Bayesian methods. The error covariance matrix is also allowed to change over time. Impulse responses in TVC-VAR models are dependent on the history of the data and the magnitude of the structural shocks and have to be evaluated by Monte Carlo integration.[22] The use of TVC-VAR models is restricted to the construction of impulse responses and counterfactual histories (see Kilian and Lütkepohl 2017). Historical decompositions and variance decompositions are not well defined, making it more difficult to interpret the model estimates.

There are a number of other caveats about the use of TVC-VAR models. First, there is evidence that the error bands for the nonlinear impulse responses in this class of models tend to be so wide that they convey essentially no information (see, e.g., Herrera and Rangaraju 2019).

Second, a common misperception in applied work is that time-variation in the estimated coefficients of a TVC-VAR model is evidence of time variation in the real world. This view is

[22] A common mistake in applied work is to report the impulse responses of the TVC-VAR model conditional on the date t estimate of the model coefficients without accounting or the expected evolution of the model coefficients.

erroneous because the parameters that govern the smooth structural change in the model coefficients are not identified in the absence of time-varying coefficients. In fact, evidence of time variation is expected due to overfitting, even when the linear model is correct.[23] There does not appear to exist a formal statistical test for the absence of time variation in TVC-VAR models.[24]

Third for computational reasons, TVC-VAR models of a given dimension can only be estimated allowing for a small number of autoregressive lags. Baumeister and Peersman (2013a), for example, restrict the maximum lag order of their VAR model to four quarterly lags. Since it is well known how important including a sufficiently large number of lags is for accurately estimating the impulse responses in global oil market models, this raises the question of whether any differences from the response estimates in linear VAR models reflect departures from linearity or unreasonably tight restriction on the largest autoregressive lag order.

Fourth, the specification and identification of the Baumeister and Peersman oil market model is superficially similar to that in Kilian and Murphy (2012). It should be noted, however, that Baumeister and Peersman (2013a) had to relax the oil supply elasticity bound in order to obtain any solutions at all for their econometric model. Their upper bound for the quarterly supply elasticity ranges from 0.6 to 1, respectively, which is unrealistically high compared with the quarterly microeconomic elasticity estimate of 0.017 reported in Newell and Prest (2019) or the elasticity bounds used in other studies.

[23] The degree of overfitting can be controlled to some extent by choosing the prior. However, this makes the estimates of TVC-VAR models highly sensitive to the specification of the prior distribution. This sensitivity is rarely reported in applied work.

[24] An informal diagnostic would be to compute the 95% joint error band for the responses to a one-standard deviation shock from the corresponding linear VAR model. This allows us to assess whether the nonlinear response functions generated by the TVC-VAR model are contained within this error band. If so, time variation is not likely to be important.

In defense of their approach, Baumeister and Peersman (2013a) suggest that the impact supply elasticity could be much higher than conventional elasticity bounds during some periods. Upon reflection, this argument is not compelling. The cost constraints that prevent oil producers from substantially adjusting production within the quarter do not vary much over time. In fact, Baumeister and Peersman's own arguments for time variation in the supply elasticity are all about why the supply elasticity may be closer to zero than suggested by pure cost considerations. They present no argument for why the supply elasticity should exceed conventional benchmarks. Thus, even if we take Baumeister and Peersman's arguments for a decline at face value, the question remains of how to explain their incredibly large estimates of the supply elasticity in the 1970s and early 1980s (with posterior median values as high as 0.85). Likewise, their estimates of the one-quarter price elasticity of oil demand for the 1970s and early 1980s (with posterior medians as low as -0.65) strain credulity.[25]

4. Common Mistakes in Estimating Oil Market VAR Models

There are a number of mistakes in estimating global oil market VAR models that are common enough to merit a separate section. One such mistake is to estimate the model on too short of a sample. The use of short estimation periods not only inflates the estimation uncertainty, but, more importantly, it calls into question the identification. The identification of structural VAR models relies on there being sufficient variation in the data driven by each shock. The shorter the estimation period, the less likely are we to encounter such variation. For example, if we focus on

[25] A partial explanation may be that the impact oil demand elasticity in Baumeister and Peersman (2013a) is incorrectly defined, as discussed earlier, and hence upward biased. This explanation does not apply to the impact supply elasticity, however.

a subperiod which is dominated by the cumulative effects of flow demand shocks, we will not be able to identify and reliably estimate the effect of supply or storage demand shocks.[26]

A second mistake is to combine data from the post-1973 era with data from the pre-1973 era. It is well documented that there was a structural break in the process governing the nominal price of oil (and hence the real price of oil) in late 1973 (see Alquist, Kilian and Vigfusson 2013). There are no proxies for the global price of oil for the pre-1973 era, but only U.S. data, and the U.S. price of oil was heavily regulated. Its evolution in levels follows a step function pattern. In growth rates, it exhibits long periods of no change, interrupted by sharp spikes. In contrast, starting in January 1974, the growth rate of the oil price behaves much like that of the price of any other internationally traded commodity. The reason is a fundamental change in the structure of the oil market that took place in late 1973. Thus, pre-1973 and post-1973 data should never be combined in the same VAR model. For the same reason, using VAR estimates from pre-1973 to pin down the prior for VAR models to be estimated on post-1973 data, as proposed by Baumeister and Peersman (2013a) and Baumeister and Hamilton (2019a) is not a good idea.

A third mistake is to rely on the WTI price of crude oil. It is well known that the price of domestically produced oil in the United States was subject to regulation until the early 1980s (see Mork 1989). It is less well known that between 2010 and 2015 arbitrage between the domestic price of oil in the United States and the price of oil in global markets broke down due to transportation bottlenecks (see Kilian 2016). Thus, the WTI price (and similar prices such as the U.S. producer price of crude oil) should not be used in modeling the global price of crude oil after 1974. Kilian (2009) proposes using the U.S. refiners' acquisition cost of crude oil imports instead as a proxy for the global price of crude oil. Alternatively, one could use the Brent price of

---

[26] A tell-tale sign of such identification problems is that estimates of the structural responses appear to change, when estimating the model on shorter subperiods.

crude oil, but the latter oil price has existed only since the mid-1980s, and hence is not a practical alternative.

A fourth mistake is to rely on much shorter lag lengths than recommended by Kilian (2009) and related studies. As discussed in Kilian and Lütkepohl (2017), the existence of long cycles in global commodity markets makes it essential to include at least two years worth of monthly lags in the VAR model. Models with shorter lags will fail to capture slowly building and declining cycles and will therefore understate the importance of flow demand shocks.[27] Likewise, the use of conventional lag order selection criteria such as the SIC or AIC is not recommended. Even the least parsimonious among these criteria tend to be downward biased in small samples. Such lag order estimates cannot be trusted. More generally, data-based model selection should be avoided, because it invalidates second-stage inference, as shown by Leeb and Pötscher (2005).

A fifth mistake is to mix nominal and real oil market variables. For example, Peersman and Van Robays (2009) propose an oil market VAR model that relates changes in the nominal price of oil to global real economic activity and global oil production. This specification is not only inconsistent with the underlying economic model, but invalidates their identification.

Another type of mistake is specific to oil market models that mix sign and zero restrictions on the impact multiplier matrix. Such oil market models have become increasingly common in recent years (e.g., Aastveit, Bjørnland and Thorsrud 2015; Kilian and Zhou 2019a,b). Candidate draws for the impact multiplier matrix in such models are constructed as $B_0^{-1} = PQ$, where $P = chol(\Sigma_u)$ is the lower triangular Cholesky decomposition of the reduced-form error

[27] Underfitting VAR models has also been shown to be a concern in studying the transmission of oil price shocks to the domestic economy (see Hamilton and Herrera 2004).

covariance matrix Σ_u and the rotation matrix Q is drawn subject to zero restrictions on selected elements of Q. The presence of these zero restrictions violates the conventional assumption that the rotation matrix Q is uniformly drawn from the space of all rotation matrices, invalidating the standard algorithm for generating posterior draws from sign-identified VAR models, as described in Rubio-Ramirez, Waggoner and Zha (2010). This problem may be addressed by reweighting the posterior draws based on the importance sampling procedure described in Arias, Rubio-Ramirez and Waggoner (2018).

Another mistake is to model the evolution of the price of oil based on U.S. oil market and macroeconomic data. Since oil is traded in global markets and the United States account for only a small fraction of the global economy and of global oil production, this approach inevitably suffers from an omitted variables problem, making it impossible to recover structural oil demand and oil supply shocks.

Finally, skepticism is called for whenever the names of "structural" shocks coincide with the names of observable model variables. This inevitably means that the shock in question is not structural. Examples include oil market studies referring to oil price shocks or real activity shocks. Likewise, studies equating innovations to real activity with oil demand shocks or innovations to oil production with oil supply shocks have not addressed the identification problem.

5. Recent Controversies About Modeling Oil Markets

Recently, some studies have claimed to have overturned the consensus based on mainstream oil market models that oil demand shocks are the primary determinant of oil price fluctuations. It is useful to review some of these claims and to put them into context.

5.1. Can Inequality Restrictions be Rejected by the Data?

As discussed in section 3.1, structural VAR models identified by inequality restrictions such as sign restrictions on impulse responses or narrative restrictions are typically evaluated by Bayesian methods based on an acceptance sampler algorithm. This means that we draw at random a larger number of structural models, but retain only those models that are admissible in that they satisfy the identifying restrictions. A common misperception among applied users is that a small fraction of admissible models means that the identifying restrictions are not supported by the data. As discussed in Kilian and Lütkepohl (2017), this argument is wrong. The fraction of the admissible model tells us nothing about the validity of the economic identifying assumptions. Nor does it indicate the reliability of the estimates. The latter depends only on the number of admissible models in the identified set.

There are different ways of illustrating this point. First, we need to recognize that the fraction of admissible models depends on the sampling algorithm. Choosing a more efficient algorithm can easily affect this fraction by a factor of 100. Whether efficient algorithms are available also depends on the structure of the model, so in general we cannot compare these fractions across models.

Second, clearly, if we dropped a valid identifying restriction, the fraction of admissible models would increase, but this does not mean that this modified model is more economically plausible than the original model. In fact, it would be less economically plausible. For example, Baumeister and Hamilton (2019a) suggest that their oil market model is superior because it is based on a much larger share of admissible models than existing models in the literature. What they fail to recognize is that this fact is largely driven by them having removed a key identifying assumption about the value of the oil supply elasticity.

Third, what a low fraction of admissible models means is that the identifying restrictions are informative. If these restrictions are valid a priori, more economic information (and hence a lower fraction of admissible models) is better by construction. In fact, omitting relevant identifying information will bias the estimates, as stressed by Kilian and Murphy (2012) and Antolin-Diaz and Rubio-Ramirez (2018), among others.

5.2. How to Define the Impact Price Elasticity of Oil Demand

Whereas the definition of the impact price elasticity of oil supply is uncontroversial, there has been some confusion in the recent literature about the definition of the impact price elasticity of oil demand. The cause of this confusion is that some researchers have uncritically applied a textbook definition of this elasticity intended for goods that are not storable. Since oil is a storable commodity, defining the oil demand elasticity based on an accounting identity that equates oil production with oil consumption at each point in time, as proposed by Baumeister and Hamilton (2019a) and Caldara et al. (2019), is obviously incorrect.

The impact price elasticity of oil demand is properly defined as the change in the use of oil in response to an oil price increase caused by an exogenous shift in oil supply. In a global oil market model, the amount of oil produced in a given period may be consumed in a refinery or put into storage.[28] Much of the storage is owned by refiners in oil-importing economies. This point has important implications in practice. Not only is the impact price elasticity of oil demand not well defined in an oil market model that does not include oil inventories, but, even when oil inventories are included among the model variables, computing this elasticity without accounting

[28] When modeling individual countries, this definition has to be augmented to include net oil exports (see Kilian 2017).

for the response of oil inventories is incorrect. As illustrated in Kilian and Murphy (2014), this tends to make the demand elasticity estimate look larger than it truly is.

This fact implies, first, that the impact price elasticity of oil demand is not identified in the oil market models of Kilian (2009) and Kilian and Murphy (2012). Thus, Baumeister and Hamilton's (2019a) claim that that the impact oil demand elasticity is implausibly large in these two models is without support. Second, VAR studies that impose priors an incorrectly defined demand elasticity estimates such as Baumeister and Hamilton (2019a) and studies that use incorrect demand elasticity estimates for choosing among alternative VAR model specifications such as Caldara et al. (2019) are misspecified. Third, Caldara et al. (2019) and Baumeister and Hamilton (2019a) mistakenly proceed as if including a proxy for changes in global oil inventories to the oil market model allows them to equate oil production and oil consumption in computing the oil demand elasticity. Fourth, Baumeister and Hamilton (2019a) suggest that a low price elasticity of oil supply necessarily implies an unrealistically high price elasticity of oil demand. This claim, which is also central to the analysis in Caldara et al. (2019) and Bruns and Piffer (2019), is invalid and driven by their use of an incorrect definition of the demand elasticity. A counterexample is the evidence in Kilian and Murphy (2014) whose model has a price elasticity of oil supply close to zero, but a lower price elasticity of oil demand than the preferred model of Baumeister and Hamilton (2019a) which is based on a much higher price elasticity of oil supply.

5.3. Is There Evidence for a Much Higher Impact Price Elasticity of Oil Supply?

Kilian and Murphy (2012) established that the effect of oil supply shocks on the real price of oil is necessarily modest, if the impact price elasticity of oil supply in model (2) is close to zero, as suggested by microeconomic theory and extraneous microeconomic evidence. It is this fact alone that ensures that oil demand shocks are the primary determinant of the variability in the real price of oil. This point has been reaffirmed in the context of model (3) by Herrera and Rangaraju (2019).

Baumeister and Hamilton (2019a) recently claimed to have overturned the consensus that oil demand shocks are the most important determinant of the real price of oil. They proposed a alternative model of the global oil market that allows the impact price elasticity of oil supply to be unbounded from above. Their posterior median estimate of 0.15 is an astronomical 22 standard errors above the extraneous elasticity estimate in Newell and Prest (2019). They show that, not surprisingly, given this large supply elasticity value, the responses of the real price of oil to oil supply shocks are much larger in their model than in conventional oil market models.[29]

Herrera and Rangaraju (2019), however, establish that this conclusion is not robust. Under any prior that bounds the impact price elasticity of oil supply in line with conventional views of the magnitude of this elasticity, the response of the real price of oil to oil supply shocks in Baumeister and Hamilton's model is similar to that obtained from Kilian and Murphy's model. In other words, the substantive conclusions of Kilian and Murphy (2014) are reaffirmed.

Thus, the magnitude of the upper bound on the value of the one-month oil supply elasticity is central for this controversy. Baumeister and Hamilton (2019a,b) suggest that there is new evidence that shows that this elasticity could be as high as 0.9. Specifically, they cite two recent studies by Caldara et al. (2019) and Bjørnland, Nordvik and Rohrer (2019). Kilian (2019a), however, demonstrates that Bjørnland et al.'s microeconomic estimate of the oil supply elasticity in the Bakken is not identified, is based on an economically implausible model, and is

[29] It should be noted that there are many problematic features of their model, as reviewed in Kilian and Zhou (2019c) and Kilian (2019b), but these features are not the main driver of their results and hence can be ignored for the purpose of this survey.

at odds with direct evidence from industry sources, while the IV supply elasticity estimates for selected oil-producing countries in Caldara et al. suffer from a violation of the exclusion restriction. Correcting for that violation produces oil supply elasticity values very close to the estimate in Newell and Prest (2019).

Thus, there is no credible evidence that Baumeister and Hamilton (2019a) overturn the consensus view about the relative importance of oil demand and oil supply shocks. Likewise, Caldara et al.'s (2019) and Bruns and Piffer's (2019) conclusion that the oil supply shock is a more important determinant of the real price of oil than oil demands shocks is incorrect. There is no support for their preferred oil supply elasticity estimates. This is not surprising because their analysis directly builds on Baumeister and Hamilton's work and hence suffers from some of the same errors. This debate illustrates the importance of incorporating economic theory, extraneous microeconomic evidence and institutional knowledge of the oil industry in the specification of oil market VAR models.

5.4. Is the Price Elasticity of Oil Demand Positive?

Most oil market VAR models including Kilian and Murphy (2014) impose the assumption that the impact price elasticity of oil demand is negative. Sockin and Xiong (2015) make the striking claim that these models are inherently misspecified because, according to their own analysis, the price elasticity of oil demand actually is positive rather than negative. This view is based on a theoretical model of informational frictions. In short, Sockin and Xiong's premise is that rising oil futures prices signal a stronger global economy. Their theoretical model postulates that economic agents have no other means of detecting whether the global economy is booming but to observe oil futures prices. This friction ensures that agents in their model habitually confuse increases in the oil price driven by oil supply shocks or storage demand shocks with increases driven by flow demand shocks.

Sockin and Xiong show that the informational content of increases in oil futures prices in their theoretical model can be so strong that it offsets the dampening effect of higher oil costs on manufacturing activity, resulting in a positive impact price elasticity of oil demand. In other words, Sockin and Xiong argue that higher oil prices induce manufacturing firms to buy more oil. Since the Kilian and Murphy (2014) model imposes a negative oil demand elasticity, it is misspecified under Sockin and Xiong's assumptions.

Sockin and Xiong apply their model to the period of the sustained commodity price boom between 2003 to mid-2008. There argue that an exogenous increase in speculative trading in oil futures markets that was reflected in higher oil futures prices caused manufacturing firms to increase their demand for raw materials such as crude oil in anticipation of an economic boom. Such a nonfundamental demand shift would look like a flow demand shock in the Kilian and Murphy (2014) model rather than a shock to speculative demand.

There are many reasons to be skeptical of the Sockin-Xiong model. First, even if we accept the premise that manufacturing firms may at times be confused about the state of the global economy, the argument that manufacturing firms collectively increased their demand for commodities and their output based on the false premise of a booming global economy for five years without realizing that there was no demand for their products is simply not credible. A model in which agents are confused for five years at a time and do not learn from their mistakes is a model of irrational expectations. In the real world, firms that make such systematic mistakes go out of business.

Second, there is no extraneous microeconomic evidence that the one-month price elasticity of oil demand is positive. Nor is there evidence that economic agents predicted a sustained economic boom between 2003 and mid-2008, as maintained by Sockin and Xiong. In fact, Kilian and Hicks (2013) document that professional real GDP growth forecasters systematically underestimated global growth between 2003 and mid-2008.

Third, the very existence of the informational friction postulated in Sockin and Xiong's model is suspect. It is by no means necessary for economic agents to directly observe oil demand and oil supply shocks in order to detect a global economic expansion or decline.[30] There are many indicators of global real economic activity that are readily available to economic agents and allow them to learn whether there actually is an economic boom or not without having to rely exclusively on industrial commodity prices (see Kilian and Zhou 2018). This makes it implausible that manufacturing firms would ever blindly rely on industrial commodity prices as indicators of the direction of the global economy. If they do not, then the mechanisms described by Sockin and Xiong (2015) are irrelevant.

Finally, it should be noted that there is no support for Sockin and Xiong's additional premise that the increased participation of financial traders in oil futures markets was an exogenous event, as opposed to an endogenous response to fundamental demand or supply pressures and rising expected returns on oil. Nor is there support for the premise that these financial traders necessarily took positions that raised the oil futures price. For a comprehensive review of this debate the reader is referred to Fattouh et a. (2013).

5.5. Has the Shale Oil Revolution Undermined the Stability of Global Oil Market Models?

[30] Sockin and Xiong (2105) motivate their use of informational frictions based on the false argument that Kilian and Murphy's (2014) model requires economic agents to directly observe both oil demand and oil supply shocks. There is, of course, no such assumption in the structural VAR literature.

An important question is whether the U.S. shale oil revolution that took place after 2008 has undermined the stability of global oil market models.[31] Kilian (2017) directly addresses this question by constructing a counterfactual for global oil production in the absence of U.S. shale oil production. He notes that the shale oil revolution can be represented as a sequence of classical shocks to the technology of producing crude oil. Kilian (2017) shows that the sequence of global flow supply shocks that would be required to remove the shale oil component from global crude oil production involves shocks that are neither unusually large nor unusually serially correlated by historical standards. Thus, there is no reason to expect these shocks to have undermined the stability of the coefficients of the structural model.

Another common misperception is that the shale oil revolution must have increased the value of the one-month price elasticity of flow supply. Not only is shale oil only a small fraction of global oil production, limiting the effects of the shale oil supply elasticity on the one-month price elasticity of oil supply in global oil market models, but Newell and Prest (2019) provide independent microeconomic evidence that the one-month price elasticity of oil supply for shale oil is very close to zero, much like the corresponding elasticity for conventional crude oil. This conclusion is also supported by direct evidence from industry sources.

5.6. How to Measure the Global Business Cycle in Oil Market Models

An important question in models of the global oil market is how to measure the global business cycle at monthly frequency. Kilian (2009) proposed a widely used measure of global real economic activity that is based on a proxy for the volume of shipping of industrial raw materials. It is well known that changes in trade volumes need not line up with changes in real output, as

[31] For a review of the U.S. shale oil revolution the reader is referred to Kilian (2016).

measured by world real GDP or world industrial production (see Kilian and Zhou 2018).[32]

The Kilian index was, in fact, constructed as an alternative to world real GDP which is not only conceptually inadequate for modeling industrial commodity markets, but poorly measured and available only at quarterly or annual frequency. The Kilian index was also designed as an alternative to the indices of global industrial production that have been favored by some recent studies.[33] Not only is the relationship between global industrial production and the volume of shipping potentially unstable over time, but the timing of these indices differs because industrial commodities tend to be shipped before changes in industrial production take place. Likewise, the evolution over time of these indices differs, because the Kilian index embodies an expectational component that is missing in indices of global industrial production because the decision to ship these raw materials is made in expectation of future industrial production (see Kilian and Zhou 2018; Kilian 2019b). Thus, there is no a priori reason for these indices to behave similarly, although it has been shown that most of the substantive results in Kilian and Murphy are robust to replacing the Kilian index by the OECD+6 industrial production index created by the OECD (see Zhou 2019).[34] Another caveat that applied users must be aware of is that, contrary to the claim in Baumeister and Hamilton (2019a), neither the Kilian index nor the global industrial production index is a measure of global income, making it impossible to define or restrict income elasticities in this type of model.

---

[32] Alternative proxies for global real economic activity that recognize that industrial commodities are traded are indices of global real industrial commodity prices (e.g., Delle Chiaie, Ferrara, and Giannone 2017; Alquist, Bhattarai and Coibion 2019).

[33] Some studies rely on factor-augmented VAR models to construct indices of global real output from panels of industrial production and/or real GDP series for many countries (see, e.g., Aastveit, Bjørnland and Thorsrud 2015). The latter approach suffers from the same conceptual drawbacks as conventional indices of global industrial production and global real GDP.

[34] Not all models are robust to the choice of the index. For example. the estimates in Baumeister and Hamilton (2019a) change substantially when replacing their industrial production index by the Kilian index (see Herrera and Rangaraju 2019).

An important question, in practice, is whether to express indices of global real economic activity in deviations from trend or in growth rates. In the case of the Kilian (2009, 2019b) business cycle index, this question does not arise because the index is stationary by construction (see Kilian and Zhou 2018). It obviously does not make sense to difference this business cycle index, although some applied studies have made that mistake.[35] When using alternative indicators of global real activity such as the index of OECD+6 industrial production originally created by the OECD, the standard approach has been to express this index in deviations from a log-linear time trend. This approach is natural if we are interested in capturing the global business cycle. The alternative of expressing indicators of global real activity as month-by-month growth rates, in contrast, eliminates long cycles in the real price of oil that are characteristic of commodity markets. It also overemphasizes the high-frequency variation in the data. As a result, log-differencing tends to downplay the importance of flow demand shocks.[36] It may also suggest that the estimates are less sensitive to increasing the lag order than is actually the case.

5.7. How to Transform the Real Price of Oil

A recurring question among applied researchers is whether to express the real price of oil in oil market models in log-levels or in growth rates. There has been no apparent trend in the log real price of oil since 1974. Thus, the conventional approach of expressing the real price of oil in log-levels has the advantage that standard frequentist inference about the estimates of the

[35] In related work, Hamilton (2019) makes the claim that there is statistical evidence against the data transformations applied by Kilian (2009, 2019b) and that the raw data underlying this index should be log-differenced. This claim is econometrically invalid, as discussed in Kilian (2019a).

[36] An alternative is to express indices of global real activity as cumulative growth rates over one year, as discussed in Kilian and Zhou (2018), or over two years, as proposed by Hamilton (2019). It is unclear what the rationale of the latter procedure is, except as descriptive statistic. By construction, cumulative growth rates do not provide a good measure of the business cycle at a given point in time.

impulse responses under weak conditions will remain asymptotically valid, even when the underlying process contains a unit root (or is possibly cointegrated with other variables) (see Inoue and Kilian 2020b). The same is true for forecast error variance decompositions at any finite horizon. In contrast, differencing the real price of oil when it actually is stationary causes these estimates to be inconsistent and inference to be invalid. Likewise, from a Bayesian point of view, inference remains valid whether the real price of oil is I(0) or I(1). Only the construction of historical decompositions requires the user to take a stand on the order of integration of the real price of oil.

Unit root tests (or for that matter so-called stationarity tests) are not able to discriminate between the I(0) and the I(1) hypothesis for realistic sample sizes, but the fact that the real price of oil has ultimately reverted back to its mean for almost four decades, defying predictions of permanent highs and permanent lows time and again, is suggestive of an I(0) process. This view is also consistent with evidence from the literature on forecasting the real price of oil, which has found that autoregressive models in log-levels tend to be more accurate out of sample than models in log-differences.

6. Non-Traditional Approaches to Identifying Oil Demand and Oil Supply Shocks

While the bulk of structural oil market models seeks to recover structural shocks based on the information set provided by a VAR model, there is a smaller literature seeking to exploit extraneous estimates of oil supply and oil demand shocks. Such extraneous shock measures may be used as an external VAR instrument, as an internal VAR instrument, or as regressors in distributed lag models or local projection models (see Plagborg-Møller and Wolf 2019).

6.1. Historical Counterfactuals for OPEC Events

Given the importance attached to OPEC actions in the early oil market literature, there has been much interest in identifying exogenous oil supply shocks in OPEC countries. For example, Hamilton (2003) proposed a simple measure of OPEC oil supply shortfalls caused by exogenous geopolitical events. Kilian (2008b) observes that not all of the events considered by Hamilton are plausibly exogenous. He also points out that Hamilton's OPEC oil supply shock measure is based on indefensible assumption about the timing and magnitude of the OPEC oil supply disruptions. Kilian (2008b) proposes an improved measure that articulates explicit counterfactuals about how the production of other OPEC members would have evolved in the absence of geopolitical events in selected OPEC member countries. Kilian's (2008b) exogenous OPEC oil supply shock series has been updated by Bastianin and Manera (2018). Unlike Hamilton, Kilian allows OPEC oil supply shocks to be negative as well as positive. He draws attention to the fact that political events such as wars paradoxically may stimulate oil production rather than necessarily curtailing it. For example, during the Iran-Iraq War, both of these countries sought to increase their oil exports in order to finance purchases of military equipment from abroad.

Kilian (2006) uses Kilian's (2008b) OPEC oil supply shock measure as an internal instrument within an extended version of the Kilian (2009) oil market model and shows that explicitly modeling OPEC oil supply shocks provides no value added. Montiel Olea, Stock and Watson (2018) establish the robustness of the results in Kilian (2009) to using the same OPEC oil supply shock measure as an external instrument.

In practice, the usefulness of both the Kilian (2008b) and the Hamilton (2003) measure of exogenous OPEC oil supply shocks for applied work is limited for two reasons. First, Kilian (2008b) stresses that all such measures lack predictive power for the real price of oil. This not only contradicts Hamilton's (2003) claim that these oil supply shocks explain major oil price fluctuations, but means that these shocks are weak instruments for the real price of oil, which complicates estimation and inference. This point has been reinforced by more formal evidence in Kilian (2008a) and Montiel Olea, Stock and Watson (2018).

Second, Kilian (2006, 2009) emphasize that OPEC members such as Saudi Arabia and to a lesser extent Kuwait and the UAE have had a history of responding to exogenous production shortfalls in other OPEC countries by expanding their own production, so the net shortfall of oil often is much smaller than it seems at first sight. Examples include the Iranian Revolution of 1978/79 and the invasion of Kuwait in 1990. Moreover, the importance of OPEC has declined since the 1970s. OPEC oil production shortfalls since the 1980s have been increasingly offset by production increases in non-OPEC countries, making it important to measure oil supply shocks at the global level rather than at the OPEC level. This evidence, along with the increasing recognition that OPEC oil supply shocks are weak instruments at best, has led to the demise of the literature on OPEC oil supply shocks caused by geopolitical events. There have been a number of other efforts in the recent literature, however, to construct extraneous estimates of oil demand and oil supply shocks.

6.2. Changes in Oil Futures Prices Around OPEC Announcements

One increasingly popular approach has been to rely on event studies. For example, Känzig (2019) uses changes in 6-month oil futures prices on days of OPEC announcements as an instrument to identify what he calls an oil supply news shock. His premise is that a change in the oil futures price at high frequency reflects a shift in oil price expectations rather than in the risk premium. These daily shocks are aggregated to monthly frequency by summing the daily shocks over a given month. An increase in oil price expectations (referred to as negative supply news by

Känzig) is associated with an immediate increase in the spot price of oil, a gradual fall in oil production and an increase in oil inventories.

It can be shown that the oil supply news shock discussed in Känzig (2019) is a special case of a storage demand shock driven by the change in oil price expectations, as discussed in Kilian and Murphy (2014). This proposition is testable. Since we know that unexpected oil supply disruptions cause a decline in oil inventories, whereas positive storage demand shocks cause an increase in oil inventory holdings, it is immediately clear from Känzig's impulse response estimates that OPEC announcements that raise oil futures prices affect the real price of oil through storage demand rather than through changes in the flow supply of oil. This conclusion is also consistent with the sharp increase in the real price of oil following the supply news shock. Although Känzig's shock measure captures only a subset of the storage demand shocks identified in recent structural oil market models, it corroborates the dynamic responses to storage demand shocks uncovered by Kilian and Murphy (2014).

Känzig's approach is not without limitations. For example, the assumption of no change in the risk premium around OPEC announcements need not be correct in practice. Likewise, the temporal aggregation of the daily shock measure to monthly frequency is ad hoc.[37] Nor is it clear which horizon of the term structure of oil futures we should focus on in constructing this shock. Most importantly, the interpretation of this shock as oil supply news is somewhat misleading. The extent to which supply news affects oil price expectations depends on expectations of oil demand. In other words, the same OPEC supply news will have different effects on oil price expectations depending on expected demand. This does not affect the validity of Känzig's

[37] For alternative approaches that take account of temporal aggregation see Ghysels (2016) and Chudik and Georgiadis (2019).

results, but their interpretation. A more natural interpretation of his shock would be as a shock to oil price expectations.

This comment also applies to other oil supply news measures discussed in the literature. For example, Arezki, Ramey and Sheng (2017) treat large oil discoveries as news shocks about future oil output. Oil discoveries by construction leave the flow of oil production unaffected for years to come. Instead, they matter because they shift expectations of future oil production and hence expectations of future oil prices. Thus, these shocks do not represent shocks to the flow supply of oil, but to storage demand.

It is worth pointing out that oil supply news measures such as giant oil discoveries are poor proxies for exogenous shifts in oil price expectations. In fact, the link between oil discoveries and oil price expectations is nonlinear in general. An unexpected giant oil discovery will not move oil price expectations, when expected oil demand is low and the market is well supplied, but may lower oil price expectations substantially, when expected oil demand is high. Just because an oil discovery has a large effect on the economy of the oil producer (and possibly on the real price of oil) in one episode does not necessarily mean that it will do so in another episode. Thus, the coefficients of a regression of changes in the real price of oil on this shock measure will be time-varying. The same concern applies to other oil supply news measures such as indices of attacks on oil shipping.

If we force the coefficients to be time-invariant, we obtain a linear approximation, the accuracy of which is sensitive to the unmodeled demand side of the oil market over the estimation period. In contrast, storage demand shocks, as measured in the Kilian and Murphy (2014) model avoid this instability by focusing on the shift in storage demand associated with a shift in oil price expectations. This approach avoids having to measure oil price expectations or having to model the nonlinear process that generates oil price expectations.

6.3. Forecast Revisions for Global Growth Forecasts

Kilian and Hicks (2013) construct proxies for global flow demand shocks based on revisions of forecasts of real GDP growth made by professional forecasters, as recorded by the Economist Intelligence Unit. They document that professional forecasters between 2003 and mid-2008 persistently underestimated global growth, mainly because they underestimated growth in emerging Asia. This pattern is consistent with the pattern of flow demand shocks recovered by the Kilian and Murphy (2014) model.

6.4. The Narrative Approach to Identifying Oil Supply and Oil Demand Shocks

An alternative is the narrative approach to identification. An early example is Cavallo and Wu (2012). Their approach involves a manual audit of articles published in the *Oil Daily*, the *Oil & Gas Journal*, and the *Monthly Energy Chronology* between 1984 and 2007. Cavallo and Wu use introspection informed by articles in these industry publications to attribute daily changes in the spot price of oil (or, alternatively, the residual of a regression of the change in the spot price on the oil futures spread) to a subset of 22 different types of oil-market related events such as OPEC announcements on oil production, U.S. oil inventory announcements, political developments in the Middle East, or oil production or transportation disruptions that are considered exogenous with respect to the oil price by the authors. If there is more than one event on a given day, say, n events, $1/n$ of the change in the price of oil on that day is attributed to each event. Cavallo and Wu then proceed to construct monthly averages of the daily oil price changes that they attribute to exogenous events. They abstract from inflation in defining their exogenous oil price shocks.

One obvious concern with this methodology is that the events in question are not necessarily the cause of these oil price changes. For example, an announcement about oil inventories (or of OPEC production plans) would be expected to have no effect on the price of oil to the extent that the announcement was expected (e.g., Ye and Karali 2016). Nor is it clear to what extent political events in an OPEC country cause the price of oil to move. Another concern is that many of the events in question are not plausibly exogenous with respect to the change in the price of oil. For example, OPEC production plans are endogenous with respect to the state of the global economy. Nor is there any justification for giving equal weight to different events. Finally, one has to have a lot of faith in the ability of the authors (and, implicitly, in the ability of the oil journalists whose articles form the basis of the analysis) to solve the underlying identification problem. The problem faced by these journalists is little different from that faced by sports commentators or stock market pundits explaining the latest results after the fact. There is a natural tendency to list all possible explanations with little regard to the consistency of the explanation over time. It is these explanations that form the raw material for the analysis in Cavallo and Wu (2012). Many of these problems carry over to more recent narrative approaches to identifying oil demand and oil supply shocks.

6.5. Text-Based Measures of Oil Demand and Oil Supply Shocks

Cavallo and Wu's (2012) analysis focuses on measuring exogenous oil price shocks associated with oil demand and oil supply shocks. It does not provide direct measures of the actual demand or supply shocks, only of their oil price impact. The direct measurement of oil supply and/or oil demand shocks based on textual analysis is discussed in Caldara, Cavallo and Iacoviello (2019) and Datta and Diaz (2019).

Caldara, Cavallo and Iacoviello's (2019) Oil Supply Shock Measure

Caldara et al. (2019) use articles in the *Oil Daily* published by the Energy Intelligence Group and the *Oil Market Report* of the International Energy Agency to classify selected disruptions in countries' oil production as exogenous because, in the judgment of the authors, these cutbacks were due to geopolitical events, natural disasters, strikes or other events that the authors consider exogenous with respect to a country's level of oil production. Ultimately, the classification of events as exogenous is a judgement call and, not surprisingly, some of Caldara et al.'s classifications have been called into question (see Kilian 2019a). It is also noteworthy that Caldara et al. restrict attention to oil supply disruptions. They do not allow for exogenous oil production increases.

Datta and Diaz' (2019) Oil Supply and Oil Demand Shock Measures

As in Cavallo and Wu (2012), the textual analysis in Caldara et al. (2019) is conducted manually by the authors. More recently, Datta and Diaz (2019) have proposed using a systematic and fully automated procedure for gathering information from news articles in two oil industry publications by the Energy Intelligence Group. Their objective is measuring the supply- and demand-driven components of oil price movements. By categorizing words in these articles into expressions linked to "oil supply", "oil demand", "increase", and "decrease", they construct oil supply and oil demand indicators, demeaned values of which they treat as proxies for oil demand and oil supply shocks. While the scope of their analysis is much more ambitious than that in Caldara et al. (2019), it is also less constrained by direct measures of oil market data.

Datta and Diaz' (2019) approach is problematic for several reasons. First, their claim that text searches allow them to circumvent delays in the availability of data on oil production and global real activity is incorrect. Given the six-month delay in the availability of global oil production data, for example, oil industry journalists clearly know as little as economists about the current level of global oil production. They have no advance knowledge of the data.

Second, there is reason to be skeptical of the reliability of these text classifications. Just because the process is automated does not mean that it reads the data correctly. For example, policymakers routinely mislabel observed oil production as “supply” and observed oil consumption as “demand”. Since oil consumption may drop in response to an oil supply disruption, it is easy to see how a mechanical text search could confuse oil supply with oil demand. Another concern is that Datta and Diaz’ method does not accommodate negations such as “not a supply increase”, which will be coded as a “supply increase”, overturning the meaning of the original text. Another problematic situation arises with texts such as “OPEC oil production will increase its production next year by less than expected”. The textual search will classify this sentence as a “supply increase”, when in reality it reflects a decline in expected oil production, which, all else equal, causes an increase in storage demand. More generally, Datta and Diaz do not separate expected changes from actual changes. In addition, they drop search terms such as “contract*” which may be activated by the words “contracts” as well as “contraction”, which risks miscounting sentiment about global demand.

Third, an index of the frequency of certain words or phrases in the press, on television, or in social media (say the phrase “decline in oil supply”) is not a shock, but an endogenous variable, calling into question Datta and Diaz’ (2019) interpretation of their data. It may seem that the unpredictable component of the demand (supply) indicator, would be a natural measure of the demand (supply) shock. This misconception arises from a semantic confusion about the terms “demand” and “supply”. The term “demand” among policymakers and in news reports is short-hand for global real activity or oil consumption. Clearly, a surprise change in these variables could be caused by any combination of oil demand or oil supply shocks. The underlying structural shocks are not identified. The same is true for unexpected changes in the "supply" of oil, which in press reports refers to oil production. Thus, the resulting "supply shock" measure would not be a structural shock either. Clearly, there is no reason for the magnitude of this index (or its unpredictable component) to be proportionate to the magnitude of actual oil supply shocks.

Fourth, as the oil market VAR literature has shown, there are many different types of oil demand shocks with different effects on the real price of oil. Lumping all these demand shocks together in textural analysis, as proposed by Datta and Diaz (2019), is not likely to produce a sensible measure of oil demand shocks.

Finally, there is no reason to presume that the press always correctly identifies supply and demand shocks. For example, the oil supply disruption that occurred as a result of Hurricane Rita and Katrina in 2005, when offshore oil platforms in the Gulf of Mexico were shut down, was negligible on a global scale. The main effect of these hurricanes was to shut down the U.S. refining industry along the Gulf Coast. This shutdown by construction represented a negative supply shock for the U.S. gasoline market, but a negative demand shock for the global oil market. Nevertheless, this event was frequently incorrectly characterized as a negative oil supply shock in the media. Another good example is Iraq's invasion of Kuwait in August 1990, which represented both an oil supply disruption and a positive shock to storage demand, reflecting expectations of future oil supply disruptions and hence rising oil prices. The latter explanation received much less attention in the media, yet has been shown to be as important as the oil supply disruption for understanding the spike in the real price of oil in 1990/91 (see Kilian and Murphy 2014). Problems such as these are pervasive in textual analysis. Likewise, it has been common in the media to attribute oil price increases (as well as decreases) to actions taken or not taken by OPEC, even when the basis of these attributions is not clear. Similarly, the media attention given to the peak-oil hypothesis in discussing rising oil prices has never matched the empirical support for this hypothesis. Yet another example is the extensive public debate about the role of speculative demand in oil markets in the 2000s, triggered by the so-called Masters Hypothesis, which at no point was supported by hard evidence (see Fattouh, Kilian, Mahadeva 2013).[38]

Caldara and Iacoviello's (2018) Geopolitical Risk Measure

In related work, Caldara and Iacoviello (2018) propose a monthly index of geopolitical risk that starts in 1985. Geopolitical risk is defined as the intensity of new coverage associated with wars, terrorist acts and tensions between states that affect the normal and peaceful course of international relations. In contrast, events such the Brexit or the trade wars launched by the Trump administration are not considered geopolitical risks. Caldara and Iacoviello's unsubstantiated premise is that the fraction of articles generated by their text search captures the extent to which readers are concerned with geopolitical risk.[39]

Spikes in the index occur, for example, during the Russian-Ukrainian military conflict, the Paris terrorist attacks, the 9/11 terrorist attack, the 1990 invasion of Kuwait, and the 2003 Iraq War.[40] Caldara and Iacoviello estimate the causal effect of changes in geopolitical risk by

[38] Such false narratives are also common in other contexts. For example, after the financial crisis the hypothesis of the labor market skills mismatch received media attention far out of proportion, even when empirical evidence suggested that this hypothesis was at odds with the data.

[39] This index is not based on any formal risk measure, as discussed in Machina and Rothschild (1987), nor is it necessarily global. It is simply an index based on a count of articles containing certain key words in eleven English-language newspapers published in the United States and the United Kingdom, divided by the total number of articles over the same time period.

[40] Not all of the variation in the index is intuitively plausible. For example, there is only a minor spike in the index in mid-1990 following the invasion of Kuwait, when the oil price rose sharply, but the index more than doubles in early 1991, when the real price of oil declined sharply because the threat to Saudi oil supplies had been largely

ordering the index first in a recursive VAR model including a range of U.S. and global macroeconomic and financial variables. The VAR estimates show that positive shocks to this geopolitical risk measure persistently lower the real price of oil. Since this shock also lowers global industrial production and a range of other indicators of the level of real activity, it is clearly not a storage demand shock, but appears to act like a flow demand shock.

This finding is not surprising. Although increases in geopolitical risk are highly relevant for oil markets and help explain sharp increases in the real price of oil, Caldara and Iacoviello's index is a poor measure for these oil market risks because comparatively few geopolitical events covered by their index directly affect the security of oil supplies. Events such as the U.S. invasions of Panama or Afghanistan, the coup attempt in Turkey, the Paris terrorist attacks, the death of Osama Bin Laden, tensions between the United States and North Korea, or the annexation of Crimea have little or no effect on expectations about global oil production.

The interpretation of positive shocks to the geopolitical index as negative flow demand shocks is not entirely accurate, however. Caldara and Iacoviello's explanation is that adverse geopolitical events raise economic policy uncertainty, increase fiscal deficits and future taxes, and weigh on consumer sentiment, reducing consumer spending. It also affects international capital flows. While this explanation may have some merit, it is incomplete. By construction, Caldara and Iacoviello's shock provides a biased estimate of the effect on the real price of oil, because some geopolitical events that shift the flow demand for oil (thus lowering economic activity and the real price of oil) also raise storage demand and/or lower the supply of oil (thus raising the real price of oil and further lowering economic activity). A case in point is the invasion of Kuwait in 1990. Thus, Caldara and Iacoviello's shock measure conflates several

---

removed. Nor does the index detect the peak of the oil tanker war in the Persian Gulf in 1988, raising questions about the methodology.

structural shocks, which makes their index is of limited relevance for modeling the global market for crude oil.

7. Conclusion

As this survey has illustrated, there is a plethora of approaches to identifying oil demand and oil supply shocks and of alternative oil market VAR models, reflecting differences in the data, the reduced-form specification, the identifying assumptions, and the estimation method. It may be tempting to summarize the consensus in the literature as the “average” of the results reported in published research. Alternatively, one may wish to argue for focusing on the most recent results in the literature. Neither of these approaches is compelling upon reflection. In oil market modeling, as in econometric modeling more generally, not all approaches are created equal. Results must be weighted based on the care that went into constructing the estimates and based on their robustness to more sophisticated modeling approaches.

What is remarkable about the oil market VAR literature to date is how robust its central findings have been across a broad range of model specifications. There are a few insights in the global oil market literature that stand out because they have been reaffirmed time and again. First, although the economics profession has been preoccupied for three decades with measuring the impact of oil supply shocks on the real price of oil for three decades, there is robust evidence that the effect of oil demand shocks on the real price of oil is quantitatively much more important than that of oil supply shocks. This is true even accounting for the role of the U.S. shale oil revolution in recent years. Second, there is no credible evidence at this point that the oil price fluctuations of the 1970s and early 1980s were primarily caused by exogenous supply disruptions in the Middle East, calling into question the treatment of these episodes in many macroeconomics textbooks. Third, it is not only shocks to the flow demand for oil that matter, but also shocks to the storage demand for oil driven by shifts in oil price expectations. Such expectations shocks were largely ignored by the literature until a few years ago. Explicitly accounting for oil price expectations helps understand how political events have shaped the oil price and provides a fresh perspective on historical events in the oil market.

This does not mean that the current models have resolved these questions once and for all. There continues to be much interest in exploring new identification schemes, new econometric approaches and new identifying information for oil market models. For example, Plagborg-Møller (2019), Lanne and Luoto (2019) and Bruns and Piffer (2019) recently proposed innovative approaches to achieving identification in structural VAR models that are relevant for oil market studies. Time will tell which of these new ideas will stand the test of time and which will remain interesting thought experiments.

References

Aastveit, K.A, Bjørnland, H.C., and L.A. Thorsrud (2015), "What Drives Oil Prices? Emerging vs. Developed Economies," *Journal of Applied Econometrics*, 30, 1013-1028.

Alquist, R., and L. Kilian (2010), "What Do We Learn from the Price of Crude Oil Futures?," *Journal of Applied Econometrics*, 25, 539-573.

Alquist, R., Bhattarai, S., and O. Coibion (2019), "Commodity-Price Comovement and Global Economic Activity," forthcoming: *Journal of Monetary Economics.*

Alquist, R., Kilian, L., and R.J. Vigfusson (2013), "Forecasting the Price of Oil," in: G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, 2, Amsterdam: North-Holland, 427-507.

Anderson, S.T., Kellogg, R., and S.W. Salant (2018), "Hotelling Under Pressure," *Journal of Political Economy*, 126, 984-1026.

Antolin-Diaz, J., and J.F. Rubio-Ramirez (2018), "Narrative Sign Restrictions for SVARs," *American Economic Review*, 108, 2802-2839.

Arezki, R., Ramey, V., and L. Sheng (2017), "News Shocks in Open Economies: Evidence from Giant Oil Discoveries," *Quarterly Journal of Economics*, 132, 103-155.

Arias, J., Rubio-Ramirez, J.F., and D.F. Waggoner (2015), "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications," manuscript, Federal Reserve Board.

Arias, J., Rubio-Ramirez, J.F., and D.F. Waggoner (2018), "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications," *Econometrica*, 86, 685-720.

Barsky, R.B., and L. Kilian (2002), "Do We Really Know that Oil Caused the Great Stagflation? A Monetary Alternative," in B. Bernanke and K. Rogoff (eds.), *NBER Macroeconomics Annual 2001*, 137-183.

Barsky, R.B., and L. Kilian (2004), "Oil and the Macroeconomy since the 1970s," *Journal of Economic Perspectives*, 18, Fall 2004, 115-134.

Bastianin A., and M. Manera (2018), "How Does Stock Market Volatility React to Oil Shocks?," *Macroeconomic Dynamics* 22, 666-682.

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

Baumeister, C., and J.D. Hamilton (2019b), "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Setting the Record Straight," manuscript, UC San Diego.

Baumeister, C., and G. Peersman (2013a), “The Role of Time-Varying Price Elasticities in Accounting for Volatility Changes in the Crude Oil Market,” *Journal of Applied Econometrics*, 28, 1087-1109.

Baumeister, C., and G. Peersman (2013b), “Time-Varying Effects of Oil Supply Shocks on the U.S. Economy,” *American Economic Journal: Macroeconomics*, 5, 1-28.

Blanchard, O., and D. Quah (1989), “The Dynamic Effects of Aggregate Demand and Supply Disturbances,” *American Economic Review*, 79, 655-673.

Bjørnland, H.C., Nordvik, F.M., and M. Rohrer (2019), “Supply Flexibility in the Shale Patch: Evidence from North Dakota,” manuscript, Norwegian Business School.

Bjørnland, H.C., and J. Zhulanova (2019), “The Shale Oil Boom and the US Economy: Spillovers and Time-Varying Effects,” manuscript, BI Business School.

Bruns, M., and M. Piffer (2019), “Bayesian Structural VAR Models: A New Approach for Prior Beliefs on Impulse Responses,” manuscript, Queen Mary, University of London.

Bützer, S., Habib, M.M., and L. Stracca (2016), “Global Exchange Rate Configurations: Do Oil Shocks Matter?,“ *IMF Economic Review*, 64, 443-470.

Caldara, D., Cavallo, M., and M. Iacoviello (2019), “Oil Price Elasticities and Oil Price Fluctuations,” *Journal of Monetary Economics*, 103, 1-20.

Caldara, D., and M. Iacoviello (2018), “Measuring Geopolitical Risk,” manuscript, Federal Reserve Board.

Cavallo, M. and T. Wu (2012), “Measuring Oil-Price Shocks Using Market-Based Information,” International Monetary Fund Working Paper No. 12/19.

Chen, S.-S. (2009), “Revisiting the Inflationary Effects of Oil Prices,” *Energy Journal*, 30, 141-154.

Chudik, A., and G. Georgiadis (2019), "Estimation of Impulse Response Functions When Shocks are Observed at a Higher Frequency than Outcome Variables," manuscript, Federal Reserve Bank of Dallas.

Datta, D.D., and D.A. Dias (2019), "Oil Shocks: A Textual Analysis Approach," manuscript, Federal Reserve Board.

Cross, J. (2019), "The Role of Uncertainty in the Market for Crude Oil," manuscript, Norwegian School of Business.

Delle Chiaie, S., Ferrara, L., and D. Giannone (2017), "Common Factors in Commodity Prices," manuscript. Federal Reserve Bank of New York.

Fattouh, B., Kilian, L., and L. Mahadeva (2013), "The Role of Speculation in Oil Markets: What Have We Learned So Far?" *Energy Journal*, 34, 7-33.

Fry, R., and A.R. Pagan (2011), "Sign Restrictions in Structural Vector Autoregressions: A Critical Review," *Journal of Economic Literature*, 49, 938-960.

Ghysels, E. (2016), "Macroeconomics and the Reality of Mixed Frequency Data," *Journal of Econometrics*, 193, 294-314.

Hamilton, J.D. (2003), "What Is an Oil Shock?," *Journal of Econometrics*, 113, 363-398.

Hamilton, J.D. (2019), "Measuring Global Economic Activity," forthcoming: *Journal of Applied Econometrics*.

Hamilton, J.D., and A.M. Herrera (2004), "Oil Shocks and Aggregate Macroeconomic Behavior: The Role of Monetary Policy: A Comment," *Journal of Money, Credit and Banking*, 36, 265-286.

Hausman, J.A., and W.K. Newey (1995), "Nonparametric-Estimation of Exact Consumers' Surplus and Deadweight Loss," *Econometrica*, 6, 1445-1476.

Herrera, A.M., and S.K. Rangaraju (2019), "The Effects of Oil Supply Shocks on U.S. Economic Activity: What Have We Learned?," forthcoming: *Journal of Applied Econometrics*.

Herwartz, H., and M. Plödt (2016), "The Macroeconomic Effects of Oil Price Shocks: Evidence from a Statistical Identification Approach," *Journal of International Money and Finance*, 61, 30-44.

Inoue, A., and L. Kilian (2013), "Inference on Impulse Response Functions in Structural VAR Models," *Journal of Econometrics*, 177, 1-13.

Inoue, A., and L. Kilian (2019), "Corrigendum to 'Inference on impulse response functions in structural VAR models' [J. Econometrics 177 (2013), 1-13]," *Journal of Econometrics*, 209, 139-143.

Inoue, A., and L. Kilian (2020a), "Joint Bayesian Inference about Impulse Responses in VAR Models," manuscript, Federal Reserve Bank of Dallas.

Inoue, A., and L. Kilian (2020b), "The Uniform Validity of Impulse Response Inference in Autoregressions," forthcoming: *Journal of Econometrics*.

Jacks, D., and M. Stürmer, M. (2019), "What Drives Commodity Price Booms and Busts?," forthcoming: *Energy Economics*.

Juvenal, L., and I. Petrella (2015), "Speculation in the Oil Market," *Journal of Applied Econometrics*, 30, 621-649.

Känzig, D. (2019), "The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements," manuscript, London Business School.

Kilian, L. (2006), "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," CEPR Discussion Paper No. 5994.

Kilian, L. (2008a), "The Economic Effects of Energy Price Shocks," *Journal of Economic*

Literature, 46, 871-909.

Kilian, L. (2008b), "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *Review of Economics and Statistics*, 90(2), 216-240.

Kilian, L. (2009), "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market", *American Economic Review*, 99, 1053-1069.

Kilian, L. (2010), "Explaining Fluctuations in U.S. Gasoline Prices: A Joint Model of the Global Crude Oil Market and the U.S. Retail Gasoline Market," *Energy Journal*, 31, 87-104.

Kilian, L. (2016), "The Impact of the Shale Oil Revolution on U.S. Oil and Gas Prices," *Review of Environmental Economics and Policy*, 10, 185-205.

Kilian, L. (2017), "The Impact of the Fracking Boom on Arab Oil Producers," *Energy Journal*, 38, 137-160.

Kilian, L. (2019a), "Facts and Fiction in Oil Market Modeling," manuscript, Federal Reserve Bank of Dallas.

Kilian, L. (2019b), "Measuring Global Real Economic Activity: Do Recent Critiques Hold Up to Scrutiny?," *Economics Letters*, 178, 106-110.

Kilian, L., and B. Hicks (2013), "Did Unexpectedly Strong Economic Growth Cause the Oil Price Shock of 2003-2008?," *Journal of Forecasting*, 32, 385-394.

Kilian, L., and T.K. Lee (2014), "Quantifying the Speculative Component in the Real Price of Oil: The Role of Global Oil Inventories," *Journal of International Money and Finance*, 42, 71-87.

Kilian, L., and H. Lütkepohl (2017), *Structural Vector Autoregressive Analysis*, Cambridge University Press.

Kilian, L., and D.P. Murphy (2012), "Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models," *Journal of the European Economic Association*, 10, 1166-1188.

Kilian, L., and D.P. Murphy (2014), "The Role of Inventories and Speculative Trading in the Global Market for Crude Oil," *Journal of Applied Econometrics*, 29, 454-478.

Kilian, L., and C. Park (2009), "The Impact of Oil Price Shocks on the U.S. Stock Market," *International Economic Review*, 50, 1267-1287.

Kilian, L., and C. Vega (2011), "Do Energy Prices Respond to U.S. Macroeconomic News? A Test of the Hypothesis of Predetermined Energy Prices," *Review of Economics and Statistics*, 93, 660-671.

Kilian, L., and X, Zhou (2018), "Modeling Fluctuations in the Global Demand for Commodities," *Journal of International Money and Finance*, 88, 54-78.

Kilian, L., and X. Zhou (2019a), "Does Drawing Down the U.S. Strategic Petroleum Reserve Help Stabilize Oil Prices?" manuscript, Federal Reserve Bank of Dallas.

Kilian, L., and X. Zhou (2019b), "Oil Prices, Exchange Rates and Interest Rates," manuscript, Federal Reserve Bank of Dallas.

Kilian, L., and X. Zhou (2019c), "Oil Supply Shock Redux?" manuscript, Federal Reserve Bank of Dallas.

Kilian, L., and T. Zha (2002), "Quantifying the Uncertainty about the Half-Life of Deviations from PPP," *Journal of Applied Econometrics*, 17, 107-125.

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

Lanne, M., and J. Luoto (2019), "Identification of Economic Shocks by Inequality Constraints in

Bayesian Structural Vector Autoregression," forthcoming: *Oxford Bulletin of Economics and Statistics.*

Leeb and Pötscher (2005), "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21, 21-59.

Lippi, F., and A. Nobili (2012), "Oil and the Macroeconomy: A Quantitative Structural Analysis," *Journal of the European Economic Association*, 10, 1059-1083.

Lütkepohl, H., and Netšunajev (2014), "Disentangling Demand and Supply Shocks in the Crude Oil Market: How to Check Sign Restrictions in Structural VARs," *Journal of Applied Econometrics*, 29, 479-496.

Machina, M.J., and M. Rothschild (1987), "Risk," in *The New Palgrave Dictionary of Economics*, Eatwell, J., Millgate, M., and Newman (eds.), London, UK: MacMillan, pp. 203–205.

Montiel Olea, J.L, Stock, J.H., and M.W. Watson (2018), "Inference in Structural Vector Autoregressions Identified with an External Instrument," manuscript, Columbia University.

Mork, K.A. (1989), "Oil and the Macroeconomy When Prices Go Up and Down: An Extension of Hamilton's Results," *Journal of Political Economy*, 97, 740-744.

Newell, R., and B. Prest (2019), "The Unconventional Oil Supply Boom: Aggregate Price Response from Microdata," *Energy Journal*, 40, 1-30.

Peersman, G., and I. van Robays (2009), "Oil and the Euro Area Economy," *Economic Policy*, 24, 603-651.

Plagborg-Møller, M. (2019), "Bayesian Inference on Structural Impulse Response Functions," *Quantitative Economics*, 10, 145-184.

Plagborg-Møller, M., and C.K. Wolf (2019), "Local Projections and VARs Estimate the Same Impulse Responses," manuscript, Princeton University.

Primiceri, G. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821-852.

Rubio-Ramirez, J.F., Waggoner, D., and T. Zha (2010), "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, 77, 665-696.

Sims, C.A., and T. Zha (1999), "Error Bands for Impulse Responses," *Econometrica*, 67, 1113-1156.

Sockin, M., and W. Xiong (2015), "Informational Frictions and Commodity Markets," *Journal of Finance*, 70, 2063-2097.

Stürmer, M. (2018), "150 years of Boom and Bust - What Drives Mineral Commodity Prices," *Macroeconomic Dynamics*, 22, 702-717.

Ye, S., and B. Karali (2016), "The Informational Content of Inventory Announcements: Intraday Evidence from the Crude Oil Futures Market," *Energy Economics*, 59, 349-364.

Zhou, X. (2019), "Refining the Workhorse Oil Market Model," forthcoming: *Journal of Applied Econometrics*.