Structural Interpretation of Vector Autoregressions with Incomplete Identification: Setting the Record Straight

August 16, 2019

Christiane Baumeister
University of Notre Dame
cjsbaumeister@gmail.com

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

ABSTRACT

A recent paper by Kilian and Zhou (2019) mischaracterizes our 2019 paper in *American Economic Review* and much of the related literature. They misstate our contribution to the literature on identification, mischaracterize the role of prior information about supply elasticity in our analysis, inaccurately describe the relation between structural elasticities and the impacts of shocks, and mischaracterize the literature on supply elasticity. Our purpose in this paper is to set the record straight.

1 Introduction.

Kilian and Zhou (2019) (hereafter KZ) offer a critique of Baumeister and Hamilton (2019) (hereafter BH). KZ's critique is so inaccurate, both in broad substance and in specific details, that we hardly recognize our paper in their description. Section 2 of this paper begins with a summary of what BH actually did. Section 3 clarifies the differences between our approach and that in Kilian and Murphy (2014). Section 4 establishes that one of KZ's misguided criticisms of BH is based on a misunderstanding of the relation between structural elasticities and the impacts of shocks. Since this misunderstanding also shows up in several other published papers, our clarifications on this issue may be of independent interest. Section 5 discusses many other inaccuracies in KZ. These include misrepresentation of the flexibility of our approach, mischaracterization of the literature on measuring supply elasticity and global real economic activity, and misleading statements about loss functions.

2 The core contribution of Baumeister and Hamilton (2019).

We stated our main purpose in writing BH in the paper's abstract:

> Traditional approaches to structural vector autoregressions can be viewed as special cases of Bayesian inference arising from very strong prior beliefs. These methods can be generalized with a less restrictive formulation that incorporates uncertainty about the identifying assumptions themselves. We use this approach to revisit the importance of shocks to oil supply and demand.

To demonstrate the claim in the first sentence, we used two examples taken from the literature on oil supply and demand. The first example was the model in Kilian (2009). This used a 3-variable vector autoregression based on world oil production, a measure of world economic activity, and the real price of oil. Kilian gave this VAR a structural interpretation by using a Cholesky identification strategy, in which the first equation was interpreted as the oil supply curve. Though his analysis was entirely frequentist, BH showed that Kilian's approach could be viewed as a special case of Bayesian inference. The prior distribution associated with this Bayesian interpretation assumes that the analyst knows with certainty before seeing the data that supply has zero response on impact to the price, but has no useful information at all about the response of demand to price. We showed that Kilian's (2009) results could be replicated exactly with a Bayesian analysis using these particular prior distributions to represent the researcher's beliefs before seeing the data.

The second example we used was the model in Kilian and Murphy (2012), which was based on the same 3-variable VAR. In this case, Kilian and Murphy's (2012) structural interpretation came from restrictions on the signs and magnitudes of certain effects. Again although Kilian and Murphy did not describe theirs as a Bayesian approach, BH showed that their results could again be reproduced as a special case of Bayesian inference. A key component of the prior distribution implicit in Kilian and Murphy's analysis was that the Bayesian knows with certainty before seeing the data that the short-run supply elasticity could not be larger than 0.0258.

BH followed these demonstrations with a section titled, "Do we really know for certain that the oil supply elasticity is less than 0.0258?" After reviewing the literature, we concluded that the answer is no – a more reasonable analysis would acknowledge some possibility that this parameter could be larger than 0.0258. The next section was titled, "Do we really know nothing about the elasticity of demand?" After reviewing the literature, we concluded that the answer was again no – we have good reason to believe that the short-run demand elasticity, like the short-run supply elasticity, is small. We suggested that the traditional approach to identification – for example, assuming we know the supply elasticity perfectly but know nothing about demand – could be replaced by a Bayesian approach in which the analyst has weak but imperfect information about both supply and demand, with this information represented in the form of probability distributions. If the information is very good, the prior distribution has a small variance. Achieving identification by imposing a particular value for the supply elasticity is a special case of Bayesian analysis as the variance of the prior distribution goes to zero. If the information is weak, the prior distribution has a very large variance, and posterior credibility sets will be very large, accurately conveying the researcher's doubts about the identifying assumptions themselves.

We illustrated how this could be done using a 4-variable VAR. Our baseline model assumed that the analyst had weak, but far from perfect, information that both supply and demand elasticities were likely small. We treated the prior information about these two magnitudes in exactly the same way. For the supply elasticity, we used a Student t distribution with location parameter 0.1, scale parameter 0.2, and 3 degrees of freedom, truncated to be positive. For the demand elasticity, we used a Student t distribution with location parameter -0.1, scale parameter 0.2, and 3 degrees of freedom, truncated to be negative.

KZ focus in detail on the prior distribution used by BH for the short-run supply elasticity. KZ claim that this distribution "really amounts to imposing positive probability mass on elasticity values larger than can be supported by extraneous evidence" (page 11) and that BH's conclusions result mainly from "the imposition of a highly unrealistic prior for the impact price elasticity of oil supply" (page 19). KZ plot this truncated Student t distribution in their Figure 2 and add annotations purporting to show the inconsistency between this distribution and previous point estimates from the literature. We reproduce that distribution in the dashed black curve in our Figure 1.

KZ say very little about the fact that BH also repeated the analysis with a prior for the supply elasticity designed to tilt very strongly in favor of the upper bound of 0.0258 that was imposed by Kilian and Murphy (2012, 2014). This prior distribution assigns 80% probability to the elasticity being uniform over $(0, 0.0258)$ and 20% to the truncated Student t(0.1,0.2,3). This distribution is plotted as the solid red curve in Figure 1. The mixture distribution has a median at 0.0158 and assigns an 81% probability that the value is less than 0.0258. BH reported how various details of our conclusions would be affected if we replaced the dashed black prior distribution with the solid red. Those details are reproduced in Table 1 below. None of our results change if we replace the Student t prior with the mixture prior.

All this was reported in BH. There we also explained the mathematics for why the results in the two columns of Table 1 are so similar:

> The Bayesian posterior distribution is a weighted average of the likelihood, with weights given by the prior density. If the prior density has a very large variance, the weights are approximately uniform over the range for which the likelihood has nonnegligible mass, and the posterior is essentially the same as the likelihood, with the prior exerting no influence on the posterior.

BH developed this point both algebraically and numerically. At no point do KZ challenge any of this mathematical analysis. Their claim that our results come mainly from "imposition of a prior on the impact price elasticity of global oil supply that attaches unrealistically large probability mass to high elasticity values" (page 2) is simply false.

KZ use the word "impose" in connection with use of a prior 43 different times in their paper. This expression is misleading and misguided. Prior information is something that is used, not imposed. How much influence the prior information has depends primarily on the variance of the prior. If the variance is large, the researcher is honestly saying that he or she has essentially no confidence in the information, and using that information has essentially no effect. It is inaccurate to describe something that has no effect on the inference as "imposing" some prior information.

The standard deviation of the truncated Student t(0.1,0.2,3) distribution is 0.25. As shown in detail in BH, this large variance means that the prior does not significantly distort the information in the likelihood. There is no sense in which this distribution *imposes* a value for the short-run supply elasticity that is bigger than 0.0258. Instead, what this distribution does is *allow* the possibility that the value could be bigger than 0.0258. By contrast, the $U(0, 0.0258)$ implicit prior in Kilian and Murphy (2012) has a standard deviation of only 0.0074. The tiny variance and dogmatic upper bound causes the prior used by Kilian and Murphy to have a huge influence on their results. We elaborate more on this point in the next section.

3 Differences between Baumeister and Hamilton (2019) and Kilian and Murphy (2014).

BH used Kilian (2009) and Kilian and Murphy (2012) (hereafter KM12) to illustrate how traditional approaches to identification can be viewed as special cases of Bayesian inference. KZ criticize BH for using KM12 for this illustration instead of using Kilian and Murphy (2014) (hereafter KM14). BH chose to use KM12 instead of KM14 because KM12 provides a clearer illustration of the methodological points we wanted to make. However, it is straightforward to address KZ's request for more discussion of the relation between our analysis and that in KM14. We do so now.

KM14, like KM12, imposed the condition that the supply elasticity has to be below 0.0258. The simplest way to demonstrate the role this plays in KM14's conclusions is to start from exactly KM14's specification – their model, their data, their algorithm exactly. Running the code publicly posted at the *Journal of Applied Econometrics* data archive generates 5 million draws for the vector of possible parameters.[1] Their algorithm then rules out various draws based on a long list of criteria. The end result of running the code is that only 16 of the original 5 million draws remain at the end of the selection process. In the upper left panel of Figure 2, we plot the histogram[2] of the short-run supply elasticity implied by this set of 16 accepted draws.

In the section of BH titled, "Do we really know for sure that the oil supply elasticity is less than 0.0258?" we noted that KM12 obtained the value of 0.0258 by dividing the increase in oil production in countries other than Iraq and Kuwait in August of 1990 (1.17%) by the increase in oil price in August of 1990 (45.3%); that is, $0.0258 = 1.17/45.3$. We then called attention to Caldara, Cavallo and Iacoviello's (2019) (hereafter CCI) observation that a primary reason that the increase outside of Iraq and Kuwait was as small as it was (1.17%) was because of the 19.5% cut from United Arab Emirates. CCI concluded from a careful analysis of statements and newspaper articles that this cut was a response to fear of military action from Iraq if U.A.E. had not immediately cut its production. In August 1990, the increase in production from countries other than Iraq, Kuwait and U.A.E. was 1.95%. If we used the response of countries other than Iraq, Kuwait, and U.A.E., instead of KM12's use of the response of countries other than Iraq and Kuwait, we would have arrived at an estimated short-run supply elasticity for this episode of $1.95/45.3 = 0.043$. Zhou (forthcoming) proposes to use 0.04 as the upper bound, and that number is now treated at various points in KZ as perfectly plausible.

The upper right panel of Figure 2 reruns the KM14 code with a single change – the upper bound of 0.0258 is replaced by 0.043. We made no changes in their code other than this.

[1] The KM14 original code and our results from running it are available at http://econweb.ucsd.edu/~jhamilton/BH4_code.zip.

[2] All histograms shown here are plotted as densities, that is, the area under each histogram is exactly unity.

Now 59 of the 5 million draws get retained.

Returning to the original discussion in BH, we continued:

> Caldara, Cavallo and Iacoviello further noted that August 1990 was but one of dozens of historical episodes like this that could have been used for such calculations. Other examples include strikes by Norwegian oil workers in 1986, attacks on Libyan oil fields in 2011, and hurricanes disrupting Mexican production in 1995 and U.S. production in 2005 and 2008.

CCI used this broad set of disruptions as instrumental variables to arrive at an estimate of the short-run supply elasticity of 0.081. KZ adopt the odd rhetorical device of using any point estimate as if it represents an upper bound on the possible value. But their formal claim from a Bayesian perspective is that there is zero possibility that the elasticity could be larger than some specified value x. Their response to any new evidence is, well, perhaps x may be a larger value than we originally supposed. The CCI estimate of 0.081 has a standard error of 0.037. If one were constructing a true upper bound based on this evidence alone, the upper bound would not be 0.081 but something like $0.081 + (2)(0.037) = 0.155$. The lower left panel of Figure 2 reruns the KM14 code with 0.155 as the upper bound. There are now 1175 retained draws.

Associated with each of the 16 values in the histogram in the upper left panel of Figure 2 is a parameter vector that implies a value for any other magnitude of interest. For example, the upper left panel of Figure 3 summarizes the 16 implied values for the demand elasticity as calculated by KM14.[3]

A fundamental question is how we are supposed to interpret the 16 different draws in the upper left panels of these figures. *By construction*, each of these draws is perfectly consistent with all the observed data and with all the specified restrictions. As noted by Baumeister and Hamilton (2015, 2018), unless the researcher has some prior information *in addition* to the restrictions themselves – and KM did not present their results as having relied on any such information – then all that the researcher has any basis reporting is what is referred to in the literature as the identified set, which is the set of all possible admissible values. The identified set by definition is larger than the full set of 16 numbers in the upper left panel. These 16 numbers are just random examples of some of the values that are included in the identified set. If we ran the code again with a larger number of draws or with a different random number seed, we would obtain a different set of retained draws.

[3] Astute readers may note that there are some observations in the lower left panel below -0.8, despite the statement in KM14 on page 462 that they imposed the bound that the demand elasticity had to be above -0.8. It turns out that this bound is not written into the code they posted on the *Journal of Applied Econometrics* data archive, and we used this code verbatim (changing only the supply elasticity bound to 0.155) to generate our Figure 3. Whether the code includes the bound is irrelevant for the upper left panel, because none of the 16 draws that pass all of KM14's other restrictions would have implied a demand elasticity below -0.8.

To plot impulse-response functions in their papers, KM would select one of the retained draws as if it could represent the set of all generated draws. KM12 (page 1179) made this choice based on the draw that implied the largest response of oil price to an oil supply shock:

> For expository purposes we choose the model with the largest response of the real price to oil supply shocks in Figure 1(b).

KM14 (page 464) based the choice on the implied elasticity of oil demand "in use":

> Solid lines indicate the impulse response estimates for the model with an impact price elasticity of oil demand in use closest to the posterior median of that elasticity among the admissible structural models obtained conditional on the least-squares estimate of the reduced-form VAR model.

Zhou (forthcoming) described yet a third way to choose a representative draw, and astonishingly asserted that this is in fact what KM14 did:

> Kilian and Murphy (2014) relied on narrative inequality restrictions for selecting the most credible model among the set of model solutions that satisfy the sign restrictions on the impulse responses (see also Kilian and Lee (2014)). I incorporate these narrative restrictions into the estimation of the model rather than imposing them based on the visual inspection of the historical decomposition of the admissible models as in Kilian and Murphy (2014).

KZ in footnote 1 repeat this reference to "narrative sign restrictions on the historical decomposition of the real price of oil utilized by Kilian and Murphy (2014)." The expression "narrative sign restrictions" appears nowhere in either KM14 or in Kilian and Lee (2014). That expression was first used by Antolín-Díaz and Rubio-Ramírez (2018), and nothing resembling their method was implemented anywhere in KM14. Zhou (forthcoming) and KZ seem to be using the expression to refer to some unspecified procedure for selecting which one of the 16 retained draws is regarded as representative.

We suggest that the most accurate interpretation of the procedure that KM14 used is obtained by running the code that the authors publicly posted in the *Journal of Applied Econometrics* data archive. If the authors today have some different ideas about what the procedure is supposed to be, that has nowhere been formally explained. The publicly posted code has a particular algorithm for selecting the curves to be plotted (namely, the procedure described in the quote from KM14 above). The seed for the random number generator in their posted code is 316. The red dotted lines in our Figure 4 summarize the effects of what KM14 call a speculative demand shock on their measure of real economic activity and on the real price of oil running this code as posted. These reproduce the graphs shown in Figure 1 of KM14. A researcher who ran this code would describe the findings as KM14 did on pages 464-465:

> a positive speculative demand shock is associated with an immediate jump in the real price of oil. The real price response overshoots, before declining gradually. The effects on global real activity and global oil production are largely negative, but small.

We reran the publicly posted code making only one change – we used a seed for the random number generator of 613 instead of 316. The blue solid lines in our Figure 4 show what these imputed effects look like when this different seed for the random number generator is used. A researcher who ran their code using a random number seed of 613 instead of 316 would describe the findings as follows:

> a positive speculative demand shock is associated with an immediate large drop in economic activity and a small positive effect on price.

The red line in the lower right panel of Figure 2 reproduces the BH mixture prior for the supply elasticity from Figure 1. This panel also shows the posterior distribution (blue histogram) that results from using that prior. In this case, the histogram is based on 1 million generated draws (in contrast to the sample sizes of 16 to 1175 in the other three panels). And in contrast to the other three panels, the probability distribution associated with this histogram is a well defined object of interest. The distribution in the lower right panel summarizes uncertainty about parameters after we combine uncertain identifying assumptions with observed data. The uncertainty associated with this distribution accurately reflects both sampling error (we have only observed a finite sample of observed data) and also uncertainty about the identifying assumptions themselves. In this case, the researcher had a strong prior belief that the supply elasticity was below 0.0258, but recognized some possibility that it could be higher. Having seen the data, the researcher significantly revises upward the possibility of a higher supply elasticity.

To summarize, KZ have the facts exactly backward. The reason that the supply elasticity in the KM14 analysis is so low is only because they *forced* this answer on the data. The reason that BH concluded that the elasticity was larger than 0.0258 was not because we forced this answer on the data, but because we *allowed* this answer. Any procedure that *allows* the possibility that the supply elasticity exceeds 0.0258 – whether our Bayesian approach, or, as seen in the first three panels of Figure 2, even the exact algorithm implemented by KM14 – will end up concluding that it is above 0.0258.

4 On the correct characterization and estimation of elasticity.

KM12 and KM14 further made a conceptual error in the way they calculate elasticities, an error that KZ curiously try to turn into a criticism of BH. Their error was in thinking that an elasticity can be inferred from a single column of the matrix of impacts of structural shocks. This would be correct in a 2-variable system but in general is incorrect when there are more than two variables. This error has been repeated by other studies that followed KM's example, including Ludvigson, Ma, and Ng (2017), Antolín-Díaz and Rubio-Ramírez (2018), Herrera and Rangaraju (2019), and Zhou (forthcoming). For this reason, explaining this error may be of broader interest.

The issue can be illustrated using a 3-equation system consisting of the log of quantity (q_t), the log of income (y_t), and the log of price (p_t),

$$q_t = \gamma y_t + \alpha p_t + u_t^s$$

$$y_t = \xi q_t + \psi p_t + u_t^y$$

$$q_t = \delta y_t + \beta p_t + u_t^d.$$

These are interpreted as the supply equation, the income equation, and the demand equation, respectively. The meaning of the parameter β (the demand price elasticity) is the answer to the question: if price were to increase by 1% with income held constant, by how much would quantity demanded change?

How do we estimate such a concept? Suppose for example that we knew the values of $\gamma, \alpha, \xi, \psi$ and that the structural shocks were mutually uncorrelated. Then the supply shock $u_t^s = q_t - \gamma y_t - \alpha p_t$ and income shock $u_t^y = y_t - \xi q_t - \psi p_t$ would be valid instruments for purposes of estimating the parameters of the demand equation, because they are correlated with y_t and p_t but uncorrelated with u_t^d:

$$\begin{bmatrix} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^T u_t^s y_t & \sum_{t=1}^T u_t^s p_t \\ \sum_{t=1}^T u_t^y y_t & \sum_{t=1}^T u_t^y p_t \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T u_t^s q_t \\ \sum_{t=1}^T u_t^y q_t \end{bmatrix}. \qquad (1)$$

Alternatively, we could estimate δ and β by maximizing the likelihood function conditional on u_t^s and u_t^y. For this example, the conditional MLE is numerically identical to the IV estimates; see the appendix for mathematical details.

More generally, using the likelihood function of the observed vector of data $\mathbf{y}_t$ along with any additional information about structural parameters is the optimal way to form inference about the structural parameters. From a frequentist perspective, when the additional information takes the form of complete identifying assumptions, maximum likelihood estimation is optimal in the sense that it achieves the smallest asymptotic variance. From a Bayesian perspective, when the additional information takes the form of Bayesian prior distributions, Bayesian inference is optimal in the sense of minimizing posterior expected loss.

Now consider the relation between the parameter β and the observed impacts of structural shocks. The above structural model can be written $\mathbf{A}\mathbf{y}_t = \mathbf{u}_t$, with the impacts of the structural shocks on the observed variables captured by the matrix

$$\mathbf{H} = \frac{\partial \mathbf{y}_t}{\partial \mathbf{u}_t'} = \mathbf{A}^{-1} \tag{2}$$
$$= |\mathbf{A}|^{-1} \begin{bmatrix} -\beta - \delta\psi & \alpha\delta - \beta\gamma & \alpha + \gamma\psi \\ -\psi - \beta\xi & \alpha - \beta & \psi + \alpha\xi \\ \delta\xi - 1 & \delta - \gamma & 1 - \gamma\xi \end{bmatrix}.$$

What would we get if we tried to estimate the demand elasticity on the basis of the ratio of the change in q_t to the change in p_t in response to a shock to supply u_t^s? For this system that ratio is given by

$$\frac{h_{11}}{h_{31}} = \frac{-\beta - \delta\psi}{\delta\xi - 1}. \tag{3}$$

In general, expression (3) is not the demand elasticity β. The reason is that if there is a shock to u_t^s, not only will it change the price p_t, but it will also change income. The size of the change in price is $|\mathbf{A}|^{-1}(\delta\xi - 1)$ and the size of the change in income is $|\mathbf{A}|^{-1}(-\psi - \beta\xi)$. From the demand curve, the change in price will lead to a change in quantity demanded of β times the change in price, namely $\beta|\mathbf{A}|^{-1}(\delta\xi - 1)$. Likewise the change in income will lead to a change in quantity demanded of δ times the change in income, namely $\delta|\mathbf{A}|^{-1}(-\psi - \beta\xi)$. The observed change in quantity demanded in response to the shock in supply is the sum of these two terms,

$$\underbrace{\beta}_{\text{response to price}} \underbrace{|\mathbf{A}|^{-1}(\delta\xi - 1)}_{\text{change in price}} + \underbrace{\delta}_{\text{response to income}} \underbrace{|\mathbf{A}|^{-1}(-\psi - \beta\xi)}_{\text{change in income}} = \underbrace{|\mathbf{A}|^{-1}(-\beta - \delta\psi)}_{\text{total change}}.$$

Dividing this by the magnitude of the change in price that results from the supply shock, $|\mathbf{A}|^{-1}(\delta\xi - 1)$, produces the result (3).

In the special case when demand does not respond to income ($\delta = 0$), expression (3) would simplify to the correct answer β. But in general, expression (3) reflects a combination of the sensitivity of demand to price and the sensitivity of demand to income.

When KM14 calculated the demand elasticity in their 4-variable model they used an incorrect expression like (3). Interestingly, to calculate the supply elasticity they used two different expressions, both of which are incorrect. Their code first calculates h_{12}/h_{32}, the ratio of the first to the third elements of the second column of $\mathbf{A}^{-1}$. This is the ratio of the change in quantity to the change in price in response to what they call a flow demand shock. The code next calculates h_{13}/h_{33}, the ratio of the first and third elements of the third column of $\mathbf{A}^{-1}$. This is the ratio of the change in quantity to the change in price in response to what they call a speculative demand shock. According to KM14's understanding, these objects could both be estimates of what they call the "supply elasticity." In practice they will be two different numbers. The way their code works is to reject a draw unless both of these proposed measures of the supply elasticity are below 0.0258.

In the correct way to use a structural model, there is a single unique magnitude implied by $\mathbf{H}$ that should be called the supply elasticity, and this is the (1,3) element of $\mathbf{H}^{-1}$.

KZ add more confusion to this issue by suggesting that the existence of inventories adds further complications to estimating elasticities that are not taken into account by BH. They claim for example on page 5: "BH use an incorrect definition of the oil demand elasticity that ignores the storability of crude oil." One question they seem to be raising here is whether or not we want to include the response of inventory demand to a change in price. KM14 prefer to use a measure that excludes this response, which they call the elasticity of demand in use. They define this concept on page 478:

> A more appropriate definition of the price elasticity of oil demand for policy questions is the elasticity in use. The latter demand elasticity is based on the change in the use of oil, defined as the sum of the change in oil production and of the depletion of oil inventories.

Let Q_t denote the level of oil production in month t and I_t^* the true level of inventories. KM14 would then define the level of oil consumed as $U_t = Q_t - \Delta I_t^*$. Consider equation (29) in BH:

$$q_t = \beta_{qy} y_t + \beta_{qp} p_t + \Delta i_t^* + \mathbf{b}_3' \mathbf{x}_{t-1} + u_{3t}^*. \tag{4}$$

Using the BH definitions $q_t = \log(Q_t) - \log(Q_{t-1})$ and $\Delta i_t^* = \Delta I_t^* / Q_{t-1}$, this equation can be rewritten

$$\log Q_t - \log Q_{t-1} = \beta_{qy} y_t + \beta_{qp} p_t + \Delta I_t^* / Q_{t-1} + \mathbf{b}_3' \mathbf{x}_{t-1} + u_{3t}^*.$$

Given the small size of monthly production changes, the left-hand side is very well approximated by $(Q_t - Q_{t-1})/Q_{t-1}$. Rearranging gives

$$U_t / Q_{t-1} - Q_{t-1} / Q_{t-1} = \beta_{qy} y_t + \beta_{qp} p_t + \mathbf{b}_3' \mathbf{x}_{t-1} + u_{3t}^*.$$

Since the monthly change in inventories ΔI_t^* is on average a small fraction of total monthly production, the left-hand side of this equation is in turn well approximated by

$$U_t / Q_{t-1} - Q_{t-1} / Q_{t-1} \simeq (U_t - U_{t-1}) / U_{t-1} \simeq u_t$$

for $u_t = \log(U_t) - \log(U_{t-1})$ the growth rate of oil in use as defined by KM14. Thus the BH equation (4) is an excellent approximation to

$$u_t = \beta_{qy} y_t + \beta_{qp} p_t + \mathbf{b}_3' \mathbf{x}_{t-1} + u_{3t}^*.$$

The parameter β_{qp}, for which our approach provides the optimal inference, is thus the magnitude in which KM14 claimed we should be interested, namely, the price elasticity of oil demand in use.

By contrast, KM14 calculated their estimate of this magnitude for month t as

$$\eta_t^{Use} = h_{31}^{-1} \frac{h_{11} Q_{t-1} - h_{41}}{Q_{t-1} - \overline{\Delta I}}.$$

where $(h_{11}, h_{21}, h_{31}, h_{41})'$ is the first column of $\mathbf{A}^{-1}$ (that is, the effect on impact of a shock to oil supply on production, economic activity, price, and inventories). They then averaged η_t^{Use} over all months t. This estimate is subject to the problems we noted above with claiming to infer a structural parameter from a single column of $\mathbf{A}^{-1}$, along with many other shortcomings. How could anyone claim that this is a better measure than the one we used? KZ nevertheless summarize this issue with their usual colorful language on page 6: "The only reason for BH to insist on using the wrong definition of the oil demand elasticity is that their econometric method... cannot be applied to oil market models subject to restrictions on the correctly defined elasticity." Had we for some reason wanted to incorporate prior information about certain ratios of elements of $\mathbf{A}^{-1}$, we would have done so using exactly the method that we detailed on pages 1894-1895.

Even more odd is KZ's insistence that one cannot talk about demand elasticity at all in a model that does not explicitly include inventories. They claim on page 6 that "the coefficients in the 3-variable models of Kilian (2009) and Kilian and Murphy (2012) are complicated transformations of the structural coefficients in the 4-variable Kilian and Murphy (2014) model, from which the oil demand elasticity value cannot be extracted." Consider the inventory demand equation (30) in BH:

$$\Delta i_t^* = \psi_1^* q_t + \psi_2^* y_t + \psi_3^* p_t + \mathbf{b}_4^{*\prime} \mathbf{x}_{t-1} + u_{4t}^*. \tag{5}$$

Substituting equation (5) into equation (4) results in

$$q_t = \beta_{qy} y_t + \beta_{qp} p_t + \psi_1^* q_t + \psi_2^* y_t + \psi_3^* p_t + \mathbf{b}_4^{*\prime} \mathbf{x}_{t-1} + u_{4t}^* + \mathbf{b}_3^{\prime} \mathbf{x}_{t-1} + u_{3t}^*.$$

Rearranging,

$$q_t = \tilde{\beta}_{qy} y_t + \tilde{\beta}_{qp} p_t + \tilde{\mathbf{b}}_4^{\prime} \mathbf{x}_{t-1} + \tilde{u}_{3t}.$$

where $\tilde{\beta}_{qp} = (\beta_{qp} + \psi_3^*)/(1 - \psi_1^*)$. This is exactly the form of the demand equation in a 3-variable system that excludes inventories. The meaning of the elasticity $\tilde{\beta}_{qp}$ in this equation is

$$\tilde{\beta}_{qp} = (1 - \psi_1^*)^{-1} \left[\frac{\partial (q_t - \Delta i_t^*)}{\partial p_t} + \frac{\partial \Delta i_t^*}{\partial p_t} \right].$$

Thus $\tilde{\beta}_{qp}$ is the price elasticity of the total demand for oil (consumption plus inventories). It represents a combined response of "demand in use" $(Q_t - \Delta I_t^*)$, inventory demand (ΔI_t^*), and any potential multiplier effect coming from a response ψ_1^* of inventory demand to total oil consumption. BH noted that a Bayesian interpretation of KM12's analysis would imply a 60% probability that $\tilde{\beta}_{qp}$ is less than -2, i.e., that a 10% increase in price results in a 20% drop in total demand for oil within the month. A Bayesian interpretation of Kilian's (2009) analysis implies a nonneglible posterior probability that a 10% increase in price results in a 50% *increase* in the total demand for oil within the month. Our conclusion from this exercise was that we should relax the strong restrictions about supply that were used by Kilian (2009) and KM12 to achieve identification, but supplement this weaker information with additional weak information about other aspects of the model.

Nothing in KZ undermines that conclusion in the slightest.

5 Discussion of other inaccuracies in Kilian and Zhou (2019).

We have documented that KZ missed the main point of BH. KZ are moreover inaccurate in their discussion of virtually every detail. Here we highlight a number of examples.

5.1 Mischaracterizations about the flexibility of BH.

KZ claim on page 5:

> applying the BH methodology to state-of-the-art oil market models ... is impossible because imposing priors on the price elasticity of oil demand requires simultaneous restrictions on multiple elements of B_0^{-1} [$\mathbf{A}^{-1}$ in the notation of BH]. This violates the default assumption of prior independence across the elements of B_0 in BH's analysis.

In point of fact, prior independence across the elements of $\mathbf{A}$ is *not* a default assumption and is in no way required by our analysis. Quoting from Baumeister and Hamilton (2018, page 48), "the resulting prior $p(\mathbf{A})$ is no longer independent across the individual elements of $\mathbf{A}$, but includes some joint information about their interaction." Simultaneous restrictions on multiple elements of $\mathbf{A}^{-1}$ feature prominently in the baseline analysis of both BH (equations (41) and (42)) and Baumeister and Hamilton (2018, equations (27)-(30)).

KZ further claim on page 3:

> Bayesian priors used in estimating sign-identified models ... may be inadvertently informative about B_0^{-1}.... BH's response to this concern is to recommend that researchers instead impose explicit priors for the elements of B_0 with the diagonal elements normalized to one.

No. Our response to this concern is to recommend that researchers use information about both $\mathbf{A}$ and $\mathbf{A}^{-1}$ to inform a fully specified structural model, as we did in both BH and Baumeister and Hamilton (2018).

5.2 Mischaracterizations about the motivations of BH.

A commonly repeated tactic of KZ is to make up an alleged motive or hidden agenda behind some detail in BH and then try to impugn our analysis on the basis of this invented motivation. For example, they write on page 11:

> their results are not driven by the use of an informative prior ... but by shifting this prior in the direction of unrealistically large supply elasticity values. Why would BH do that? Hamilton for many years has been advocating the position that oil supply shocks have been the main driver of oil price fluctuations.... Thus, it stands to reason that BH understood the importance of increasing the value of this elasticity in their effort to undermine existing evidence that oil demand shocks are the main driver of oil price fluctuations. They accomplished this objective mainly by imposing prior information on the distribution of the prior probability mass that reflects their personal beliefs rather than extraneous information.

Well, for starters, whether we use the prior distribution that is plotted as the dashed black line in Figure 1 or the prior distribution that is plotted as the solid red line has zero effect on the portion of price movements accounted for by supply shocks, as seen in rows 6-9 of Table 1. Whatever our reason for proposing the dashed black line in Figure 1 as a reasonable prior, it can not be because we wanted to "undermine existing evidence that oil demand shocks are the main driver of oil price fluctuations." We are moreover unaware of any monotonic mapping, such as the one implied by this quote from KZ, between specified elements of the prior and various objects of potential interest to researchers in complicated models like this.

Another example of this unfortunate debate tactic appears on page 2: "their choice of econometric method ... is designed to inflate the effects of oil supply shocks and to reduce the effect of oil demand shocks." KZ are quite mistaken in the objective they attribute to us in writing the paper. Our primary purpose in writing the paper was to discuss structural interpretation of VARs with incomplete identification. We used oil supply and demand as a way to illustrate this theme, not as the primary reason for writing the paper. That is why we chose the title for BH that we did, why we wrote the abstract to BH that we did, and why we chose the title for the current paper that we did. BH was primarily about structural inference with incomplete identification. The initial response of the editor was that the paper had too much methodology, and we should highlight more the differences between our empirical results and some other results in the literature. In response to this request, we added and emphasized empirical results more in the version that was ultimately published. Much of what KZ suggest was our "real reason" for writing the paper did not even appear in the first draft!

5.3 Mischaracterizations of the literature on the supply elasticity.

Our understanding of the results of several papers on supply elasticity differs from that in KZ. They offer on page 15 the following interpretation of the conclusions of Bjørnland, Nordvik and Rohrer (2019): "It should be noted that in the May 2019 version of the Bjørnland et al. study the corresponding supply elasticity for conventional oil is only 0.03 (and again not statistically significant), while the supply elasticity for shale oil is -0.12 (which actually is the wrong sign)." These words from KZ are intended to summarize the results from the following regression (standard errors in parentheses):

$$\Delta q_{it} = \underset{(0.005)}{-0.36}\Delta q_{i,t-1} + \underset{(0.01)}{0.03}\Delta q_{i,t-1}s_i + \underset{(0.05)}{0.03}\Delta p_t - \underset{(0.05)}{0.15}\Delta p_t s_i$$
$$+ \underset{(0.20)}{0.07}(\Delta p_t - \Delta f_{t,t+3}) + \underset{(0.27)}{0.76}(\Delta p_t - \Delta f_{t,t+3})s_i + X_t + \lambda_y + \mu_i + \rho_{i,t} + e_{it}.$$

Here q_{it} is the log of production from well i in month t, p_t is the log of the price of crude oil in month t, $s_i = 1$ if well i is in the shale region and 0 otherwise, $\Delta f_{t,t+3} = \log(F_{t,t+3}) - \log(F_{t-1,t+2})$ is the percent change in the price of a 3-month-ahead futures contract in month t, X_t captures effects of a set of macroeconomic control variables, λ_y captures year fixed effects, μ_i captures well fixed effects, and $\rho_{i,t}$ captures fixed effects based on the age of well i in month t. When KZ say that this regression implies a supply elasticity of 0.03 they are looking only at the coefficient on Δp_t. This corresponds to the answer to the question, how much does production of a conventional well change in response to an increase in the current price that is exactly matched by an increase in the futures price (in other words, assuming that $\Delta p_t - \Delta f_{t,t+3} = 0$)? But the striking feature of this regression, and indeed the focus of Bjørnland et al.'s paper, is the large coefficient on $\Delta p_t - \Delta f_{t,t+3}$. This large and dramatic response to a temporary change in price, especially for shale wells, is a direct challenge to the conclusion of Anderson, Kellogg, and Salant (2018). Here is how Bjørnland and coauthors summarized their findings in the abstract to their paper: "While output from conventional wells appear non-responsive to price fluctuations in the short-term, we find supply elasticity to be positive and in the range of 0.3-0.9 for shale oil wells, depending on wells and firms characteristics."

KZ dismiss Bjørnland et al.'s findings on the grounds that they only refer to North Dakota, and likewise dismiss the observation of BH that production of Saudi Arabia has exhibited huge, rapid adjustments to changing demand conditions on the grounds that this refers only to Saudi Arabia. KZ instead highlight the findings of Anderson, Kellogg and Salant (2018), even though this study used only wells in Texas. In the regression reported in the online appendix to Anderson et al., the coefficient relating the change in the log of production of Texas well i in month t to the month t change in the log of the front-month futures price is 0.083 with a standard error of 0.036. Thus if KZ had reported the results of Anderson et al. following the same principle they used to summarize the Bjørnland et al. findings, they would have said that the Anderson et al. estimate of the elasticity is 0.083, not zero. It is only when Anderson et al. added together the coefficient on the current month's price change with the coefficient on the previous month's price change that they arrived at a coefficient near zero (namely 0.0009 with a standard error of 0.034).

In our view, the best estimate of the *global* within-month supply elasticity is the Caldara, Cavallo, and Iacoviello (2019) instrumental-variables estimate of 0.081 with a standard error of 0.037 that we described in Section 3. Their proposal was that exogenous shifts in supply of countries other than Saudi Arabia – for example, hurricanes in the Gulf of Mexico or a strike by Norwegian oil workers – could be used as an instrument for how Saudi Arabia responds to changes in price. Aggregating across countries, they looked at the combined response of all countries not directly involved in the shock to the price change associated with the shock. KZ dismiss their evidence on page 14 with the assertion that "we need an exogenous shift in the oil demand curve to identify the oil supply elasticity, not an exogenous shift in oil supply." This makes no sense. From the perspective of an individual producer, the incentives are the same whether the price has increased because of higher demand or because of reduced supply somewhere else. Indeed, the CCI approach is just a generalization and extension of the method by which KM12 obtained their estimate of 0.0258 based on the response of countries other than Iraq and Kuwait to the oil price increase in August 1990 when Iraq invaded Kuwait. The main difference is that CCI want to perform this IV regression with a sample size of 29 observations while KM12 do the regression with a sample size of one.

Note moreover that CCI found that most of this supply response in these episodes comes from OPEC, and especially Saudi Arabia. CCI's estimates are in fact perfectly consistent with the conclusion of Anderson et al. (2019) that the short-run response of U.S. supply to the current price may be more modest.

Even more fundamentally, KZ's entire discussion of the literature on the supply elasticity avoids addressing the core question. The question is not, "what is the best estimate of the short-run supply elasticity?" Rather, the question they need to answer in order to defend a prior of the form used by KM12 and KM14 is, "what is the value of x such that we are 100% certain the short-run elasticity could be no larger than x?" They do not emerge from their discussion of the literature with a proposed value for x to replace KM's insistence that $x = 0.0258$. The reason is that the literature is not amenable to summarizing in terms of any particular value for x. KZ mention standard errors of a particular study only when the standard error suggests the value could be zero, without noting that the same logic implies the elasticity could also be larger than x. Perhaps the best they could defend is the position on page 18 that we should use a "prior distribution for the supply elasticity that reaches its maximum at zero." Note that this is exactly what the solid red curve in Figure 1 does.

5.4 Mischaracterizations of the literature on measuring global economic activity.

KZ claim on page 7 "BH substitute an arguably inferior measure of global real activity for the standard measure based on Kilian (2009)." Hamilton (forthcoming) outlined a number of reasons why the Kilian measure should not be used. KZ respond only to one of these in their footnote 3, where they note that Kilian (2019) has now corrected the mathematical error that was discovered by Hamilton (forthcoming) in the index that was used by Kilian and Zhou (2018). Nowhere do KZ address the many other concerns raised by Hamilton (forthcoming). (1) Both the original index used by Kilian and Zhou (2018) and the corrected index proposed by Kilian (2019) imply that the cyclical component of world real economic activity reached a lower point in 2016 than during the financial crisis or for that matter any historical recession. (2) The cyclical component of BH's measure of global economic activity has a correlation with annual real GDP growth of 0.88, whereas the measures used by Kilian and Zhou (2018) and Kilian (2019) have a correlation of 0.10, not statistically significantly different from zero. (3) The Kilian and Zhou (2018) measure does not make a statistically significant contribution to forecasting the price of any commodity. The Kilian (2019) measure does only slightly better. By contrast, the BH measure is significantly helpful for forecasting almost every commodity. (4) The Kilian and Zhou (2018) and Kilian (2019) measures assume that there is a linear time trend in their respective measures of real shipping costs. This assumption is inconsistent with the observed data based on the tests reported in Hamilton (forthcoming).

On page 8, KZ criticize BH for addressing measurement error in inventories but not in world economic activity with their usual tactic: "The apparent reason that BH choose to focus on measurement error in oil inventories exclusively is that they seek to lower the role for storage demand shocks and increase the role of oil supply shocks in their model." We note that KM14 have a section titled "How Accurate Are the Oil Inventory Data?" but no section titled "How Accurate Is the Proxy for Real Economic Activity?". For our part, BH explained on page 1888 why we focused on the quality of the inventory proxy:

> (i) there are no data on OECD crude oil inventories, and so the series is extrapolated from OECD petroleum product inventories; (ii) there are no data even for OECD product inventories before 1988, requiring numbers for this earlier period to be further extrapolated from the growth rate of U.S. petroleum product inventories; (iii) OECD petroleum product consumption only accounts for 60% of world petroleum product consumption on average over 1992-2015, so even if we had an

accurate measure of OECD crude inventories, it likely represents little more than half of the world total.

KZ further criticize the absence of serial correlation in our representation of measurement error, stating on page 8, "BH take the view that any arbitrary measurement error specification is better than none. It is not clear what the basis of that view is." The basis for this view is that the absence of measurement error is a special case of the model we estimate, the special case being when the variance of the measurement error is zero.

5.5 Mischaracterizations about the loss function.

KZ write on pages 4 and 5 that the Bayesian inference in BH is "optimal only conditional on their particular choice of the loss function.... If we do not accept that loss function, the median response functions reported by BH are economically and statistically meaningless." The loss function we use is quite standard. But the logical jump from "not optimal" to "meaningless" leaps over quite a chasm. For example, with a loss function based on minimizing the absolute value of deviations (the one we favor and that in fact is conventionally employed), the optimal estimate is the posterior median. On the other hand, if the loss function is based on minimizing a quadratic function defined *jointly* over the full vector of impulse responses, the optimal estimate is the pointwise posterior mean (see Baumeister and Hamilton, 2018, Section 2.1). But even if one preferred the latter loss function, one would never say that the posterior median is "meaningless." Indeed, in practice the two estimators, the median and the mean, are typically quite close to each other.

KZ pick up this theme again on page 9: "The Kilian and Murphy model is evaluated using the econometric methodology of Inoue and Kilian (2013, 2018)." Inoue and Kilian never wrote down an explicit loss function that would motivate their procedure. The implicit loss function that would motivate their approach assumes zero loss if one gets *every detail* of the model exactly right and unit loss otherwise, regardless of how close or far from the truth the estimate is. Baumeister and Hamilton (2018) discussed why we prefer our approach to that in Inoue and Kilian. These issues aside, note that Inoue and Kilian wrote on page 11, "Our approach in this paper is explicitly Bayesian in nature." To calculate a Bayesian posterior distribution you need to start with a Bayesian prior distribution. But a Bayesian prior distribution over the structural parameters is nowhere even mentioned, let alone defended, in either KM14 or Zhou (forthcoming). As we noted in Section 3, in the absence of such a prior distribution, there is no meaning to the probability distributions represented by the histograms in the first three panels of Figure 2, and no justification for applying the Inoue-Kilian procedure, or any other loss function, to these histograms as if they represented a Bayesian posterior distribution. Moreover, it is immediately clear from Figure 4 that there exists *no loss function whatever* according to which the code posted by KM14 could be claimed to have produced an optimal inference. If the red line in Figure 4 is "optimal," the blue line clearly can not be.

5.6 Other mischaracterizations of BH.

There are many other inaccuracies in KZ. We mention a few of these now.

KZ describe our approach on page 1 as "standard Bayesian estimation methods for structural VAR models (see Sims and Zha 1998)." Sims and Zha were proposing a method for models that are completely identified. By contrast, our proposal is to use Bayesian priors in place of conventional identifying assumptions. Such a suggestion is nowhere to be found in Sims and Zha.

Also on page 1 KZ claim "They argue that this approach nests existing Bayesian methods for structural VAR analysis as a special case." No. What we claimed (and demonstrated) was that our approach nests existing *frequentist* approaches to identification as special cases. That is why the two examples we chose to illustrate this point, namely Kilian (2009) and KM12, were frequentist, not Bayesian analyses.

Next on page 1 KZ write, "BH are also incorrect in asserting that earlier studies did not impose all relevant identifying information." Nowhere in BH will one find us suggesting researchers should have imposed more identifying information. That is the exact opposite of our message. Instead, our message was that identifying assumptions can be relaxed and represented as imperfect identifying information, and that once one does so, there are benefits to incorporating imperfect identifying information from a variety of sources to assist the inference.

On page 9 KZ write, "they ignore the existence of a well-documented structural break in these data in late 1973." Quite the opposite: BH devoted four full pages to this question. See pages 1896-1898 and 1906.

6 Conclusion.

Kilian and Zhou (2019) misunderstood the main contribution of Baumeister and Hamilton (2019), which was to develop a flexible empirical framework that nests frequentist identification strategies as a special case and that lends itself to applications in many contexts besides the oil market. They try hard to argue that Baumeister and Hamilton (2019) did not offer a valid or useful approach to analyzing the global oil market. Every one of their criticisms of BH is without merit.

References

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

Antolín-Díaz, Juan, and Juan F. Rubio-Ramírez (2018). "Narrative Sign Restrictions for SVARs," *American Economic Review* 108: 2802-2829.

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

Baumeister, Christiane, and James D. Hamilton (2017). "Structural Interpretation of Vector Autoregressions with Incomplete Information: Revisiting the Role of Oil Supply and Demand Shocks," NBER Working Paper 24167.

Baumeister, Christiane, and James D. Hamilton (2018). "Inference in Structural Vector Autoregressions when the Identifying Assumptions Are Not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations," *Journal of Monetary Economics* 100: 48-65.

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

Bjørnland, Hilde C., Frode Martin Nordvik, and Maximilian Rohrer (2019). "Supply Flexibility in the Shale Patch: Evidence from North Dakota," working paper, Norwegian Business School.

Caldara, Dario, Michele Cavallo, and Matteo Iacoviello (2019). "Oil Price Elasticities and Oil Price Fluctuations," *Journal of Monetary Economics* 103: 1-20.4

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

Herrera, Ana Maria, and Sandeep Rangaraju (2019). "The Effect of Oil Supply Shocks on US Economic Activity: What Have We Learned?" *Journal of Applied Econometrics,* forthcoming.

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

Kilian, Lutz, and Thomas 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, Lutz, and Daniel 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(5): 1166-1188.

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

Kilian, Lutz, and Xiaoqing Zhou (2019). "Oil Supply Shock Redux?" Working paper.

Ludvigson, Sydney C., Sai Ma, and Serena Ng (2017). "Shock Restricted Structural Vector-Autoregressions," NBER Working Paper 23225.

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

Appendix: The relation between IV and MLE in a simple 3-variable example.

KM14 and KZ, along with Ludvigson, Ma, and Ng (2017), Antolín-Díaz and Rubio-Ramírez (2018), Herrera and Rangaraju (2019), and Zhou (forthcoming), proposed to estimate elasticities using the responses to specified structural shocks. We formalized the intuition behind this idea in equation (1) which characterized the instrumental-variables estimates of elasticities in a simple 3-variable example. IV is closely related to (and in general, is less efficient than) full-information maximum likelihood, which is the frequentist analog to Bayesian inference about structural parameters using the likelihood function. In this appendix we show that, for this simple 3-variable example, IV turns out to be identical to maximum likelihood. Our purpose in doing so is to help explain why analysis based on the likelihood function implied by the structural model (whether frequentist or Bayesian) is the correct and indeed the optimal way to estimate elasticities.

Characterization of IV estimate.

Expression (1) can be rewritten

$$\begin{bmatrix} \sum_{t=1}^T u_t^s y_t & \sum_{t=1}^T u_t^s p_t \\ \sum_{t=1}^T u_t^y y_t & \sum_{t=1}^T u_t^y p_t \end{bmatrix} \begin{bmatrix} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^T u_t^s q_t \\ \sum_{t=1}^T u_t^y q_t \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{bmatrix} \begin{bmatrix} \sum_{t=1}^T q_t y_t & \sum_{t=1}^T q_t p_t \\ \sum_{t=1}^T y_t y_t & \sum_{t=1}^T y_t p_t \\ \sum_{t=1}^T p_t y_t & \sum_{t=1}^T p_t p_t \end{bmatrix} \begin{bmatrix} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{bmatrix} \begin{bmatrix} \sum_{t=1}^T q_t q_t \\ \sum_{t=1}^T y_t q_t \\ \sum_{t=1}^T p_t q_t \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{bmatrix} \begin{bmatrix} \sum_{t=1}^T q_t q_t & \sum_{t=1}^T q_t y_t & \sum_{t=1}^T q_t p_t \\ \sum_{t=1}^T y_t q_t & \sum_{t=1}^T y_t y_t & \sum_{t=1}^T y_t p_t \\ \sum_{t=1}^T p_t q_t & \sum_{t=1}^T p_t y_t & \sum_{t=1}^T p_t p_t \end{bmatrix} \begin{bmatrix} 1 \\ -\hat{\delta}_{IV} \\ -\hat{\beta}_{IV} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{6}$$

We can partition $\mathbf{A}$ as

$$\underset{(3\times 3)}{\mathbf{A}} = \begin{bmatrix} \underset{(2\times 3)}{\mathbf{\Gamma}} \\ \underset{(1\times 3)}{\boldsymbol{\eta}'} \end{bmatrix}$$

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{bmatrix}$$

$$\boldsymbol{\eta}' = \begin{bmatrix} 1 & -\delta & -\beta \end{bmatrix}.$$

This allows (6) to be written compactly as

$$\boldsymbol{\Gamma}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\eta}}_{IV} = \mathbf{0} \tag{7}$$

for $\hat{\boldsymbol{\Omega}}$ the sample variance-covariance matrix of the observed data:

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} T^{-1}\sum_{t=1}^{T} q_t q_t & T^{-1}\sum_{t=1}^{T} q_t y_t & T^{-1}\sum_{t=1}^{T} q_t p_t \\ T^{-1}\sum_{t=1}^{T} y_t q_t & T^{-1}\sum_{t=1}^{T} y_t y_t & T^{-1}\sum_{t=1}^{T} y_t p_t \\ T^{-1}\sum_{t=1}^{T} p_t q_t & T^{-1}\sum_{t=1}^{T} p_t y_t & T^{-1}\sum_{t=1}^{T} p_t p_t \end{bmatrix}.$$

Characterization of MLE.

Next consider maximum likelihood estimation. Conditional on the (2×1) vector $\mathbf{z}_t = \boldsymbol{\Gamma}\mathbf{y}_t$, the remaining randomness in $\mathbf{y}_t$ can be summarized in terms of the observed scalar $w_t = \boldsymbol{\gamma}_\perp'\mathbf{y}_t$ where $\boldsymbol{\gamma}_\perp$ is the (3×1) vector that is orthogonal to the rows of $\boldsymbol{\Gamma}$ and whose third element is normalized to be unity.[4] If $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{D})$, then $\mathbf{y}_t$ is multivariate Normal and $w_t|\mathbf{z}_t \sim N(\boldsymbol{\pi}(\delta,\beta)'\mathbf{z}_t, v)$ where given $\boldsymbol{\Gamma}$, the (2×1) vector $\boldsymbol{\pi}$ is a known function of (δ,β). To calculate this function, notice that $\mathbf{y}_t = \mathbf{A}^{-1}\mathbf{u}_t$ so $w_t = \boldsymbol{\gamma}_\perp'\mathbf{y}_t = \boldsymbol{\gamma}_\perp'\mathbf{A}^{-1}\mathbf{u}_t$. Since the elements of $\mathbf{u}_t$ are mutually independent, the coefficients in the population projection of w_t on (u_t^s, u_t^y) are given by the first two terms of the vector $\boldsymbol{\gamma}_\perp'\mathbf{A}^{-1}$:

$$\boldsymbol{\pi}(\delta,\beta)' = \boldsymbol{\gamma}_\perp'\mathbf{A}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{8}$$

Given $\boldsymbol{\Gamma}$, this is a known function of (δ,β).

The conditional MLE of (δ,β) is the value that maximizes

$$-(T/2)\log(2\pi) - (T/2)\log v - \frac{\sum_{t=1}^{T}[w_t - \boldsymbol{\pi}(\delta,\beta)'\mathbf{z}_t]^2}{2v}$$

or equivalently the value that minimizes $\sum_{t=1}^{T}[w_t - \boldsymbol{\pi}(\delta,\beta)'\mathbf{z}_t]^2$. Note that in general the value of $\boldsymbol{\pi}$ that minimizes this sum of squared residuals is the OLS estimate

$$\hat{\boldsymbol{\pi}}_{OLS}' = \left(\sum_{t=1}^{T} w_t\mathbf{z}_t'\right)\left(\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'\right)^{-1}. \tag{9}$$

[4] That is,

$$\begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \\ 0 & 0 & 1 \end{bmatrix}\boldsymbol{\gamma}_\perp = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\boldsymbol{\gamma}_\perp = \begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \\ 0 & 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Note that $\boldsymbol{\gamma}_\perp$ is known from $\boldsymbol{\Gamma}$ and does not depend on (δ,β).

We claim that if we choose $(\hat{\delta}_{MLE}, \hat{\beta}_{MLE})$ such that

$$\hat{\boldsymbol{\eta}}'_{MLE}\left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{z}'_t\right)\left(\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}'_t\right)^{-1} = \mathbf{0}', \tag{10}$$

then (8) will satisfied:

$$\hat{\boldsymbol{\pi}}'_{OLS} = \boldsymbol{\gamma}'_{\perp}\begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \\ 1 & -\hat{\delta}_{MLE} & -\hat{\beta}_{MLE} \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{11}$$

To verify (11), first define the (3×2) matrix

$$\begin{aligned}\hat{\boldsymbol{\Pi}} &= \left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{z}'_t\right)\left(\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}'_t\right)^{-1} \\ &= \left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}'_t\boldsymbol{\Gamma}'\right)\left(\sum_{t=1}^{T}\boldsymbol{\Gamma}\mathbf{y}_t\mathbf{y}'_t\boldsymbol{\Gamma}'\right)^{-1}\end{aligned} \tag{12}$$

so that for example

$$\hat{\boldsymbol{\pi}}'_{OLS} = \boldsymbol{\gamma}'_{\perp}\hat{\boldsymbol{\Pi}}. \tag{13}$$

Premultiply (12) by $\hat{\mathbf{A}}_{MLE}$:

$$\begin{bmatrix} \boldsymbol{\Gamma} \\ \hat{\boldsymbol{\eta}}'_{MLE} \end{bmatrix}\hat{\boldsymbol{\Pi}} = \begin{bmatrix} \boldsymbol{\Gamma}\left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}'_t\boldsymbol{\Gamma}'\right)\left(\sum_{t=1}^{T}\boldsymbol{\Gamma}\mathbf{y}_t\mathbf{y}'_t\boldsymbol{\Gamma}'\right)^{-1} \\ \hat{\boldsymbol{\eta}}'_{MLE}\left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}'_t\boldsymbol{\Gamma}'\right)\left(\sum_{t=1}^{T}\boldsymbol{\Gamma}\mathbf{y}_t\mathbf{y}'_t\boldsymbol{\Gamma}'\right)^{-1} \end{bmatrix}. \tag{14}$$

The first two rows of (14) will be recognized as the (2×2) identity matrix, and the third row is zero by the proposed choice for $\hat{\boldsymbol{\eta}}_{MLE}$. Thus

$$\hat{\mathbf{A}}_{MLE}\hat{\boldsymbol{\Pi}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Premultiplying by $\boldsymbol{\gamma}'_{\perp}\hat{\mathbf{A}}^{-1}_{MLE}$ gives

$$\boldsymbol{\gamma}'_{\perp}\hat{\boldsymbol{\Pi}} = \boldsymbol{\gamma}'_{\perp}\hat{\mathbf{A}}^{-1}_{MLE}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

or using (13),

$$\hat{\boldsymbol{\pi}}'_{OLS} = \boldsymbol{\gamma}'_{\perp}\hat{\mathbf{A}}^{-1}_{MLE}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

as claimed in (11).

Demonstration of equivalence.

Note that since $\left(\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'\right)$ has full rank, condition (10) can equivalently be written

$$\begin{aligned}\mathbf{0}' &= \hat{\boldsymbol{\eta}}_{MLE}'\left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{z}_t'\right)\\ &= \hat{\boldsymbol{\eta}}_{MLE}'\left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right)\boldsymbol{\Gamma}'.\end{aligned}$$

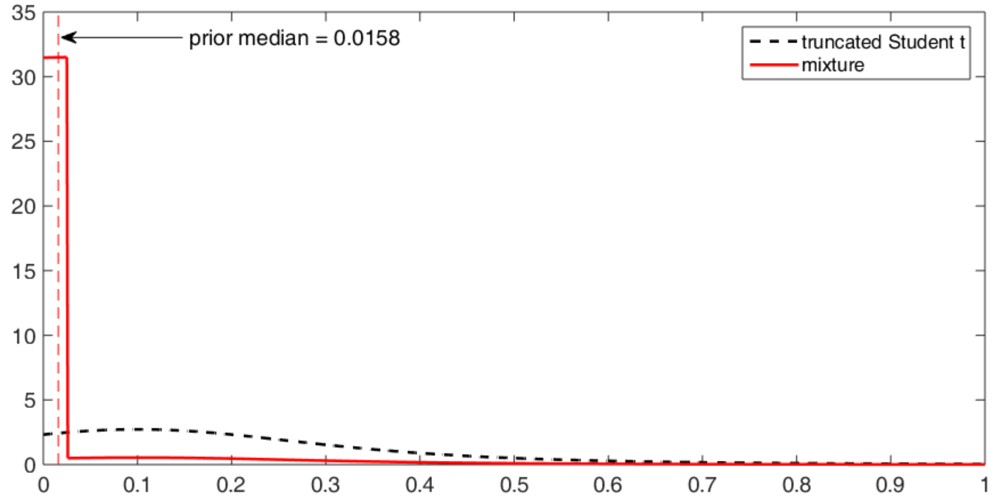
This is simply the transpose of (7), confirming that for this example MLE is numerically identical to IV.

Table 1. Comparison of results obtained using truncated Student $t(0.1,0.2,3)$ prior with those using the mixture prior.

Parameter of interest	**Student *t***	**Mixture**
(1) Posterior median of short-run supply elasticity	0.15	0.15
(2) Posterior median of short-run demand elasticity	-0.35	-0.35
(3) Effect of oil supply shock that raises real oil price by 10% on economic activity 12 months later	-0.50	-0.50
(4) Effect of oil consumption demand shock that raises real oil price by 10% on economic activity 12 months later	0.13	0.13
(5) Effect of oil inventory demand shock that raises real price of oil by 10% on economic activity 12 months later	-0.36	-0.35
(6) Percent of observed oil price increase during June 1990 -Oct 1990 attributed to oil supply shocks	46.1%	46.4%
(7) Percent of observed oil price increase during Jan 2007- June 2008 attributed to oil supply shocks	47.1%	47.5%
(8) Percent of observed oil price decrease during June 2014-Jan 2016 attributed to oil supply shocks	38.1%	38.5%
(9) Percent of observed oil price increase during Feb 2016-Dec 2016 attributed to oil supply shocks	30.7%	31.1%

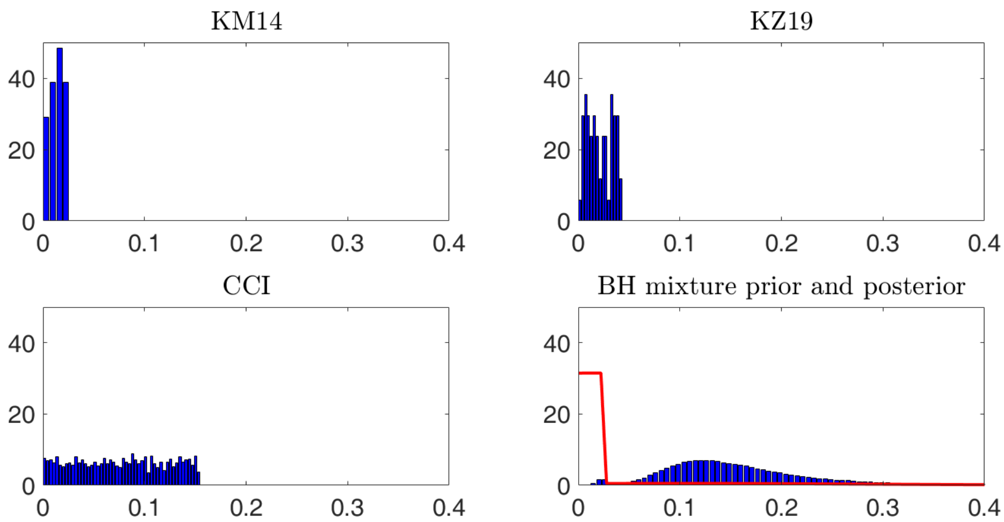
Source: All numbers taken from Tables 3 and 4 in Baumeister and Hamilton (2019).

Figure 1. Two of the priors for short-run supply elasticity used in BH.



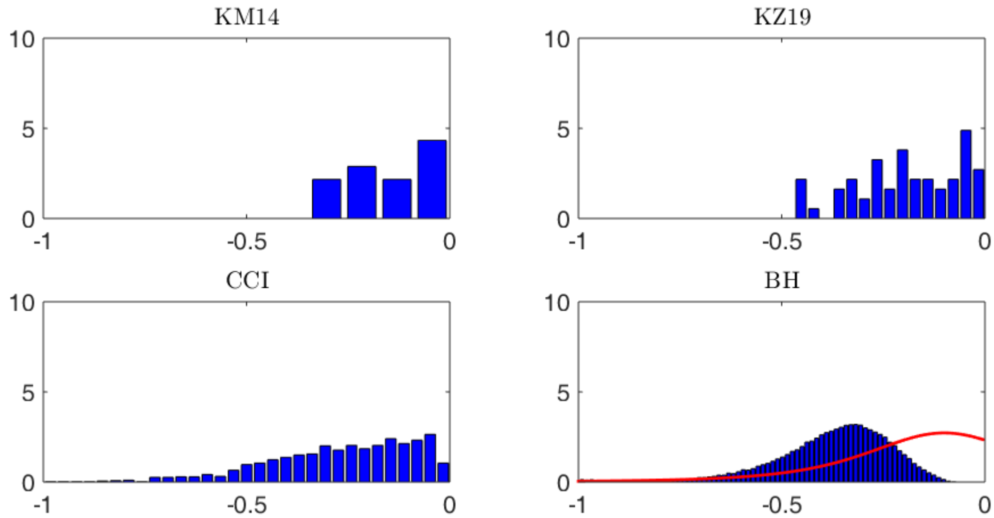
Notes to Figure 1. Dashed black: truncated Student *t*(0.1,0.2,3); solid red: mixture that puts weight 80% on *U*(0,0.0258) and 20% on the truncated Student *t*.

Figure 2. Distribution of short-run supply elasticity implied by the Kilian and Murphy (2014) procedure under three different upper bounds, and prior and posterior distribution implied by the Baumeister and Hamilton (2019) procedure when the mixture prior is used.



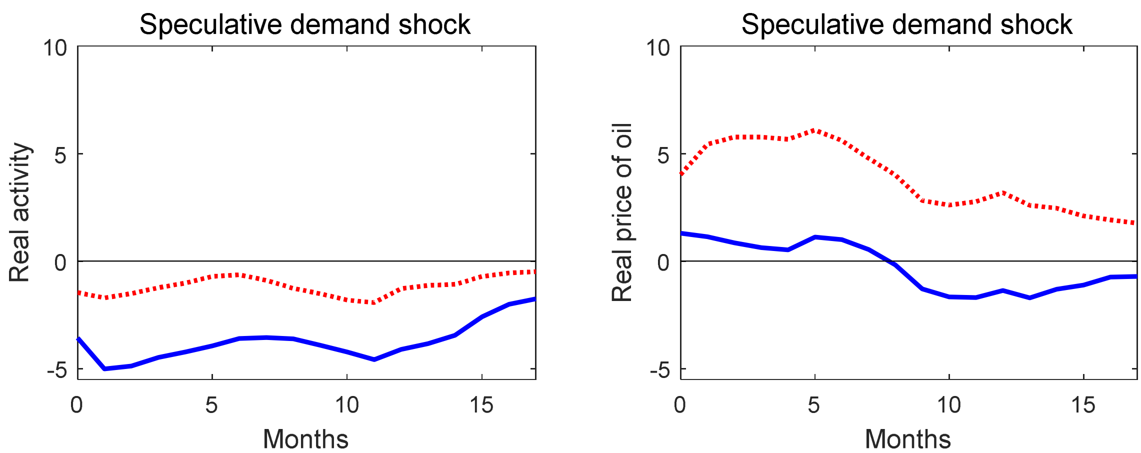
Notes to Figure 2. Upper left: distribution of draws in original Kilian and Murphy (2014) code. Upper right: distribution when upper bound of 0.0258 is replaced by 0.043. Lower left: distribution when upper bound of 0.0258 is replaced by 0.155. Lower right, prior (red) and posterior (blue) for Baumeister and Hamilton's (2019) analysis using the mixture prior.

Figure 3. Distribution of short-run demand elasticity implied by the Kilian and Murphy (2014) procedure under three different upper bounds on the supply elasticity, and prior and posterior distribution for demand elasticity implied by the Baumeister and Hamilton (2019) procedure when the mixture prior for supply elasticity is used.



Notes to Figure 3. Upper left: distribution of draws of demand elasticity in original Kilian and Murphy (2014) code. Upper right: distribution when upper bound of 0.0258 on supply elasticity is replaced by 0.043. Lower left: distribution when upper bound of 0.0258 is replaced by 0.155. Lower right, prior (red) and posterior (blue) for demand elasticity in Baumeister and Hamilton's (2019) analysis using the mixture prior for supply elasticity.

Figure 4. Effects of "speculative oil demand shock" for the Kilian and Murphy (2014) specification and data set using two different seeds for the random number generator.



Notes to Figure 4. Left panel: effect on real activity. Right panel: effect on real price of oil. Red dotted lines: seed = 316, which was the original seed used by Kilian and Murphy (2014) and which reproduces panels (3,2) and (3,3) In Kilian and Murphy's Figure 1. Blue solid lines: seed = 613. Source: Baumeister and Hamilton (2017).