CESifo WORKING PAPERS

7902
2019

October 2019

Facts and Fiction in Oil Market Modeling

Lutz Kilian

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
www.cesifo-group.org/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: www.CESifo-group.org/wp

CESifo *Working Paper No. 7902*
Category 10: Energy and Climate Economics

Facts and Fiction in Oil Market Modeling

Abstract

Baumeister and Hamilton (2019a) assert that every critique of their work on oil markets by Kilian and Zhou (2019a) is without merit. In addition, they make the case that key aspects of the economic and econometric analysis in the widely used oil market model of Kilian and Murphy (2014) and its precursors are incorrect. Their critiques are also directed at other researchers who have worked in this area and, more generally, extend to research using structural VAR models outside of energy economics. The purpose of this paper is to help the reader understand what the real issues are in this debate. The focus is not only on correcting important misunderstandings in the recent literature, but on the substantive and methodological insights generated by this exchange, which are of broader interest to applied researchers.

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

Keywords: oil supply elasticity, oil demand elasticity, IV estimation, structural VAR, Bayesian inference, oil price, global real activity.

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

October 8, 2019
The views expressed in this paper are my 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. I thank Brian Prest, Ana Maria Herrera, Atsushi Inoue, Martin Stürmer, Mine Yucel, and Xiaoqing Zhou for helpful discussions.

1. Introduction

In a recent note, Baumeister and Hamilton (2019a, henceforth BH) assert that every critique of the oil market model in Baumeister and Hamilton (2019b) by Kilian and Zhou (2019a, henceforth KZ) is without merit. In addition, they make the case that key aspects of the economic and econometric analysis in the widely used oil market model of Kilian and Murphy (2014) and its precursors are incorrect. Their critiques are also directed at other researchers who have worked in this area and, more generally, extend to research using structural VAR models outside of energy economics. The purpose of this paper is to help the reader understand what the real issues are in this debate. The focus is not only on correcting important misunderstandings in the recent literature, but on the substantive and methodological insights generated by this exchange, which are of broader interest to applied researchers.

The question of how to model oil markets may seem esoteric to many economists at first, but has important implications for how oil-importing and oil-exporting economies respond to global oil price fluctuations and for how policymakers should respond to these oil price fluctuations. Baumeister and Hamilton (2019b) concluded that oil supply shocks are more important drivers of the real price of oil and that they are much more recessionary for the U.S. economy than suggested by earlier oil market studies including Kilian (2008, 2009) and Kilian and Murphy (2014). Their conclusion, however, is highly sensitive to a priori assumptions about what constitutes a reasonable value for the one-month price elasticity of oil supply.

In this paper, I provide new evidence that BH's agnostic prior specification for that elasticity is inconsistent with extraneous elasticity estimates even after accounting for estimation uncertainty. This conclusion holds even after accounting for the recent rise in U.S. shale oil production. My discussion draws on recent supply elasticity estimates for U.S. oil producers in

Newell and Prest (2019), which is the most comprehensive microeconomic study of oil supply elasticities to date. I further explain why the alternative supply elasticity estimates in Bjørnland, Nordvik and Rohrer (2019) and Caldara, Cavallo and Iacoviello (2019) that BH appeal to are overstating the oil supply elasticity. When using economically plausible priors motivated by microeconomic evidence, BH's approach generates results quite similar to those based on the model proposed by Kilian and Murphy (2014), highlighting the importance of utilizing extraneous identifying information in specifying structural VAR models. My analysis illustrates that prior beliefs on structural parameters must be selected carefully in applied work.

It is also worth emphasizing that there is no evidence that explicitly modeling uncertainty about identifying restrictions as proposed by BH makes any difference in practice, compared to mainstream approaches to Bayesian inference for structural VAR models (see Herrera and Rangaraju 2019).[1] This evidence (and related evidence in the literature) confirms that the substantive results of Kilian and Murphy (2014) are robust to changes in the estimation period, data choice, loss function, model specification, and econometric approach, provided the prior on the oil supply elasticity is economically plausible. It also reaffirms the conclusion that oil demand shocks are the dominant driver of the real price of oil and that the recessionary effect of oil supply shocks is modest (e.g., Baumeister and Kilian 2016a).

In addition, I explain why BH’s assertion that earlier oil market models with low oil supply elasticities imply implausibly large demand elasticities, which has been interpreted by some applied researchers as an indication that these models are misspecified, is not correct. This view is not only contradicted by the estimates in Kilian and Murphy (2014) who report a lower oil demand elasticity than Baumeister and Hamilton (2019b), while maintaining an oil supply

[1] A similar conclusion was reached by Lanne and Luoto (2019a), who examined the application of the same methodology to labor markets in Bauneister and Hamilton (2015).

elasticity close to zero. It can also be shown that BH's argument, which has been reiterated by a number of recent studies, is flawed because they apply an incorrect definition of the oil demand elasticity to earlier studies. Nor is there support for BH's suggestion that earlier structural VAR studies misunderstood the implications of elasticity priors for inference or misinterpreted the elasticity concept. The latter point is relevant more generally for other structural VAR applications involving bounds on elasticities. Finally, I address BH's concerns about the validity of the Kilian index of global real economic activity used in many oil market studies including Kilian and Murphy (2014), to the extent that these concerns have not already been addressed in Kilian (2019).

The remainder of this paper is organized as follows. Section 2 reviews how Kilian and Murphy (2014) originally evaluated their oil market model, how subsequently Bayesian methods of inference developed by Inoue and Kilian (2013, 2019) have been applied to this model, and how the latter approach differs from the Bayesian approach favored by BH. Section 3 examines BH's concern that the impulse response estimates in Kilian and Murphy (2014) are not robust and addresses their claim that this study did not employ narrative sign restrictions. Section 4 reviews the debate over the magnitude of the one-month price elasticity of oil supply. It explains why KZ's assessment of this evidence differs sharply from BH's and discusses how KZ accounted for the uncertainty in microeconomic elasticity estimates. Section 5 reviews the role of elasticity priors in oil market modeling. In particular, it addresses the questions of how to define the one-month oil demand elasticity, whether this elasticity can be recovered from the oil market models of Kilian (2009) and Kilian and Murphy (2012), why the construction of the elasticities in the Kilian and Murphy (2014) model makes sense, and why seemingly agnostic priors such as BH's prior for the oil supply elasticity are actually highly influential. Section 6 addresses

Hamilton's (2019) renewed critique of the merits of the Kilian (2009) index of global real economic activity. Section 7 discusses how BH's contribution fits into the literature. The concluding remarks are in section 8.

2. Understanding the Empirical Evidence on the Kilian and Murphy (2014) Model

The structural VAR models of Kilian and Murphy (2012, 2014) are without doubt among the most widely used and studied oil market models in the literature.[2] It is useful to lay out the broader context of Bayesian inference for these models.

Estimation and Inference in Popular Oil Market Models

Whereas the oil market model of Kilian (2008, 2009) was estimated and evaluated using frequentist econometric methods, oil market models based on sign restrictions have been routinely evaluated using the Bayesian approach to estimation and inference. Thus, BH are by no means the first to have relied on Bayesian methods for evaluating oil market models. Rather they provide an alternative to existing Bayesian approaches already used in the literature.

An important challenge in Bayesian inference is how to summarize the posterior density of the impulse response estimates. When Kilian and Murphy (2014) wrote their paper, they were well aware of the conceptual failings of commonly used summary statistics for posterior draws from sign-identified models such as the so-called posterior median response function recommended by BH. The problems with inference based on quantiles of individual impulse responses are well documented in the literature (e.g., Sims and Zha 1999; Fry and Pagan 2011; Kilian and Murphy 2012; Inoue and Kilian 2013). For a review, I refer the reader to the textbook of Kilian and Lütkepohl (2017).

[2] Recent examples include Inoue and Kilian (2013), Baumeister and Kilian (2014a), Bützer, Habib, and Stracca (2016), Kilian (2017), Bruns and Piffer (2018), Antolin-Diaz and Rubio-Ramirez (2018), Herrera and Rangaraju (2019), Zhou (2019), and Lanne and Luoto (2019b).

Since there was no coherent econometric approach to evaluating the posterior draws at the time, Kilian and Murphy (2014) focused on the set of admissible draws evaluated at the mean of the posterior draws for the reduced-form parameters. Under their diffuse prior for the slope parameters, this amounts to conditioning on the MLE. Likewise, Kilian and Murphy (2012) conditioned on the MLE in illustrating the problem of identification in sign-identified models. The focus in Kilian and Murphy (2012) was on how alternative assumptions affect the set of admissible models and therefore statistical summary measures including the median response function. In contrast, Kilian and Murphy (2014) made the case that among the model draws selected based on the sign restrictions, one model can be externally validated. Their empirical analysis focused on the latter model.

Subsequently, Inoue and Kilian (2013, 2019) developed an economically sensible and statistically coherent Bayesian alternative to the use of median response functions that yields a unique most likely admissible model as well as joint credible sets for the impulse response functions of that model.[3] Inoue and Kilian (2013) and Herwartz and Plödt (2016) applied this new approach to the Kilian and Murphy (2012) model and confirmed the substance of the earlier conclusions. Herrera and Rangaraju (2019) and Zhou (2019) applied the same method to the Kilian and Murphy (2014) model and concluded that even accounting for estimation uncertainty, the substantive conclusions of Kilian and Murphy can be replicated.[4] When confronted with this evidence, BH's response is to question the merits of Inoue and Kilian's approach to Bayesian inference. Their arguments are not persuasive, as shown next.

[3] Inoue and Kilian (2013) appeared in print before Kilian and Murphy (2014), but was written after the Kilian and Murphy study.

[4] BH imply that there must be something wrong with the Kilian and Murphy models because of the small number of admissible models. It should be noted that comparing the number of admissible models in Kilian and Murphy (2012, 2014) conditional on the MLE with that based on the posterior in the BH model is misleading. BH should have compared their results with the results in Herrera and Rangaraju (2019), for example, who evaluate the posterior of the Kilian and Murphy (2014) model.

Why the Approach to Bayesian Inference in Inoue and Kilian (2013) Makes Sense

There are two differences between the approach of Inoue and Kilian and that of BH. One is the statistic of interest. The other is the loss function applied to that statistic. It is useful to elaborate on these differences.

As to the statistic of interest, Inoue and Kilian focus on inference on the set of admissible structural models, where each model is represented by the joint set of all structural impulse responses in this model up to a certain horizon. Inference is based on the posterior density of the admissible structural models. In contrast, BH focus on inference on individual impulse responses and construct response functions and error bands by stringing together quantiles of the marginal posterior distribution of impulse responses, ignoring the dependence of these impulse responses over time and across response functions. Their approach is akin to arguing that in studying a system of regression equations it is sufficient to base inference on t-tests for each coefficient in a given equation, ignoring that the t-statistics are dependent within and across equations.

As to the loss function, Inoue and Kilian evaluate the VAR model at the mode of the density of the admissible structural models, which allows them to construct a unique and economically interpretable most likely model. They also construct joint highest posterior density credible sets based on the 68% most likely admissible models. BH criticize that Inoue and Kilian did not make explicit their loss function. That seems hardly necessary, given that the mode has a long tradition in Bayesian inference (e.g., Koop 1996; Zha 1999; Waggoner and Zha 2012, Plagborg-Møller 2019). After all, few studies articulate the loss function underlying the posterior median response function.

In contrast, BH evaluate the absolute loss of each individual impulse response coefficient and sum the results. They stress that the posterior median response function is optimal under this loss function. That point is correct. Of course, Inoue and Kilian's estimate by construction is equally optimal under their loss function, so the real question is what object to apply the loss function to and how to select among alternative loss functions. What BH do not make explicit is that their loss function implies that the user has no interest in the dynamics of the impulse response functions or in the dependence across impulse response function. As documented at length in Kilian and Lütkepohl (2017), these dependencies actually are the primary interest of most users of structural VAR models, so BH's loss function (and hence the use of the median response function) is unappealing.

BH take issue with KZ's assessment that median response functions are economically and statistically meaningless. As summarized in Kilian and Lütkepohl (2017), they are economically meaningless because they conflate the dynamic responses implied by the admissible structural models and they are statistically meaningless because the vector of pointwise medians is not a proper measure of the central tendency of a vector of random variables (see Fry and Pagan 2011; Inoue and Kilian 2013). BH concede that one might prefer to evaluate the set of admissible models under quadratic loss, which calls for reporting the vector of posterior means of the impulse responses. The latter approach provides a valid measure of central tendency, but still conflates the dynamic responses of different admissible structural models. Only Inoue and Kilian's approach solves the latter problem. The same concerns also apply to vectors of historical decompositions and sequences of structural shocks. For example, the practice of treating the vector of posterior medians of the oil supply shock sequence obtained from the VAR model as the "oil supply shock series", as proposed by BH, is inherently misleading. An additional problem when working with forecast error variance decompositions is that the vector of posterior medians violates the adding-up constraint (see Kilian and Zhou

2019c). Likewise, changes in vectors of posterior medians of historical decompositions of a VAR variable need not add up to the total change in the variable of interest.

Finally, while it is correct that posterior median response functions and the posterior mean of the response functions tend to be similar in practice, this is not necessarily true for the response functions of the most likely model, as has been illustrated in the literature (see Kilian and Lütkepohl 2017). BH do not report results based on the latter methodology and hence do not know whether their results are robust.

Why BH's Bayesian Approach Is Not Superior to Existing Bayesian Approaches

More generally, BH argue that the mainstream approach to conducting Bayesian inference for sign-identified structural VAR models, as represented by Uhlig (2005) and Rubio-Ramirez, Waggoner and Zha (2010), upon which Inoue and Kilian (2013, 2019) build, or by Arias, Rubio-Ramirez, and Waggoner (2018) and Antolin-Diaz and Rubio-Ramirez (2018), is misguided. BH even make the surprising claim that it is not possible to construct posterior distributions for structural impulse responses using the traditional approach. It appears that BH are alone in the profession in this assessment, given that numerous studies, including those referenced above, have done just that.

The key concern that motivated the analysis in BH is that the traditional approach of combining priors for the reduced-form parameters with a prior on the rotation matrix may inadvertently result in an informative prior for the structural impulse responses. Such outcomes are indeed a possibility, although it has yet to be shown that this is a serious practical concern in oil market models. The fact that BH's methodology, when applied to their own oil market model generates impulse response estimates that are similar to estimates based on the traditional approach, as long as similar priors on the one-month price elasticity of oil supply are used, suggests that this concern is overrated (see Baumeister and Hamilton 2019b; Herrera and Rangaraju 2019).

More importantly, as noted by KZ, BH's alternative approach far from addressing the concern about an inadvertently informative prior on the structural impulse responses suffers from exactly the same conceptual problem as the traditional approach. The problem is that it is not possible to have a prior on the structural VAR coefficients without implicitly specifying a prior on the structural impulse responses. For example, if one constructs an explicit prior directly on the parameters of the structural model, the implied prior for the structural responses will inevitably be informative in ways that are difficult to anticipate, since the structural impulse responses are nonlinear transformations of these structural parameters. Likewise, if one combines selected restrictions on the structural parameters with selected restrictions on the structural impact multiplier matrix, as discussed by BH, this fact does not solve the problem of an inadvertently informative prior about the structural impact multiplier matrix.

The only way to address this concern would be to construct an explicit prior on the structural impulse responses without restricting the structural parameters, but that is precisely what BH's methodology cannot do. In other words, this methodology was not designed to be applied to models that rely on identifying information on the structural impact multiplier matrix, which encompasses most sign-identified VAR models in the literature. For example, it cannot be applied to the oil market models of Kilian and Murphy (2012) and Kilian and Murphy (2014), which rely exclusively on restrictions on the structural impact multiplier matrix. An additional challenge for BH's approach are the cross-equation restrictions on the structural impulse responses embodied in the latter model.

Thus, while I am glad to see that BH were able to establish the robustness of the conclusions in both Kilian (2009) and in Kilian and Murphy (2012) based on their methodology, what is strikingly absent from their paper is an application of their methodology to the state-of-the-art oil market model of Kilian and Murphy (2014). Instead, BH take issue with the analysis in Kilian and Murphy (2014).[5]

3. BH's Critique of the Kilian and Murphy (2014) Model

A central message of BH is that the original estimates of the Kilian and Murphy (2014) model are not robust to simulation error, that the Kilian and Murphy (2014) model is misspecified, and that the elasticity bounds employed by Kilian and Murphy are incorrect. It is useful to examine the first two points before discussing the broader claims made by BH about the oil market literature and the role of oil supply elasticities.

The Impulse Response Estimates in Kilian and Murphy (2014) Can Be Replicated

BH report impulse responses constructed using the original data and code of Kilian and Murphy (2014), as posted in the Journal of Applied Econometrics data and code archive, for two different random seeds. Whereas the impulse response estimates based on the original seed match exactly those reported in Kilian and Murphy (2014), those based on the alternative seed in some cases differ in magnitude from the response estimate focused on by Kilian and Murphy, although the responses have the same sign. BH leave the reader with the impression that the empirical results in Kilian and Murphy (2014) are not robust and cannot be replicated with a different seed.[6]

---

[5] Note that the "replication code" for Kilian and Murphy (2012) posted on Baumeister's homepage actually is not for the Kilian and Murphy (2012) model at all. It is for a different model specification that was designed to resemble the original model specification as closely as possible using the BH methodology.

[6] BH also express concern that there are not more admissible models. One reason is that Kilian and Murphy's original code was computationally inefficient. More efficient code that generates more admissible models for the same number of draws was available to BH, but was not used. The other reason is that Kilian and Murphy (2014), unlike subsequent studies, conditioned on the MLE. Comparing the number of admissible draws conditional on one value of the reduced-form parameters to the number of admissible draws obtained when drawing from the posterior of the reduced-form coefficients is inherently misleading.

BH’s claim is astonishing, given how many other studies have confirmed the substantive findings of Kilian and Murphy (2014) using a variety of different econometric methods, different data sets, different sample periods and even extensions of the original model (e.g., Kilian and Lee 2014; Baumeister and Kilian 2014a, 2016b; Kilian 2017; Herrera and Rangaraju 2019; Zhou 2019; Känzig 2019; Cross 2019; Kilian and Zhou 2019b,c).

One reason why BH have difficulties reproducing the original impulse response estimates is that they do not implement the full estimation procedure employed by Kilian and Murphy (2014). In particular, they fail to impose the additional narrative sign restrictions on the historical decomposition of the real price of oil that Kilian and Murphy used to ensure the external validity of their preferred model estimate. In contrast, Zhou (2019), using the same data, but imposing these narrative sign restrictions, was able to replicate the impulse responses and historical decompositions in Kilian and Murphy (2019) without difficulty. Zhou also showed that Kilian and Murphy’s key result about the relative importance of oil supply and oil demand shocks as drivers of the real price of oil is invariant to which admissible model solution one focuses on.

Moreover, closer inspection reveals that BH’s alternative estimate appears to be roughly within the range of the conventional posterior quantile error band reported in Figure 1 of Kilian and Murphy (2014). Because this error band is based on the code also used by BH that does not incorporate the additional narrative sign restrictions, it may be used to assess the variability of their impulse response estimator subject to the caveats about the construction of error bands discussed earlier. Figure 1 suggests that the magnitude of the alternative response estimates encountered by BH is within the range of what we would expect in the absence of narrative sign restrictions. Thus, BH’s alternative estimate in no way invalidates the analysis in Kilian and Murphy (2014).

Kilian and Murphy (2014) Employed Narrative Sign Restrictions

BH's response is to deny that Kilian and Murphy (2014) employed narrative sign restrictions on the historical decomposition to select the most credible model among the set of models that satisfy the sign restrictions on the impulse responses. BH insist that they did not find the expression "narrative sign restriction" in the paper or in the replication code provided by the authors and suggest that Kilian and Murphy must have changed their mind about their procedure without telling anyone. BH further claim that nothing resembling narrative sign restrictions was implemented anywhere in Kilian and Murphy (2014). These claims are misleading.

Of course, the original paper did not use the term "narrative sign restrictions", which did not exist at the time, but it discussed how the draws for the admissible models were "externally validated" by verifying that the model estimates match external evidence about what has been driving the real price of oil during selected episodes. This point was discussed both in Kilian and Murphy (2014) and in the companion paper by Kilian and Lee (2014). For example, Kilian and Lee (2014), in reviewing the support for Kilian and Murphy's (2014) preferred model, note that:

> "… one can externally validate the fit of the model. There are several episodes for which we have extraneous evidence from industry specialists such as Terzian (1985) or Yergin (1992) that speculation took place in physical oil markets. A natural joint test of the structural model and of the inventory data is to compare its historical decomposition against this external evidence. The model passes this test. For example, it detects surges in speculative demand in 1979 following the Iranian Revolution, in 1990 around the time of the invasion of Kuwait, and in late 2002 in anticipation of the Iraq War, as well as large declines in speculative demand in 1986 after the collapse of OPEC and in late 1990 when the U.S. had moved enough troops to Saudi Arabia to forestall an invasion by Iraq (see Kilian, 2008[a]," (p. 74)

Similar statements can be found in Kilian and Murphy (2014, p. 460, 469). Given that BH's candidate solution based on their alternative seed has not been externally validated, Kilian and Murphy (2014) would not have considered it a legitimate estimate.

BH are correct that the external validation procedure discussed in these papers was not contained in the code we provided. Since the number of models satisfying the sign restrictions in Kilian and Murphy (2014) is small, this procedure was originally implemented manually by inspecting the historical decompositions for the real price of oil. This approach is obviously infeasible when considering a much larger number of model draws, but Zhou (2019) shows how one can incorporate this external validation procedure into the code. Zhou demonstrates that the original findings in Kilian and Murphy (2014) can be replicated, whether on the original data or on extended data. Zhou (2019) describes how to operationalize this procedure:

> "Motivated by the reasoning in Kilian and Murphy (2014, p. 460, 469) and Kilian and Lee (2014, p. 74), I postulate (1) that storage demand shocks cumulatively raised the log real price of oil by at least 0.2 (or approximately 20%) between May and December 1979, consistent with anecdotal evidence of a dramatic surge in inventory building in the oil market during that time, (2) that storage demand cumulatively lowered the log real price of oil by at least 0.15 between December 1985 and December 1986, after OPEC collapsed, and (3) that storage demand shocks raised the log real price of oil by at least cumulatively between June 1990 and October 1990, reflecting market expectations that Iraq would invade its neighbors. Flow supply shocks are assumed to have raised the log real price of oil cumulatively by at least 0.1 between July and October of 1990, reflecting the invasion of Kuwait and the cessation of Iraqi and Kuwaiti oil production in early August. Finally, the cumulative effect of flow demand shocks on the log real price of oil between June and October of 1990 is bounded by 0.1, given that the oil price spike of 1990 was not associated with the global business cycle."

Zhou (2019) also clearly explains how to impose these inequality restrictions, stressing that the external validation procedure in Kilian and Murphy (2014) was an early example of narrative sign restrictions, as recently proposed by Antolin-Diaz and Rubio-Ramirez (2018). Antolin-Diaz and Rubio-Ramirez also explicitly note that "narrative information in the context of the oil market was used by Kilian and Murphy (2014) to confirm the validity of their proposed identification" (p. 2803) and that Kilian and Murphy (2014) "impose[d] sign restrictions on the historical decompositions" (p. 2807). The same point is discussed in Kilian and Lütkepohl (2017, section 13.6.5). While the external validation procedure was perhaps not as clearly explained in the original paper as it should have been, owing in part to space constraints imposed by the journal, BH can hardly claim to have had no knowledge of the link between external validation and narrative sign restrictions.[7] The facts are that the impulse response estimates in Kilian and Murphy (2014) can be replicated and that Kilian and Murphy (2014) used narrative sign restrictions.

4. The Oil Supply Elasticity Debate

The magnitude of the oil supply elasticity is a key determinant of the relative importance of oil supply and oil demand shocks for the real price of oil (e.g., Kilian and Murphy 2012). KZ appeal to economic theory and microeconomic evidence to bound the value of this elasticity. BH assert that KZ misrepresented the empirical evidence on microeconomic estimates of the oil supply elasticity. They argue that the magnitude of oil supply elasticity estimates in the literature and the uncertainty surrounding these estimates justifies their choice of a diffuse oil supply elasticity prior allowing for elasticity values in the range $[0,\infty]$. It is useful to review the evidence on the value of the one-month oil supply elasticity with special attention to the limitations of the studies by Bjørnland, Nordvik and Rohrer (2019) and Caldara, Cavallo and Iacoviello (2019) that BH base their arguments on. As the baseline, I start by reviewing the microeconomic evidence provided by Newell and Prest (2019).

The Evidence in Newell and Prest (2019)

Microeconomic estimates of the oil supply elasticity are informative for the identification of oil market models because they constitute extraneous evidence. BH suggest that KZ do not understand the need to account for the uncertainty in microeconomic estimates of the one-month price elasticity of oil supply and incorrectly treat point estimates as upper bounds. They clearly

[7] Interestingly, even without these narrative restrictions, BH could have replicated the Kilian and Murphy (2014) results, if they had used state-of-the-art econometric methods for evaluating the posterior model draws rather than conditioning on the MLE, as demonstrated in Herrera and Rangaraju (2019).

misunderstood our reasoning. It is useful to illustrate this point based on the U.S. oil supply elasticity estimates reported in Newell and Prest (2019).

Whereas earlier studies focused on U.S. oil producer data from selected regions such as Texas or North Dakota, Newell and Prest (2019) include data from all major oil producing regions in the United States, making it the most comprehensive study to date. Newell and Prests's preferred estimate of the one-quarter oil supply elasticity for conventional crude is 0.017 (with a standard error of 0.006). Their estimate is close to the benchmark provided by the theoretical analysis in Anderson, Kellogg and Salant (2018) who showed within an equilibrium model that the short-run oil supply elasticity is zero if adjusting oil production is costly, as tends to be the case in practice. It is important to keep in mind that the quarterly oil supply elasticity estimate in Newell and Prest (2019) is an upper bound on the one-month price elasticity of oil supply that we are ultimately interested in. The upper bound of 0.04 considered in Zhou (2019) and other recent studies for the one-month supply elasticity is four standard errors above this quarterly point estimate. BH's posterior estimate of 0.15 is an astronomical 22 standard errors above this point estimate.

One might think that this conclusion would be changed when incorporating the one-month supply elasticity of shale oil producers. A common view is that the latter elasticity is at least as large as that for conventional crude oil. There are two points that must be kept in mind, however. One point is that shale oil did not exist for much of the estimation period considered in oil market studies. Shale oil production took off only in late 2008 and from a very low level. Thus, supply elasticity estimates for conventional crude are much more relevant for oil market modeling than elasticity estimates for shale oil. The other point is that shale oil accounted only for a small share of world oil production even in the years after 2008, so we need to consider appropriately weighted averages in inferring the implied global oil supply elasticity.

Newell and Prest (2019) report an estimate of -0.022 (with a standard error of 0.013) for U.S. shale oil producers. This estimate is close to zero and not statistically significant. This result does not contradict the widely held view that shale oil producers are nimbler in responding to market conditions than conventional producers. It simply means that this response takes more than one month even for shale oil producers. If we take the U.S. estimates as representative for oil producers in the world, given a share of 8% for shale oil production in global oil production at the end of the estimation sample, this implies a global one-month oil supply elasticity of under 0.016.[8] We do not know the covariance of the elasticity estimates for conventional oil and for shale oil, but it seems unlikely that the upper bound of a confidence interval for this quarterly global oil supply elasticity estimate would exceed the bound of 0.04 proposed by Zhou (2019).

The global oil supply elasticity estimate of 0.016 implied by the results in Newell and Prest (2019) is difficult to reconcile with the posterior median estimate of 0.15 for the one-month oil supply elasticity reported in BH. One would be hard pressed to ascribe the difference to estimation uncertainty in the micro estimates. Moreover, BH's supply elasticity priors (both the baseline prior and the alternative prior they discuss) allow for oil supply elasticity values approaching infinity. Such priors are not credible economically, if we take the micro evidence seriously.

BH do not mention the evidence in Newell and Prest (2019) for all U.S. oil producers. They instead reference a study by Bjørnland et al. (2019) that focuses on a smaller data set including oil producers in North Dakota only. I will first present my interpretation of their

[8] Specifically, $0.92 \times 0.017 + 0.08 \times 0 = 0.0156$, where we treat the shale oil supply elasticity as zero rather than using the point estimate of -0.022.

estimates and then consider BH's view.[9]

The Evidence in Bjørnland et al. (2019)

Bjørnland et al.'s estimate of the one-month price elasticity of oil supply for conventional oil producers, obtained by regressing the change in conventional oil production in North Dakota on the change in the real price of oil, is 0.03 (with a standard error of 0.05).[10] The corresponding estimate of the one-month price elasticity of oil supply for shale oil is -0.12 (with no standard error reported).[11] If we take these estimates as representative for the world, again assuming a share of 8% for shale oil production in world production, the implied global elasticity would be 0.018, which is far below the bound of 0.04 in Zhou (2019) and is also below the bound of 0.0258 proposed by Kilian and Murphy (2012).[12] If instead we treat the shale oil supply elasticity as effectively zero, we arrive at a global supply elasticity of 0.028, which is still below the bound of 0.04 in Zhou (2019). Bjørnland et al. do not report the standard error of the combined estimate, making it difficult to relate this estimate to the prior specification favored by BH.

How to Interpret the Oil Futures Spread Regressor in Bjørnland et al. (2019)

In response to this conclusion, BH argue that KZ are misreporting the evidence in Bjørnland et al. (2019).[13] BH dispute that the coefficient of 0.03 is the correct measure of the one-month oil supply elasticity for conventional crude oil. They suggest that the one-month oil supply elasticity also depends on the oil futures price, adopting an argument made by Bjørnland et al. (2019)

---

[9] The elasticity estimates reported in different versions of this paper fluctuate widely. Here we focus on the version of this paper discussed by BH.

[10] This estimate is higher than that obtained by Newell and Prest (2019), but also has a larger standard error, reflecting the smaller estimation sample.

[11] To be precise, the standard error of the elasticity of shale oil producers is the sum of the coefficient on the change in the real price of oil and the coefficient on the change in real price of oil interacted with the shale oil dummy. Since no covariance is reported in the table, the standard error for this sum cannot be inferred from the table.

[12] Specifically, $0.92 \times 0.03 + 0.08 \times (-0.12) = 0.018$.

[13] In the interest of space, my response focuses on the results for conventional crude oil. The arguments for shale oil are analogous.

whose analysis differs from other studies in that they include in addition the change in the 3-month oil futures spread in their regression. Bjørnland et al.'s interpretation is that the one-month oil supply elasticity is measured by the sum of the coefficient on the real price of oil (0.03) and the coefficient on the change in the oil futures spread (0.07), implying a statistically insignificant elasticity of 0.1 for conventional oil producers. This argument is not persuasive.

Bjørnland et al. assume a unit root in both the spot price and the futures price of oil. Given that the spread is clearly stationary and hence these prices must be cointegrated, it is unclear why the authors include the first difference of this spread in their regression. This specification choice has important consequences. Note that Bjørnland et al.'s regression,

$$\Delta q_{it} = \beta_1 \Delta p_t + \beta_2 \Delta (p_t - f_t^3) + ... + e_{it}, \tag{1}$$

where q_{it} denotes the log of oil production, p_t is the log of the spot price of oil, and f_t^3 is the log of the 3-month oil futures price, may be equivalently rewritten as

$$\Delta q_{it} = \alpha_1 \Delta p_t + \alpha_2 \Delta f_t^3 + ... + e_{it}, \tag{2}$$

where $\alpha_1 = \beta_1 + \beta_2$. Bjørnland et al. suggest that $\widehat{\alpha}_1$ represents the price elasticity of oil supply. Even if there were a compelling economic rationale for augmenting the standard regression specification by the regressor Δf_t^3, this regression would be problematic from an econometric point of view. The correlation between Δp_t and Δf_t^3 is 98%, creating multi-collinearity and undermining the identification of α_1. The high correlation between these regressors is the likely reason that including Δf_t^3 among the regressors changes the estimate of α_1 substantially compared to setting $\alpha_2 = 0$. Thus, the large elasticity values reported by Bjørnland et al. based on oil futures spreads are not credible.

In fact, there is no compelling reason for including oil futures prices in regression (2) in the first place. Bjørnland et al. think of the oil futures price as a proxy for oil price expectations. There actually is no reason to explicitly model oil price expectations when estimating the oil supply elasticity. If there are exogenous shifts in expectations about future oil prices, this will cause a change in storage demand, which in turn shifts the spot price. The model already captures this effect by including changes in the spot price and allows producers to respond to this type of shock. Thus, the only way for changes in the futures spread to affect oil production directly would be for producers to respond to higher expected oil prices by storing oil below the ground rather than extracting it. For conventional oil, for technological reasons, this is not an option (see Newell and Prest 2019). For shale oil, one could drill, but not frack a well in anticipation of rising prices. Drilled, but not yet completed wells are known as DUCs. However, the contemporaneous correlation between the growth in the number of DUCs in the Bakken and the oil futures spread is only -0.03, suggesting that this effect is small. Moreover, if a producer wants to execute the DUC option in response to higher oil prices, it still takes between four and twelve weeks to complete the well, so the production response in the current month will be negligible at best (see Golding 2019). This fact alone shows that the one-month shale oil supply elasticity estimates of up to 0.9 reported in Bjørnland et al. (2019) are physically impossible.[14]

The IV Oil Supply Elasticity Estimate of Caldara et al. (2019)

Perhaps in recognition of the limitations of the micro evidence in Bjørnland et al. (2019), BH emphasize that they base their case mainly on IV estimates of the oil supply elasticity reported by Caldara et al. (2019) based on country-level evidence for the period 1985-2015. Caldara et al.'s preferred IV elasticity estimate is 0.081 (with a standard error of 0.037). A two-standard

[14] BH also suggest that Bjørnland et al.'s estimates refute the formal theoretical results of Anderson, Kellogg and Salant (2018), which established that the short-run oil supply elasticity is close to zero, when there are adjustment costs in oil production. A more sensible interpretation would be that Bjørnland et al.'s estimates are difficult to reconcile with economic theory.

error confidence band based on this estimate would just barely include the posterior median estimate of the oil supply elasticity reported by BH, although it would not justify allowing much larger elasticity values in their elasticity priors. It is therefore useful to examine this approach in more detail.

Consider a regression of global oil production on the real price of oil. The problem with instrumenting for the real price of oil using exogenous global oil supply disruptions is that the oil supply elasticity depends on the slope of the oil supply curve, which is revealed when the demand curve exogenously shifts along the supply curve. In contrast, when using exogenous global oil supply disruptions as an instrument, these regressions actually measure the demand elasticity. Thus, the choice of the instrument matters a great deal.

It may seem that at the level of individual oil producers it would not make a difference whether the oil price increases due to an oil supply disruption or due to increased demand for oil, but in general it does. For example, Saudi authorities made it clear in the 2000s that they would not respond to oil price increases driven by what they perceived to be shifts in speculative demand for oil, although they have always been willing to respond to exogenous oil supply disruptions driven by geopolitical events. Thus, it is essential to instrument for changes in the price of oil in estimating the oil supply elasticity.

Caldara at al. (2019) propose focusing on the response of oil production in a given country to supply disruptions in other oil-producing countries under the maintained assumption that all oil producers have the same elasticity. Their instrument for the real price of oil consists of a time series of oil supply disruptions in the United States, Mexico, Venezuela, Norway, Iran and various Arab oil producing countries that are considered exogenous by the authors. The elasticity estimate cited by BH relates to the response of all oil producers not directly involved in a given oil supply disruption. Caldara et al. also report elasticity estimates for Saudi Arabia, OPEC excluding Saudi Arabia, and non-OPEC countries that allow for group-specific elasticities.

To understand Caldara et al.'s approach consider the example of an exogenous oil supply disruption in Norway. As the Norwegian oil is removed from the market, other oil producers will experience an exogenous increase in the demand for their oil. Thus, this oil supply disruption may be viewed as a demand shifter for oil producers other than Norway at the country level. This argument, however, is clearly not correct for Saudi Arabia's response to geopolitical oil supply disruptions in other OPEC countries. Since Saudi Arabia aims to offset such disruptions, both its oil supply curve and its oil demand curve will shift in response to such an event, which violates the exclusion restriction required for IV estimation. Since many of Caldara et al. 's exogenous oil supply disruptions take place in OPEC countries, their IV analysis is questionable. This example suggests that the approach taken by Caldara et al. will overestimate the Saudi supply elasticity. The same concern applies more generally to other OPEC producers with spare capacity such as Kuwait and the UAE that have often acted in line with Saudi Arabia in offsetting geopolitical oil supply disruptions.

A simple back-of-the-envelope calculation based on a specific episode of exogenous variation in the real price of oil helps illustrate this point. Between June 2014 and December 2014, the price of oil fell by 44%. There is a debate about the extent to which this price decline was caused by the unexpected rise of U.S. shale oil versus unexpected declines in global demand. Either way this unexpected decline was exogenous from Saudi Arabia's point of view. Given that Saudi Arabia did not respond to any exogenous geopolitical events during this half year, the Saudi production response can be used to cleanly identify the oil supply elasticity. Given the cumulative decline in Saudi oil production of 0.6%, the implied semi-annual Saudi oil supply elasticity is -0.6/-44 = 0.014, which is much lower than Caldara et al.'s statistically insignificant one-month oil supply elasticity estimate of 0.212 for Saudi Arabia. The corresponding semi-annual oil supply elasticity estimate for OPEC is also zero for all practical purposes.

The use of oil supply shock instruments is not only a problem when interpreting the Saudi elasticity estimate and the elasticity estimate for OPEC excluding Saudi Arabia. It is also a problem for the elasticity of all oil producers that are not affected by a given supply disruption, because that set of oil producers includes many OPEC oil producers. As a result, the elasticity estimate favored by BH is systematically overstated. Not surprisingly, the elasticity of non-OPEC countries, which is not affected by this problem, is essentially zero (-0.004) with a standard error of 0.023, yielding an upper bound on the oil supply elasticity of only 0.05, in line with the arguments in KZ. Thus, one cannot give credence to the "elasticity" estimate of 0.081 implied by Caldara et al.'s panel regression or its standard errors.

Moreover, the discussion in Kilian (2008a) suggests that Caldara et al.'s measurement of the exogenous oil supply disruptions is overly simplistic. It suffers from much the same conceptual problems as the Hamilton (2003) OPEC oil supply shock measure, and their classification of the shocks is highly subjective, casting further doubt on the exercise and is sensitive to the definition of the instrument. For example, under the narrow definition of the instrument in Caldara et al. (2019), the supply elasticity estimate drops to 0.054 (with a standard error of 0.019). BH's posterior median estimate of 0.15 exceeds that point estimate by about five standard errors. Moreover, if we eliminate the drop in oil production in the UAE in 1990 on the grounds that this decision was likely unrelated to the invasion of Kuwait, as discussed next, the combined elasticity drops to 0.029, close to the bound of Kilian and Murphy (2012).

The Oil Supply Elasticity Bound of Kilian and Murphy (2012)

Recently, there has been much confusion about the derivation of the original oil supply elasticity bound of 0.0258 in Kilian and Murphy (2012). Our discussion of Caldara et al.'s (2019) IV approach helps understand the rationale for this bound. Kilian and Murphy's approach relied on the fact that the oil supply disruption of August 1990 represented a shift in the demand for oil producers outside of Iraq and Kuwait. These countries' oil-demand curve was further shifted by the sharp rise in storage demand, reflecting expectations that Iraq would invade Saudi Arabia next. Thus, the oil demand curve shifted along the oil supply curve of oil producers not directly affected by the outbreak of the Persian Gulf War. It may seem that therefore the ratio of the percent change in oil production outside Iraq and Kuwait (Δq) over the percent change in the real price of oil (Δp) in August 1990 would be an estimate of the one-month price elasticity of oil supply. This is not the case because. at the same time, the Saudi supply curve shifted to the right along the demand curve, when Saudi Arabia, as the supplier of last resort, responded directly to the OPEC oil supply disruption by expanding its own oil production, along with other oil producers. This created an additional increase in Δq and a decline in Δp, causing the ratio $\Delta q / \Delta p$ to be larger than would be the case in response to the demand shift only. Thus, the ratio $\Delta q / \Delta p = 0.0258$ reported by Kilian and Murphy (2012) and used in many subsequent studies represents an upper bound on the one-month price elasticity of oil supply rather than an estimate of this elasticity. Kilian and Murphy (2012) further stressed that the existence of ample spare capacity in global oil production at the time as well as rare unanimity among oil producers in August1990 about the need to offset the shortfall caused by the war made this supply shift larger than would typically be the case. As we already showed, the implied bound is indeed larger than direct estimates of the one-month price elasticity of oil supply.

Caldara et al. (2019) argue that this bound is misleading. This is not a disagreement about the facts, but about their interpretation. The UAE had produced more oil than allowed by OPEC quotas since late 1989 causing increasing pressure by other OPEC numbers to reign in the UAE oil production. Caldara at al. suggest that the decline in UAE oil production in August 1990, after the UAE finally succumbed to OPEC pressure to lower its oil production in July 1990, must be attributed to a speech by Saddam Hussein on July 17, 1990, threatening some unspecified retribution if the UAE did not reduce its oil production. Caldara at al. thus call for the exclusion of the UAE from the set of countries not directly affected by the war that started on August 2, 1990, which would raise the bound on the one-month oil supply elasticity from 0.0258 to 0.045.

However, Caldara et al.'s argument is not persuasive for two reasons First, by Caldara et al.'s own account, the UAE already agreed to lower its oil production at the OPEC meeting in Jeddah on July 11 several days before Saddam Hussein's speech. Second, at no point was there an immediate military threat to the UAE, which has no direct border with Iraq. Iraq lacked the ability to effectively project military force across the Persian Gulf to the UAE by air or sea, given the presence of U.S. and other opposing forces in the region. In any case, none of the substantive conclusions in Kilian and Murphy (2012, 2014) change when relaxing the upper bound to 0.045, as has been demonstrated in a number of studies.

5. The role of elasticity priors in oil market modeling

BH also suggest that KZ mischaracterized the role of the elasticity priors in oil market modeling and their relationship with the impact of oil supply shocks. I will address each of their points in turn.

How to Correctly Define the Price Elasticity of Oil Demand

BH continue to suggest that the oil market models of Kilian (2009) and Kilian and Murphy

(2012) imply highly implausible one-month price elasticities of oil demand as large as -2 compared to the estimate of -0.35 based on their own oil market model. In response to this point, KZ pointed out that the definition of the price elasticity of oil demand that BH apply to the models of Kilian (2009) and Kilian and Murphy (2012) is incorrect because it ignores the existence of changes in oil inventories. KZ also stressed that it is not possible to recover the properly defined elasticity of oil demand from these models. Computing this elasticity requires the extended model of Kilian and Murphy (2014). The estimate of the price elasticity of oil demand in the latter model is -0.25 and perfectly reasonable, despite having imposed a supply elasticity bound of 0.0258. BH own demand elasticity estimate of -0.35 is actually larger in magnitude than Kilian and Murphy's.

Previously, BH suggested that how one defines the price elasticity of oil demand is simply a matter of taste. BH now have changed their position and make the case that the measure of the oil demand elasticity in their own oil market model actually (at least approximately) corresponds to the proper elasticity definition originally introduced by Kilian and Murphy (2014). This, of course, raises the question of why BH even in their latest paper continue to compare their demand elasticity estimate to the incorrectly defined oil demand elasticity based on the models of Kilian (2009) and Kilian and Murphy (2012). Clearly, that approach is deceptive. Moreover, it should be noted that the approximation appealed to by BH is poor. It is well known from the evidence in Kilian and Murphy (2014) and Herrera and Rangaraju (2019) that, in models for which both elasticities can be computed, the incorrectly computed impact price elasticity of oil demand is substantially higher in magnitude than the correctly computed oil demand elasticity, so BH’s claim that the distinction does not matter in practice is at odds with the evidence.

Can the Properly Defined Oil Demand Elasticity Be Derived from the Models of Kilian (2009) and Kilian and Murphy (2012)?

As pointed out in KZ, the properly defined price elasticity of demand that allows for storage cannot be recovered from oil market models such as Kilian (2009) or Kilian and Murphy (2012) that - unlike the Kilian and Murphy (2014) model - do not include oil inventories. BH dispute this fact, but their counterargument is logically flawed. They make the case that in their own oil market model the coefficient measuring the price elasticity of demand captures the proper demand elasticity, even after dropping oil inventories from their model. Without getting into the details of this result, BH's argument is missing the point. It does not disprove or otherwise address the statement in KZ, which did not relate to the BH oil market model, but to the Kilian (2009) and Kilian and Murphy (2012) oil market models. Thus, the assertion that the latter oil market models imply implausibly large demand elasticities, which has been interpreted by some applied researchers as an indication that these models are misspecified, is without basis.

More generally, this discussion shows that BH's assertion that low oil supply elasticities necessarily map into implausibly large oil demand elasticities, which later was accepted as the premise for the work of Caldara et al. (2019), is not supported by the facts. This point is far from an esoteric detail. For example, it calls into question Caldara et al.'s (2019) conclusion that setting the oil supply elasticity to 0.1 in oil market models implies that oil supply and oil demand shocks are equally important drivers of oil price fluctuations. This point was already shown to be incorrect in Kilian and Murphy (2012). Correcting such misperceptions is important for the direction of future work on oil markets.

Why the Elasticity Definition in Kilian and Murphy's (2014) Model Makes Sense

The perhaps most interesting contribution of BH is a clarification of the differences in the definition of the elasticity concept used by BH in their own oil market model and the definition used by Kilian and Murphy (2014) and many other recent oil market studies. My discussion of this point focuses on the one-month price elasticity of oil supply. BH define the oil supply elasticity as the impact response of oil production to an increase in the real price of oil triggered by an exogenous demand shift, holding constant not only the remaining structural shocks, but also all other variables in the model such as global real economic activity and oil inventories. This parameter is uniquely measured in their model.

In contrast, Kilian and Murphy (2014) define the one month price elasticity of oil supply as the impact response of oil production to the increase in the real price of oil triggered by an exogenous demand shift, allowing global real activity and oil inventories to respond contemporaneously to the exogenous demand shift. This elasticity measure corresponds to the ratio of the impact responses of global oil production and of the real price of oil to a given exogenous demand shock, while all other structural shocks remain zero.[15] This is a natural generalization of the supply elasticity concept in the textbook two-variable model, which requires that there be no other exogenous variation but the shift in demand.

The key difference is that BH's elasticity definition is a theoretical construct that one is not likely to observe in the data since both global real activity and oil inventories will in general move on impact in response to a demand shock. For example, the elasticity bound derived by Kilian and Murphy (2012) - or for that matter the alternative bound discussed by Caldara et al. (2019) - does not hold constant the remaining model variables. The same is true for the microeconomic estimates of the oil supply elasticity discussed in section 4 and for the IV estimates in Caldara et al. (2019). Thus, it makes sense to choose an elasticity definition that

[15] Since there is more than one demand shock in this model, the implied estimate of the oil supply elasticity is not unique. Kilian and Murphy impose the same bound on the implied supply elasticities, but do not force their values to be identical, since doing so would greatly complicate the econometric analysis.

corresponds to empirical elasticity estimates in the literature, which is what Kilian and Murphy (2014), along with many other researchers, have done. This approach is internally consistent.

In contrast, BH are in no position to appeal to any of the microeconomic elasticity estimates (or elasticity bounds) discussed in section 4 in motivating their prior specification because these estimates are inconsistent with their elasticity definition. Of course, the analysis in Herrera and Rangaraju (2019) suggests that this difference in the definition of the impact oil supply elasticity makes little difference for the model estimates in practice.

How Informative are BH's Oil Supply Elasticity Priors?

BH claim to be unaware of the link between the value of the oil supply elasticity and the importance of oil supply shocks for the real price of oil. This claim is surprising since BH seem to be well aware of the studies of Kilian and Murphy (2012), Herrera and Rangaraju (2019) and Zhou (2019) which studied the impact of bounds on the oil supply elasticity on the impulse responses. In defense of their decision to abandon all oil supply elasticity bounds, BH insist that a diffuse prior on the value of the oil supply elasticity is agnostic and hence will help recover the unknown elasticity value from the data. The problem is that seemingly agnostic priors on the oil supply elasticity can be highly influential when working with sign-identified oil market models, as shown by Kilian and Murphy (2012) in the context of global oil market models.

BH stress that they are not imposing any restrictions on the oil supply elasticity. They even go to the trouble of proving this result. KZ agree on this point. KZ's concern instead is that BH should have restricted that elasticity in line with extraneous information about oil supply elasticities from the microeconomic literature and with insights from economic theory. This remains true even after accounting for estimation uncertainty in the microeconomic estimates. In other words, KZ object to BH not imposing the relevant identifying restrictions on the oil supply elasticity.

BH suggest that KZ “say very little” about the fact that BH also reported results for an alternative oil supply elasticity prior that assigns 80% probability to the elasticity being bounded between 0 and 0.0258 and 20% probability to the elasticity being unbounded on the positive side. Their reply shows that BH fail to appreciate KZ’s point that both their baseline prior and this alternative prior share the same problem that they leave the value of the oil supply elasticity unrestricted from above. The support of both of these priors extends from zero to infinity. By construction, any prior that allows for values of the oil supply elasticity that are implausibly large, given the microeconomic evidence, is inappropriate. Kilian and Murphy (2012) already showed that oil supply shocks have potentially very large effects on the real price of oil, when allowing for unbounded oil supply elasticity values. BH merely illustrate this well-known finding.

Thus, the disagreement between BH and other researchers is not about whether we need to model uncertainty about identifying restrictions. Nor is it primarily about how to conduct Bayesian inference in structural VAR models or about the specification of the oil market model, although there are divergent views on this point as well. At its core, the difference in views between BH and KZ comes down to how plausible the oil supply elasticity values are that BH wish to allow for and that KZ insist should not be allowed for. The answer depends one's views about how to read the micro evidence on the one-month global oil supply elasticity. I have already elaborated in section 4 on why the evidence does not support BH's views on the economic plausibility of large one-month oil supply elasticity values.

6. The Merits of the Kilian Index of Global Real Economic Activity

BH selectively introduce ad hoc measurement error in the modeling of oil inventories, while ignoring measurement error in other model variables. KZ explained in detail why that approach is questionable. BH's retort is to criticize Kilian and Murphy (2014) for having discussed the reliability of the oil inventory data, but not of their global real activity data. Since Kilian and Murphy (2014) first introduced changes in oil inventories into global oil market models, it is not surprising that they devoted a section to discussing these data. In contrast, their measure of global real activity had been introduced in Kilian (2008b, 2009) and was well established by 2014, so there was no need to have a section on this index.[16] However, I am happy to address the main concerns about this index raised by BH in this section, because using an appropriate measure of the global business cycle is a pre-condition for identifying the role of demand and supply shocks in industrial commodity markets (see Kilian and Zhou 2018). My discussion is based on the index as reported in Kilian (2019).

Specifically, BH reiterate four claims recently made by Hamilton (2019), namely that (1) the fact that the Kilian index reaches its lowest level in 2016 implies a deeper recession in 2016 than at any other time in history; (2) that the cyclical component of BH's measure of global industrial production has a higher correlation with world real GDP than the Kilian index; (3) that the Kilian index is not helpful in forecasting real commodity prices; and (4) that the linear trend specification underlying the construction of the Kilian index is rejected by statistical tests. I will briefly address each of these arguments:

(1) The fact that the level of the Kilian business cycle index in early 2016 briefly drops below the level in late 2008 for the reasons discussed in Kilian and Zhou (2018) does not imply a bigger recession in 2016 than in 2008. The NBER business cycle dating committee identifies the months when the economy reaches a peak of activity and later months when the economy reaches a trough. The time in between is a recession, defined as a period when economic activity is contracting. Thus, the depth of a recession is measured by the extent to which real activity declines from peak to trough, not by the lowest level of real activity during the recession. Using

[16] Further discussion of this index and alternative proxies for global real activity can be found in Kilian (2009), Kilian (2019) and Kilian and Zhou (2018).

the NBER definition of a recession, the decline in 2016 is too short to be called a recession at all, and its magnitude is only about one third of the decline in late 2008.

In response to this point, Hamilton (2019) recently changed his argument, further muddying the waters. First, he misstates the starting date of the brief drop in the Kilian index in early 2016 as July 2015, implying that this episode lasted longer than it did and overstating the magnitude of the decline. Second, Hamilton now draws attention to the decline in the index from December 2013 to February 2016, ignoring that the sharp drop in early 2016 does not reflect cyclical variation, but represents an outlier that was quickly reversed. When excluding this outlier, there is indeed a sustained decline in the index in 2014 and 2015, consistent with a wide range of other indicators, as discussed in Kilian and Zhou (2018), but this decline is only half as large as that in late 2008. Third, the NBER explicitly states that "recessions start at the peak of a business cycle and end at the trough". Hamilton suggests that it is difficult to apply the NBER definition to the Kilian business cycle index because peaks and troughs are open to interpretation. It is not clear why dating the business cycle is harder for this index compared to other indices.

Finally, Hamilton (2019) suggests that Kilian's (2009) philosophy of measuring the business cycle precludes defining recessions as done by the NBER. In support of this strange argument, he cites Kilian (2009) as stating that the index is proportionate to deviations of the level of real activity from trend. This statement was intended to draw attention to the fact that numerical values of the Kilian index have no inherent meaning, only its relative changes over time. The fact that Kilian (2009) measures the level of real activity relative to trend, however, in no ways precludes applying the NBER definition of a recession. In fact, the discussion of global booms and global recessions in Kilian (2009, p. 1057) is fully consistent with the NBER definition. There is no support for Hamilton's insinuation that there is a disconnect between the analysis in Kilian (2009) and in Kilian (2019).

(2) The Kilian index was designed for modeling the business cycle in industrial commodity markets. It is a proxy for changes in 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. The Kilian index was, in fact, constructed as an alternative to world real GDP because world real GDP is not only poorly measured, but is an inappropriate measure of global real activity in industrial commodity markets. Thus, the validity of the Kilian index does not depend on being a good proxy for (or predictor of) world real GDP or, for that matter, BH's measure of global industrial production.[17]

(3) Hamilton (2019) did not conduct any forecasting exercise at all, but only reported results for the in-sample fit of some ad hoc regressions. His analysis is woefully inadequate by the standards of the existing literature on forecasting real commodity prices, and his results are contradicted by other recent studies (e.g., Baumeister and Kilian 2014b; Garratt, Vahey and Zhang 2019)

[17] This is not the focus of my reply, but it should be noted that Hamilton's mixed frequency regression analysis is highly questionable. Hamilton not only suppresses lags in dealing with the Kilian index, but his regressions are akin to regressing the annual growth rate on the monthly output gap. Not surprisingly, his empirical results are contradicted by other studies such as Ravazzolo and Vespignani (2019).

(4) The statistical tests used by Hamilton (2019) to reject the linear trend specification are invalid. Hamilton presents results of tests of the I(1) null and tests of the I(0) null. He reports being unable to reject the unit root null using the ADF test, but being able to reject the null of stationarity about a linear deterministic time trend using the KPSS test of Kwiatkowski et al. (1992).

This type of confirmatory analysis was fashionable in the 1990s, but has been shown to be misleading. The intuition is simple. It is evident that the data underlying the Kilian index are highly persistent. The apparent existence of long cycles in these data is, in fact, what motivated the analysis in Kilian (2009). For such data, the finite-sample power of tests of the unit root null based on autoregressions tends to be negligible. Thus, the fact that Hamilton cannot reject the unit root null is not surprising. Since the null distribution and the distribution under the alternative of this test overlap to a large extent, we cannot discriminate between these hypotheses based on the data. Because the null hypothesis is protected from rejection in classical hypothesis testing, we necessarily fail to reject the null in this case. This does not mean that the data support the null hypothesis, but that the data are not informative about the hypothesis of interest.

This raises the question of how the KPSS test can yield such decisive results simply by reversing the null of the test. After all, it is still true that the null distribution and the distribution under the alternative of this test largely overlap. Caner and Kilian (2001) trace the tendency of tests of the I(0) null to reject in such situations to the fact that asymptotic critical values for these tests have been constructed under the null of white noise. If these critical values are applied to stationary, but persistent time series, the KPSS test will suffer from potentially severe size distortions. Caner and Kilian demonstrate that rejection rates under the null as high as 70% are not uncommon in applied work, when using asymptotic critical values.

Caner and Kilian (2001) show that addressing this problem requires the user to bootstrap the regression model under the null of the best fitting stationary, but persistent process (possibly with bias corrections as in Kilian (1999)). The resulting bootstrap critical values are invariably higher, resulting in non-rejections of the null of trend stationarity, consistent with the evidence from tests of the unit root null. In fact, Caner and Kilian show by simulation that the power of the size-corrected bootstrap version of the KPSS test is even lower than the already low power of the standard augmented Dickey-Fuller test.

Subsequently, this problem was studied in depth from a theoretical point of view by Müller (2005) who used the device of a local-to-unity framework to represent situations in which conventional tests are uninformative. To quote from the abstract of Müller's paper:

"Tests of stationarity are routinely applied to highly autocorrelated time series. Following Kwiatkowski et al. (J. Econom. 54 (1992) 159), standard stationarity tests employ a rescaling by an estimator of the long-run variance of the (potentially) stationary series. This paper analytically investigates the size and power properties of such tests when the series are strongly autocorrelated in a local-to-unity asymptotic framework. It is shown that the behavior of the tests strongly depends on the long-run variance estimator employed, but is in general highly undesirable. Either the tests fail to control size even for strongly mean reverting series, or they are inconsistent against an integrated process and discriminate only poorly between stationary and integrated processes compared to optimal statistics."

In fact, Müller (2008) proves the impossibility of statistically discriminating between the I(0) and I(1) hypothesis, even with an infinite amount of data, so what Hamilton claims to have done is plainly impossible. In short, his claim that statistical tests show that the construction of the Kilian index is invalid is without basis.

7. How Does BH's Contribution Fit into the Literature?

When reading BH, it is not easy to keep track of what their value added is. For example, BH seem conflicted about the value added of their econometric approach. KZ characterized BH's approach as extending the structural VAR framework of Sims and Zha (1998) by specifying a prior distribution on all the structural parameters of the model. BH dispute this interpretation, asserting that their main contribution is not to generalize Sims and Zha (1998), but to relax the frequentist approach to identification. Not only is the term "frequentist" not used in Baumeister and Hamilton (2019b), but this reinterpretation of their own work seems confused. Identification in economic models derives from economic reasoning. It is not inherently frequentist or Bayesian. Kilian (2009), for example, make the case for a specific recursive structure of the structural impact multiplier matrix based on economic arguments. These identifying restrictions remain the same whether the model is estimated and inference is conducted using frequentist or Bayesian methods. Likewise, inequality restrictions on impulse response are not inherently frequentist. In fact, most applications of such VAR models are Bayesian. What BH actually do is to propose a specific approach to parameterizing uncertainty about identifying restrictions within a Bayesian framework.

Nor are BH precise about what their substantive value added is. For example, they give the misleading impression of being the first researchers to have considered the possibility that the impact price elasticity of oil supply may be larger than the bound of 0.0258 proposed by Kilian and Murphy (2012). Actually, Kilian and Murphy (2012, section 3.4) explored elasticity values of 0.05, 0.08 and much higher. Kilian and Murphy showed that their results are quite robust to imposing larger elasticity bounds than supported by their evidence. More recent studies have confirmed this point for the model of Kilian and Murphy (2014) (see Zhou 2019; Herrera and Rangaraju 2019).

Likewise, BH's rhetorical question of whether we really know nothing about the impact price elasticity of oil demand falsely implies that previous studies failed to impose any further identifying information about this elasticity. It ignores that bounds on the impact price elasticity of oil demand have been standard in the literature, ever since Kilian and Murphy (2014) proposed bounding this elasticity by zero from above and by extraneous microeconomic estimates of the long-run elasticity from below (e.g., Kilian and Lee 2014; Kilian 2017; Kilian and Zhou 2019b,c; Zhou 2019; Herrera and Rangaraju 2019; Cross 2019).[18]

In earlier versions of their work, BH asserted that Kilian (2009) and Kilian and Murphy (2012, 2014) failed to impose all relevant identifying assumptions about the oil demand and oil supply elasticities. However, they could never explain what this additional identifying information is and how they would have used that information to come up with a prior for the elasticities. In response to Kilian and Lütkepohl (2017) making this point, BH subsequently dropped this argument completely and proposed a diffuse oil supply elasticity prior instead, arguing that there is no harm in using diffuse elasticity priors for identification.

Based on this prior, BH concluded that oil supply shocks are a much more important determinant of the real price of oil than earlier studies such as Kilian and Murphy (2012, 2014) suggested. In fact, their baseline prior for the oil supply elasticity can be shown to resemble the posterior of this elasticity obtained from the Kilian and Murphy (2012) model, when imposing no identifying bounds at all on the oil supply elasticity. Kilian and Murphy (2012) already

[18] The demand elasticity bound of -0.8 proposed by Kilian and Murphy (2014) based on extraneous microeconomic estimates of the long-run price elasticity of gasoline demand is conservative, given that the price elasticity of gasoline demand tends to be higher than the price elasticity of oil demand.

demonstrated that this approach results in economically implausible models being considered admissible, invalidating summary statistics computed from the posterior, so BH's results are not surprising. BH's alternative prior is equally unbounded from above and generates identical estimates and conclusions. As shown by Herrera and Rangaraju (2019), when imposing any reasonably tight oil supply elasticity bound, these results are greatly diminished, and the effect of oil supply shocks on the real price of oil in BH's model are not larger than in other oil market models.

Thus, the debate launched by BH is not about relaxing the assumption of a one-month oil supply elasticity of zero made in Kilian (2009). Nor is it about whether we should allow for uncertainty in the elasticity value. That was already done by Kilian and Murphy (2012, 2014). Rather, at its core, the debate is about whether one-month oil supply elasticity values of 0.15, of 0.9, or of ∞, for example, all of which BH consider a priori plausible, can be defended from an economic point of view. It goes without saying that allowing for an infinite elasticity in the prior specification makes no economic sense. Even if this oil supply elasticity is only 0.15 (0.3), however, this implies that a 10% unexpected price increase caused by higher demand is associated with an increase of 1.5% (3%) in global oil production within one month. Even without considering the microeconomic evidence, such increases seem unrealistically large. Oil producers may be able to announce plans to increase production, but materially changing actual production on such short notice is difficult. This is a question where knowledge of the oil industry can help immensely in understanding what is feasible and what is not. As discussed in Newell and Prest (2019, p. 16), : "once a well has been drilled, its flow rate is determined primarily by geology and is therefore largely beyond the operator's control." (p. 16). The microeconomic estimates I presented in section 4 and the theoretical results in Anderson et al.

(2018) are consistent with this view. BH do not present credible new evidence in support of higher elasticities.

8. Conclusion

After controlling for the prior, there is no evidence that BH's alternative methodology generates substantively different results from the mainstream Bayesian methods of evaluating sign-identified structural VAR models discussed in Rubio Ramirez, Waggoner and Zha (2010), Inoue and Kilian (2013, 2019), and Arias et al. (2018), among others. Instead, the differences in results from earlier studies are driven by BH's data and modeling choices and, most importantly, by their prior specification for the one-month oil supply elasticity. I explained why that prior specification (and why BH's posterior estimate of that elasticity) is at odds with extraneous evidence. When using economically plausible priors, BH's approach generates similar results to those based on the model proposed by Kilian and Murphy (2014), as has been demonstrated by Herrera and Rangaraju (2019). Thus, what this debate has shown is that there is no credible evidence that oil supply shocks are more important than suggested by earlier oil market studies. Rather the conclusions of Kilian and Murphy (2014) are reaffirmed. Explicitly modeling the uncertainty, as proposed by BH, appears to make no difference compared with the conventional Bayesian and frequentist modeling approaches used in the existing literature.

This is not to say that the literature on modeling oil markets can only be advanced within the framework of Kilian (2009) or Kilian and Murphy (2012, 2014). There continues to be much interest in exploring new identification schemes, new econometric approaches and new identifying information for oil market models (e.g., Stürmer 2018; Antolin-Diaz and Rubio-Ramirez 2018; Känzig 2019; Lanne and Luoto 2019; Bruns and Piffer 2019). It is healthy for a literature to continuously question its foundations. Some of these new ideas will stand the test of time and some will not, but the profession advances in the process. There has also been recent work to extend the scope of global oil market models in new directions by incorporating exchange rates, interest rates and changes in strategic petroleum reserves (e.g., Kilian and Zhou 2019b,c). Another important challenge, going forward, is how to adopt oil market models to the rapidly growing importance of shale oil in the global market place. KZ made the case that the growth of U.S. shale oil production for the time being has not invalidated the current class of VAR oil market models, but this question has to be evaluated on a continuous basis.

At the same time, there is an ongoing effort to develop better global data for modeling oil markets. For example, Delle Chiaie, Ferrara and Giannone (2016), Ravazzolo and Vespignani (2019), and Alquist, Bhattarai and Coibion (2019) have proposed new and innovative measures of global real economic activity that complement existing approaches (see Kilian and Zhou 2018). There is even the prospect that new, more reliable measures of changes in oil inventories based on satellite data may help refine the conclusions from oil market models. Big data may also revolutionize the way we think about modeling oil shipping markets (e.g., Regli and Nokimos 2019). It would be surprising if these efforts did not improve our understanding of oil markets and their relationship with the economy over time.

References

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

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.

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.

Baumeister, C., and J.D. Hamilton (2015), "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 83, 1963-1999.

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

Baumeister, C., and J.D. Hamilton (2019b), "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 L. Kilian (2014a), "Do Oil Price Increases Cause Higher Food Prices?," *Economic Policy*, 80, 691-747.

Baumeister, C., and L. Kilian (2014b), "What Central Bankers Need to Know about Forecasting Oil Prices," *International Economic Review*, 55, 869-889.

Baumeister, C., and L. Kilian (2016a), "Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us," *Journal of Economic Perspectives*, 30, 139-160.

Baumeister, C., and L. Kilian (2016b), "Understanding the Decline in the Price of Oil since June 2014," *Journal of the Association of Environmental and Resource Economists*, 3, 131-158.

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

Bruns, M., and M. Piffer (2018), "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.

Caner, M., and L. Kilian (2001), "Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate," *Journal of International Money and Finance*, 20, 639-657.

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 (2016), "Common Factors in Commodity Prices," manuscript. Federal Reserve Bank of New York.

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

Garratt, A., Vahey, S. P. and Zhang, Y. (2019), "Real-Time Forecast Combinations for the Oil Price," *Journal of Applied Econometrics*, 34, 456-462.

Golding, G. (2019), "Don't Expect U.S. Shale Producers to Respond Quickly to Geopolitical Supply Disruption," https://www.dallasfed.org/research/economics/2019/1003.

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*.

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.

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

Kilian, L. (1999), "Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Responses," *Review of Economics and Statistics*, 81, 652-660.

Kilian, L. (2008a), "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. (2008b), "The Economic Effects of Energy Price Shocks," *Journal of Economic Literature*, 46, 871-909.

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

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

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 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), "Oil Supply Shock Redux?" manuscript, Federal Reserve Bank of Dallas.

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

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

Koop, G. (1996), "Parameter Uncertainty and Impulse Response Analysis," *Journal of Econometrics*, 72, 135-149.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159-178.

Lanne, M., and J. Luoto (2019a), "Useful Prior Information in Sign-Identified Structural Vector Autoregression: Replication of Baumeister and Hamilton (2015)," manuscript, University of Helsinki.

Lanne, M., and J. Luoto (2019b), "Identification of Economic Shocks by Inequality Constraints in Bayesian Structural Vector Autoregression," forthcoming: *Oxford Bulletin of Economics and Statistics.*

Müller, U.K. (2005), "Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series," *Journal of Econometrics*, 128, 195-213.

Müller, U.K. (2008), "The Impossibility of Consistent Discrimination between I(0) and I(1) Processes," *Econometric Theory*, 24, 616-630.

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

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

Ravazzolo, F., and J.L. Vespignani (2019), "A New Monthly Indicator of Global Real Economic Activity," forthcoming: Canadian Journal of Economics.

Regli, and N. Nomikos (2019), "The Eye in the Sky – Freight Rate Effects of Tanker Supply," *Transportation Research Part E*, 125, 402-424.

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 (1998), "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39, 949-968.

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

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

Terzian, P. (1985), *OPEC. The Inside Story*, Zed Books, London.

Uhlig, H. (2005), “What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure,” *Journal of Monetary Economics*, 52, 381-419.

Waggoner, D., and T. Zha (2012), “Confronting Model Misspecification in Macroeconomics,” *Journal of Econometrics*, 171, 167-184.

Yergin, D. (1992), *The Prize: The Epic Quest for Oil, Money & Power*, Simon and Schuster, New York.

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

Zha, T. (1999), “Block Recursion and Structural Vector Autoregressions,” *Journal of Econometrics*, 90, 291-316.