Tailoring Restrictions for SVARs in the Oil Market

William Swanson
U.S. Energy Information Agency

Thomas Lee
U.S. Energy Information Agency

March 23, 2022

Abstract

The proliferation of methods to constrain and identify Structural Vector Autoregressions (SVARs) has improved the precision of inference for this type of model. Some papers propose the 'kitchen sink'-approach to identification, and without a clear measure of accuracy to suggest otherwise, they produce set-identified SVARs using a broad mixture of restrictions. Considered together, such restrictions may be at best complementary or at worst contradictory. To help clean house, we propose in this paper two measures of conditional forecast *accuracy* for unobserved structural shocks. In a series of simulations, we show that these measures can be used to optimize a model's restrictions and bring it closer to the true data generating process. We then apply the measures to several well known SVARs for the oil market and argue that the often-imposed upper bound restriction on the price elasticity of oil supply should be relaxed by around half.

1 Introduction

The real oil price is a key variable for the model-based macroeconomic projections of governments and professional forecasters. The framework of reduced-form Vector Autoregressions (VARs) can generate unconditional oil price forecasts with lower Mean Squared Prediction Error (MSPE) and improved directional accuracy over the no-change forecast. When coupled with a set of identifying restrictions, a key feature of the VAR is that it can generate structural interpretations of the oil price and other macroeconomic aggregates. The enhanced Structural-VAR (SVAR) brings a host of powerful interpretative tools for generating hypothetical scenarios and sorting through past historical events, but it also brings potential for the pitfalls of mis-identifcation and bias. In forecasting hypothetical scenarios, for instance, the possible outcomes can vary widely by identification scheme. This leaves applied economists wanting for a metric of out-of-sample forecast accuracy analogous to MSPE for unconditional forecasts. While there is siginficant progress in that direction (e.g Clark and

McCracken (2014), Antolin-Diaz, Petrella, and Rubio-Ramírez (2021)), to our knowledge, there are no measures for forecast accuracy conditional on unobserved structural shocks.

This paper develops two measures to address this gap in the literature. The results here, we hope, can guide SVAR selection and help fine-tune some categories of identifying restrictions. In a series of simulations, we show how both measures move us closer to the true parameters of the underlying data process. Minimizing the values of our measures will also minimize the distance between the model's fitted parameters and the true ones driving the data. Importantly, this will also minimize the squared distance between the model's structural shocks and the true ones. We then apply our new measures to five well known SVARs in the oil market literature, each with a unique and robust set of identifying assumptions. The papers we study are: Kilian (2009), Kilian and Murphy (2012), Baumeister and Hamilton (2019), Antolín-Díaz and Rubio-Ramírez (2018), Kilian and Murphy (2014). But our results are applicable even more broadly. The last section of this paper focuses on the framework in Antolín-Díaz and Rubio-Ramírez (2018), where we apply our measures to optimize over the bounds on the price elasticity of oil supply and show that the optimal bound implied coincides almost exactly with recent microeconomic estimates from North Dakota (see Bjørnland, Nordvik, and Rohrer (2021),Zhou (2020)).

Our first task is to carefully define the target against which we can measure the SVAR's results. This is a delicate question that some authors have touched on (e.g. Lucas (1980) and McCloskey (1998)) and we should carefully define the intended scope of this paper before going too far into the details. While the reduced-form VAR can generate forecasts for a world that is measurable against observed data, the structural model's key feature is that it gives us insights into a world that is *not* observed. That is, it provides insights into a world that is consistent with a theoretical framework that lives in the mind of the researcher. Our goal here is *not* to measure the 'realism' of any particular theoretical model imposed in an SVAR. Rather, our goal is to derive a measure for the relative ability of any set of restrictions to bring an SVAR closer to the researcher's ideal theoretical model, given that they have already chosen one.

We use the 'narrative' framework in Antolín-Díaz and Rubio-Ramírez (2018) to define a set of restrictions on the relative magnitudes and signs of historical structural shocks implied by an SVAR's impulse response functions. This approach, known as narrative restrictions, makes use of our consensus narrative understanding of events to build a set of identifying restrictions on a model's structural shocks. In this paper, we build our own set of nine narrative restrictions around three major global events. We can then measure how much of a 'surprise' our set of restrictions generates for the likelihood functions of the five oil market-SVAR baseline models mentioned above. If the amount of surprise is large, that is, if the restrictions are highly unlikely to be consistent with the SVAR's likelihood function, then the baseline model is performing poorly by our measures. Alternatively, if a baseline model's likelihood function regards the restrictions as highly likely, the model is performing well.

Our measures attempt to quantify this 'surprise'. We use the Kullback-Leibler (KL) estimate of divergence for our headline results, and we show how it is related to a simple expected value of a conditional forecast satisfying the restriction set. Zhou (2011) also considers the usefulness of KL divergence for Bayesian models, but their work is limited to considering unconditional forecasts. Antolin-Diaz et al. (2021) uses KL as a measure for the likelihood of an event, but never suggests using it in the reverse as we do here, i.e. using an event to inform us about the likelihood of a model.

Upon developing the measures and algorithms to compute them, we show in a series of simulations that as a baseline model is increasingly restricted, the structural shocks move closer to representing the true unobserved shocks driving the data. We vary the upper bound on the price elasticity of oil supply over a broad range to see if our measures can be used to glean information about the true elasticity of oil supply underlying the simulated data. We then plug in the optimal upper bound into the baseline model and, when the optimal upper bound is imposed, the model correctly estimates the mean of the elasticity for the underlying data process. There are two main takeaways from this simulation. First, restrictions generally help the baseline model move closer to the target model with restrictions. Second, for each baseline model, we can find a set of parameters that minimizes our measures, and that optimal set will also center the baselines' posterior distribution over the true DGP's parameters.

In this paper's empirical application, increasing the elasticity bound to 80% above a commonly used threshold of 0.0258 reveals that the model of Antolín-Díaz and Rubio-Ramírez (2018) performs 20% better. Remarkably, Zhou (2020) also concludes that the one-month elasticity for oil supply should be increased to our prescribed level, but they use very different information to arrive there. Citing data from oil producers in North Dakota (Bjørnland et al. (2021)), Zhou (2020) shows that the largest credible micro-econometric estimate using reduced form regressions for this elasticity is 0.04, which aligns almost exactly with our optimal bound.

We then perform an informal experiment to test how different the Antolín-Díaz and Rubio-Ramírez (2018) model is under 0.04 versus 0.0258. We run a hypothetical scenario: *what happens if we have another Covid19 outbreak like we did in February and March of 2020?* We use the approach in Baumeister and Kilian (2014) and pull the aggregate demand shock from those months of 2020, and apply it to an unconditional forecast of the oil price going forward 18 months. The results of the scenario's impact on the level of the oil price are modest. With our optimized model, the scenario predicts an extra five point drop in the oil price relative to the baseline. These results are not dramatic, but they seem reasonable.

This paper proceeds with a description of the theoretical framework and revisits the identification problem for SVARS in Section 2. Section 3 prepares the definition for the restriction sets, and Section 4 employs these sets to arrive at our first measure for forecast accuracy conditional on unobserved shocks. Section 5 builds on this progress to develop a second measure, and Section 6 tests both measures' effectiveness in a laboratory environment. Section 7 is the main empirical application to the oil market and Section 8 concludes.

2 Framework

We consider Structural Vector Autoregressive (SVAR) models that have the form,

$$y_t' A_0 = \sum_{l=1}^{p} y_{t-l}' A_l + c + \varepsilon_t' \text{ for } 1 \leq \text{t} \leq \text{T} , \tag{1}$$

where y_t is a $n \times 1$ vector of variables in the regression. The $n \times n$ matrix A_0 are the structural paramaters that describes the contemporaneous relationships between the variables in y_t. Unexpected structural shocks, ε_t, hit the system each period as an $n \times 1$ vector, and the matrix A_l is an $n \times n$ matrix of parameters for $0 \leq l \leq p$ governing the role of lagged information, and p is the number of lags. The vector c is a $1 \times n$ set of constant parameters and T is the sample size. Structural shocks, conditional on past information, are Gaussian with mean zero and covariance matrix $E\left(\varepsilon_t \varepsilon_t'\right) = I_n$, where I_n is the $n \times n$ identity matrix.

Stacking the lagged y_t and A_l matrices, this model can be written in a tractable way,

$$y_t' A_0 = x_t' A_+ + \varepsilon_t' \text{ for } 1 \leq \text{t} \leq \text{T} \tag{2}$$

where $m = np + 1$, and $A_+' = \left[A_1', \ldots, A_p', c'\right]$ has dimensions $m \times n$. Lagged information is stacked into $x_t' = \left[y_{t-1}', \ldots, y_{t-p}', 1\right]$. Conditional on lagged data, the expected value of the structural shocks is $E(\varepsilon_t | \left\{y_{t-1}\right\}_{t-j, j>0}) = 0$.

The system in Equation (2) is defined in terms of the structural parameters $\theta = (A_0, A_+)$, however in many practical applications, only the reduced form parameters (B, Σ) in Equation (3) below can be directly estimated. There is a clear relationship between the structural parameters and reduced form ones, since $B = A_+ A_0^{-1}$ and the $u_t' = \varepsilon_t' A_0^{-1}$. Therefore $\Sigma = E\left(u_t u_t'\right) = \left(A_0 A_0'\right)^{-1}$ is the covariance matrix of the reduced form model errors. The conditional expectation of the reduced form errors is $E(u_t| \{y_{t-1}\}_{t-j,j>0}) = 0$.

$$y_t' = x_t'B + u_t' \text{ for } 1 \leq \text{t} \leq \text{T} \tag{3}$$

The mapping between the reduced form parameters and the structural ones is not unique. This gives way to a well known identification problem for SVARs, and to resolve it, some sort of extra identifying restrictions should be imposed to estimate A_0. The issue arises because there are too many free variables in θ. With $n(n+1)/2$ distinct values in Σ, but n^2 in A_0, there are at least $n(n-1)/2$ degrees of freedom. Adding restrictions to the acceptable values of A_0 has been fruitful for much of the literature on this topic and there are now many well-researched paths to identification (e.g. Alquist, Kilian, and Vigfusson (2011), Antolin-Diaz et al. (2021)).

Some identification methods are especially common for oil market SVARS. For example, Sims et al. (1986) assumes that A_0 is a Cholesky factor of Σ, thereby forcing the unidentified parameters of A_0 to be zero. Another common approach is to put sign restrictions on the entries of A_0^{-1} following Uhlig (2005) and Rubio-Ramirez, Waggoner, and Zha (2010). This method has proliferated widely and we include a brief description of the procedure in the Appendix. Yet another popular strategy is to impose restrictions on the relative magnitudes and signs of the structural shocks implied by the historical decompositions following Antolín-Díaz and Rubio-Ramírez (2018).

3 Measuring the Success of an SVAR

This paper's first step to measuring an SVAR's success at generating realistic insights is to carefully define the target against which we can measure the SVAR's results. As we mentioned in the Introduction, while the reduced-form VAR in Equation (3) can create forecasts for a world that can be measured and compared against observed data, the structural model's key feature is that it gives us insights into what is not observed. That is, it provides insights into a world that is consistent with a theoretical framework that lives in the mind of the researcher. Our goal here is not to measure the 'quality' or the 'realism' of any particular theoretical model imposed in an SVAR. Rather, our goal is to derive a measure for the relative ability of any set of restrictions to bring an SVAR closer to the researcher's ideal theoretical model, given that they have already chosen one.

If there is a restriction that is known to be true of the world and consistent with the theoretical model in mind, then surely it should be used to improve the structural model. This is consistent with Uhlig (2017)'s forceful first principle of sign-restricted SVARs: *if you* *know it, impose it*. Unfortunately this is almost never the case, since adding restrictions to a model can add significant technical complexity. The reverse is much more often the case, where the researcher is required to impose things they are not sure about so that the model is tractable. In this paper, we point to some identifying information that can be directly imposed on the model without adding complexity, and we suggest a method for making use of some of it in a tractable way.

A narrative accounting of events through time is sometimes the first information that is accessible to researchers. Antolín-Díaz and Rubio-Ramírez (2018) makes use of this narrative information explicitly to develop a method for identifying model parameters by restricting an SVARs structural shocks to align with our narrative understanding of events. This seems like a rich source of information, since after all, there are many examples where economic theory was derived from a narrative accounting of historical events[1], it seems appropriate to return there to find empirical identifying restrictions. We will now fix some ideas and define a simple function that is the starting point for narrative restrictions. We start by re-arranging Equation (2) so that it is a function of the data and structural parameters

$$\varepsilon_t = y_t' A_0 - x_t' A_+. \tag{4}$$

We can use this to define a function $g(.)$ so that,

$$\varepsilon_t = g(y_t; x_t, \theta) \tag{5}$$

and for any draw of θ, the function $g(.)$ is invertible.

$$y_t = g^{-1}(\varepsilon_t; x_t, \theta) \tag{6}$$

Suppose we have a narrative accounting of event i that continues for h_i periods, then the structural shocks around that event are $\varepsilon^i = \varepsilon^i_{s_i}, ..., \varepsilon^i_{s_i+h_i}$. The event i is part of the broader set of narrative events v, so that $i \in v$, and the full set of structural shocks on which we feel comfortable putting restrictions is ε^v. Following Antolín-Díaz and Rubio-Ramírez (2018), we set restrictions on the signs of the structural shocks and the relative contribution of structural shocks as measured the historical decomposition.

Using their framework, we can apply three general types of restrictions:

- **Type 1**: Shock-sign, e.g. Shock h in time t is negative (positive).
- **Type 2**: Contribution, e.g. the contribution of shock h in time t is greater (less)

[1] In a well known example, Keynes (1939) derives a theory of liquidity and investment from his personal observations throughout the unfolding of the Great Depression.

than any other shock to shifts in variable k that period.

- **Type 3**: Sum Contribution, e.g. The contribution of shock h in time t is greater (less) than then sum of absolute values of all other shock-contributions that period for variable k that period.

All the restrictions can be stacked together for all the events in our target set v and this defines the vector $\phi(.)$. When all the restrictions of $\phi(.)$ are met, then every element of the vector is greater than 0. Another way to say this is,

$$\phi\left(\theta, \varepsilon^{i}\right) > 0_{qx1} \tag{7}$$

where q is the number of restrictions. Notice that $\phi(.)$ is also a function of the structural parameters, because historical decompositions are functions of both structural shocks and the model's structural parameters. Equation (7) can now be written using (5) to express the function $\phi(.)$ in terms of the observed data. This ensures the function is continuous with respect to θ and aids in the numerical integration performed later in this paper. The indicator function $\Phi(.)$ below in Equation (8) takes the value of 1 if the constraints are met on all restricted shocks ε^{v} and will be 0 otherwise.

$$\Phi\left(\theta, y^{v}, x^{v}\right) = \mathbf{1}\left[\phi\left(\theta, g(y^{v}; x^{v}, \theta)\right) > 0_{qx1}\right] \tag{8}$$

4 A First Measure: Simple Average

A reasonable first question about the baseline set of model parameters θ is: *how likely is it that the model will generate structural shocks consistent with all the restrictions in* $\phi(.)$? The answer is not definitive about the quality or accuracy of the baseline model, but the answer may be informative nonetheless. This section of the paper will formalize the question and show how to calculate the answer. In later sections, the answer to this question turns out to be a component of the Kullback-Leibler divergence measure discussed in Section 5.

We are then looking for the expected value of the function $\Phi(.)$ over the posterior distribution of parameters $\theta|y^T$ for our baseline model. Letting $\pi(.)$ be a probability distribution function, we can write this expected value as,

$$E_{\theta|y^T}\left[\Phi\left(\theta, y^{\nu}, x^{\nu}\right)\right] = \int \pi\left(\Phi(\theta, y^{\nu}, x^{\nu}) = 1|y^T, \theta\right) \pi\left(\theta|y^T\right) d\theta$$

The Success Rate, $\bar{S}_{\theta|y^T}$, is defined as the expected value that the shocks fails to violate the set of restrictions in $\phi(.)$ for any draw of $\theta|y^T$ given that data y^T has been observed.

$$\bar{S}_{\theta|y^T} = E_{\theta|y^T}\left[\phi\left(\theta, y^T\right)\right]$$

A key feature of $\bar{S}_{\theta|y^T}$ is that it is easy to calculate. The measure can be approximated by a simple sum of the draws of θ that satisfy the restrictions divided by the total number of draws. Provided that the total number of draws N is sufficient to approximate the limiting distribution then the algorithm for deriving this measure is,

Algorithm 1. *The following algorithm makes independent draws from the posterior of θ to calculate $\bar{S}_{\theta|y^T}$ using the narrative restrictions in $\phi(.)$.*

1. Take N draws from the posterior distribution of the baseline model's parameters, $\theta|y^T$.
2. Combine each draw with the observed data y^T to calculate $\Phi(x^v, y^v, \theta)$.
3. Sum the values for $\Phi(x^v, y^v, \theta)$ over all the draws, and divide by the total number of draws, N, to find $\bar{S}_{\theta|y^T}$.

5 A Second Measure: Kullback-Leibler

This paper has so far defined a set of narrative restrictions $\phi(.)$ that the researcher is sure are correct, and they are willing to impose on the baseline model. This section defines a measurement for the 'surprise' this set of restrictions generates for the likelihood function of the baseline model. If the amount of surprise is large, that is, if the restrictions are highly unlikely to be consistent with the baseline model, then the baseline needs improvement or should be avoided.

This 'surprise' can be calculated directly from the distance between likelihood functions with and without the restrictions imposed, i.e. $\pi\left(y^T|\theta\right)$ and $\pi\left(y^T|\theta, \Phi(\theta, y^\nu, x^\nu)\right)$. For this section, distance is measured in informational units or 'bits' that is similar to euclidean distance. If the distance from the baseline model is larger than what an alternative model yields, then the baseline model is farther away from agreeing with the restrictions. Furthermore, if satisfying the restrictions is a necessary condition for having the correct model, then the alternative model should be clearly preferred over the baseline. There is no distance too large or small in absolute terms that we recommend in this section, rather, the relevant criteria we suggest is that the preferred model is closer to meeting the restrictions in $\phi(.)$.

One way to measure this distance is to calculate the Bayes' factor, which is the total probability of observing the data y^T under the narrative restrictions versus observing the data without the restrictions, i.e. Bayes Factor $= \pi\left(y^T|\Phi(\theta, y^\nu, x^\nu)\right)/\pi(y^T)$. This ratio can be rewritten as $\pi\left(\Phi(\theta, y^\nu, x^\nu)|y^T\right)/\pi(\Phi(\theta, y^\nu, x^\nu))$. Notice that the numerator of this factor corresponds with the measure $\bar{S}$ derived in Section 4, but the denominator is the total unconditional probability of meeting the restrictions, and is not as easy to calculate as the measure proposed now, the the Kullback-Leibler divergence that is defined as the informational difference between the generic distributions, g and h.

$$D_{KL} = \int g\left(e^T\right)\log\left(\frac{g(\varepsilon^T)}{h(z)}\right)dz \tag{9}$$

Applying this to our probability functions, $\pi\left(\varepsilon^T|x^T, \theta, \Phi\left(\theta, y_v, x_v\right)\right)$ and $\pi\left(\varepsilon^T|x^T, \theta\right)$, for some variable vector of structural shocks ε^T is,

$$D_{KL} = \int \pi\left(\varepsilon^T|x^T, \theta, \Phi\left(\theta, y_v, x_v\right) = 1\right)\log\left(\frac{\pi\left(\varepsilon^T|x^T, \theta, \Phi\left(\theta, y_v, x_v\right) = 1\right)}{\pi\left(\varepsilon^T|x^T, \theta\right)}\right)d\varepsilon^T \tag{10}$$

The fraction inside the log function can be simplified considerably. Employing Bayes-theorem and re-arranging terms allows us to re-write this as the ratio of likelihoods for satisfying the restrictions given information on the existing shocks, versus without this information.

$$\frac{\pi\left(\varepsilon^T|x^T, \theta, \Phi\left(\theta, y_v, x_v\right) = 1\right)}{\pi\left(\varepsilon^T|x^T, \theta\right)} = \frac{\pi\left(\Phi\left(\theta, \varepsilon^v\right) = 1|\varepsilon^T, x^T, \theta\right)}{\pi\left(\Phi\left(\theta, \varepsilon^v\right) = 1|x^T, \theta\right)} \tag{11}$$

The denominator in (11) can be written as below, and it will be is invariant with respect to the distribution of shocks, so we can move it outside the integral for D_{KL}.

$$\pi\left(\Phi\left(\theta, y^v, x^v\right) = 1|x^T, \theta\right) = \int \pi\left(\varepsilon^v, \Phi\left(\theta, \varepsilon^v\right) = 1|x^T, \theta\right)\pi\left(\varepsilon^T\right)d\varepsilon^T \tag{12}$$

$$= \omega\left(\theta, y^T\right) \tag{13}$$

Using Equations (12) and (11), we can re-write our measure for the case when $\omega\left(.\right) > 0$,

$$D_{KL} = \int \pi\left(\varepsilon^T|x^T, \theta, \Phi\left(\theta, y_v, x_v\right) = 1\right)\log\left(\frac{\pi\left(\Phi\left(\theta, \varepsilon^v\right) = 1|\varepsilon^T, x^T, \theta\right)}{\omega\left(\theta, y^T\right)}\right)d\varepsilon^T \tag{14}$$

Notice that, given lagged values x^T and the stuctural shocks ε^T, the function $\Phi(.)$ is 1 with probability 1. If the structural shocks do not satisfy $\Phi(.)$, then the distribution used to integrate will take on zero mass, i.e. $\pi\left(\varepsilon^T|x^T, \theta, \Phi\left(\theta, y_v, x_v\right) = 1\right) = 0$. Since $\lim_{z\to 0} z\log\left(z\right) = 0$, the integrand takes the value of 0 at that point. Also note that structural shocks in period t are independent from lagged values of the data, x^T and have been normalized to be independent of the parameter draw θ. Therefore we can use these insights to simplify D_{KL} considerably,

$$D_{KL} = \log\left(\frac{1}{\omega\left(\theta, y^T\right)}\right) \tag{15}$$

This measure still varies for each draw of θ, so to arrive at a final summary measure that we can report and compare between baseline models, one more integration is performed across all the parameter draws to compute its expected value,

$$\bar{D}_{KL} = E_{\theta|y^T}\left[D_{KL}\left(\theta, y^T\right)\right] = \frac{1}{\bar{S}_{\theta|y^T}} \int\limits_{\theta\in\Theta} \log\left(\frac{1}{\omega\left(\theta, y^T\right)}\right) \pi\left(\theta|y^T\right) d\theta \tag{16}$$

where the set $\Theta = \{\theta|\Phi\left(x^v, y^v, \theta\right) = 1\}$.

Algorithm 2 *The following algorithm makes independent draws of ε^T from the standard normal distribution and from the posterior of θ to calculate D_{KL}.*

1. Take N draws from the posterior distribution of $\theta|y^T$.

2. Check the constraints in $\Phi(.)$ for each of the N-draws and save those that pass the restrictions. Count the number that pass, P.

3. Then for each of the saved draws i:

   - Draw M new shock series for T periods, $\tilde{\varepsilon}^T$, from the unconditional distribution for structural shocks, which in this case is Standard Normal $N\left(0, I\right)$.
   - For each M-draw, use $\tilde{\varepsilon}^T$ together with θ to check the constraints, and count the number of successes, S_i.
   - Calculate and save the measure $S_i/M \approx \omega\left(\theta, y^T\right)$

4. Calculate $\frac{N}{P}\sum_1^P -log(S_i/M) \approx E_{\theta|y^T}\left[D_{KL}\left(\theta, y^T\right)\right]$

6 Simulation of the Measures

To arrive at an upper bound on the usefullness of our measures $\bar{S}$ and $\bar{D}_{KL}$, in this section we estimate a series of SVARs using calibrated simulated data. Unlike real data, of course, the parameters for this simulated data are known. Having calibrated the data generating process and produced the simulated data, we re-estimated an innocent SVAR and calculate our measures $\bar{S}$ and $\bar{D}_{KL}$ using a two simple narrative restrictions. We impose shocks in the simulate data for one time period only–the period of the narrative test. The shocks we impose have two characteristics (1) there is a negative one standard deviation shocks to oil supply and (2) the shock to demand has a larger total contribution to the oil price. We then define the narrative restrictions to reflect these known characteristics.

The SVARs that we test come in four varieties, spanning from very unrestricted to heavily restricted. The Type 1 SVAR has imposed on it only the restriction that supply curve be upward sloping, while the Type 4 SVAR has the added restrictions that the demand curve be downward sloping, the elasticity of oil supply be bounded from above and below, and the IRFs be constrained to mimic the true IRFs up to three periods out. We expect our measures $\bar{S}$ and $\bar{D}_{KL}$ to show that the as the models become more restricted, in general, the structural shocks move closer to representing the true unobserved shocks behind the data generating process.

Then within each type of SVAR (i.e. 1,2,3 and 4), we vary the upper bound on the price elasticity of oil supply over a broad range to see if our measures $\bar{S}$ and $\bar{D}_{KL}$ can be used to glean information about the true elasticity of oil supply underlying the simulated data. Effectively, we perform a simple optimization by grid search. We vary the upper bound over a broad range of values and select the bound that minimizes $\bar{D}_{KL}$ and $1/\bar{S}$. We then plug in the optimal upper bound into baseline model and, when the optimal upper bound is imposed, the model correctly estimates the mean of the elasticity for the underlying data process.

The data generating process is calibrated to oil price and oil production data, $\{qo_t, rpo_t\}_{t=1}^{T}$, and the details on the data transformation to arrive at the final data set are presented in the Appendix. We use a standard framework for estimating the Bayesian VAR by assuming that the reduced form parameters, B and $\Sigma|B$, are distributed multivariate Normal and Inverse-Wishart, respectively. Using this parameterization is particularly convenient because the priors are conjugate, so the posterior distribution is immediately available for sampling without the need for extra computation. We follow Inoue and Kilian (2013) in defining our initial values for the prior distribution's hyper-parameters (v_0, S_0, B_0, N_0).

$$vec(B)|\Sigma \sim N\left(vec(B_0), \Sigma \otimes N_0^{-1}\right) \tag{17}$$

$$\Sigma \sim IW_n\left(v_0 S_0, v_0\right) \tag{18}$$

Here, N_0 is $np \times np$, S_0 is $n \times n$ and $v_0 > 0$. The posterior is

$$vec(B)|\Sigma \sim N\left(vec(\bar{B}_T), \Sigma \otimes N_T^{-1}\right) \tag{19}$$

$$\Sigma \sim IW_n\left(v_T S_T, v_T\right) \tag{20}$$

where $v_T = T + v_0$, $N_T = N_0 + X'X$, $\bar{B}_T = N_T^{-1}(N_0\bar{B}_0 + X'X\hat{B})$, $S_t = S_0 v_0/v_T + \frac{T}{v_t}\hat{\Sigma} + \frac{1}{v_T}(\hat{B} - \bar{B}_0)'N_0 N_T^{-1}X'X(\hat{B} - \bar{B}_0)$, $\hat{B} = (X'X)^{-1}X'Y$ and $\hat{\Sigma} = (Y - X\hat{B})'(Y - X\hat{B})/T$.

Upon estimating the parameters of the data, we then define a set of structural coefficients $A_0^\star$, being sure that they satisfies the expected features of a structural model for the oil market. Most importantly, we impose that $A_0^\star$ satisfies sign restrictions on all of its elements. For example, $\alpha_{11}^\star$ is the element in the first row and first column of $(A_0^\star)^{-1}$ and it should be negative because a positive supply shock corresponds to a decrease in oil supply. The set of restrictions on $(A_0^\star)^{-1}$ we impose are,

$$a_{11} < 0 \ , \ a_{12} > 0 \ , \ a_{21} > 0 \ , \ a_{22} > 0.$$

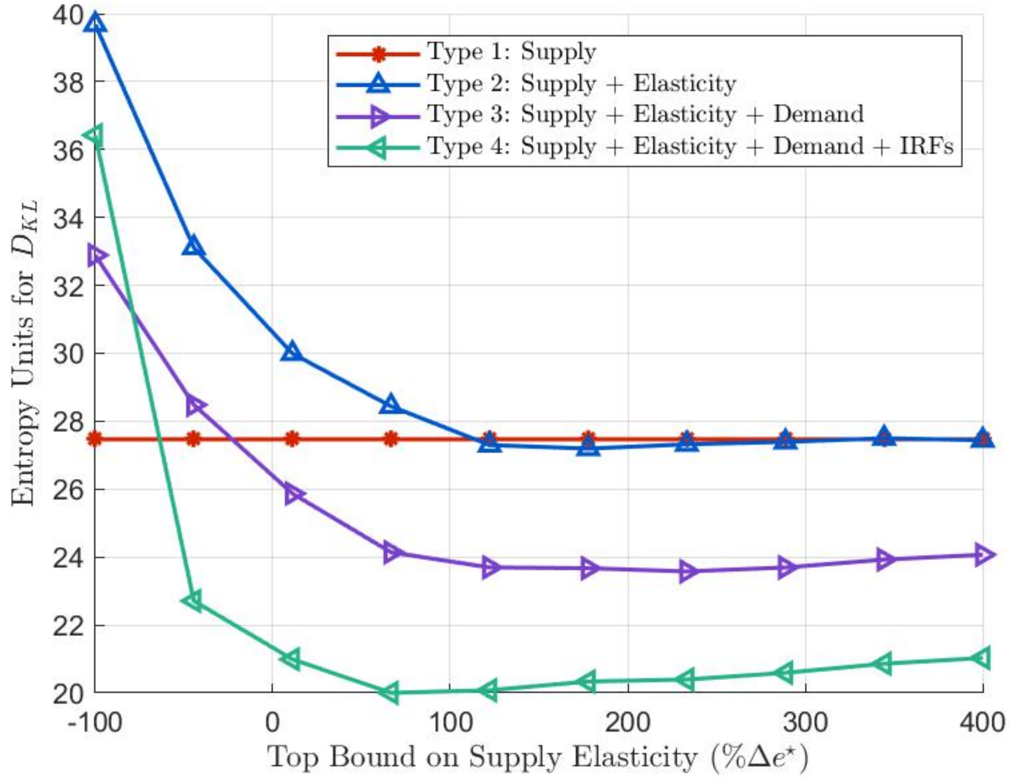
The simulated set of data, $\tilde{y}^T$ looks very similar to the real data, $y^{T,\star}$, and we estimate the model parameters using expressions in (17) and (19) on that data. It closely resembles the original estimated model $\theta^\star$ that we used to create it, as expected.

Our main results for each of the baseline models, Type 1 through Type 4, are presented in Table 1. The measures $\bar{S}$ and $\bar{D}_{KL}$ are in the far right columns of the table, but equally importantly, three other measures are reported in the center of the table that track how closely the model mimics the true data generating process. The table shows that all these measures move together and align with expectations. The first column reported in the table is $MSE^\star = \text{trace}(\tilde{\varepsilon}^{T'}\tilde{\varepsilon}^T)/T$, or the mean sum of squared errors that gives a measure of squared distance between the estimated and true structural shocks. The measure $e - e^* = \alpha_{21}/\alpha_{22} - \alpha_{21}^\star/\alpha_{22}^\star$ is a simple average distance between the estimated and the true elasticity of oil supply. Finally, the most general measure is $||A - A^\star||$, measuring the euclidean distance between the estimated and true elements of the short-term shock impact matrix, A_0^{-1}. This value is calculated as $||\,(A_0)^{-1} - (A_0^\star)^{-1}\,||$. Measures D_{KL} and S are calculated according to (9) and (16).

There are two main takeaways from this simulation. First, restrictions generally help the baseline model move closer to the target model, and this is clearly expressed in the measures $\bar{S}$ and $\bar{D}_{KL}$, as well in the first three columns of 1. Looking to that table, as we move down each column, for a given elasticity bound the measures in all five columns indicate less distance from the DGP. This is visually expressed Figure 1 that likewise shows a decreasing measure of D_{KL} for each baseline model across the full grid of elasticity bounds. Even so,

Figure 1

Simulated Results for D_{KL} as the Number of Model Restrictions Increase



while this is generally true, it is not always true. Notice that the excessively tight elasticity bounds for the Type 4 Baseline cause lower performance than the Type 3 Baseline for the same bound. In this case, the added restrictions for Type 4 compounded with the tighter bounds. Second, for each baseline model, we can find a set of parameters that minimizes our measures S and D, and that set of parameters will also center the baselines posterior distribution over the true DGP's parameters. Figure 2 plots the D_{KL} measure alongside other unobservable measures the span of elasticity bounds, and it clearly shows that all measures track each other closely.

7 Applying the Measures to the Oil Market

The two measures are calculated across five different well-cited SVAR restrictions schemes that span the varieties of approaches currently being used to identify structural shocks in the oil market. We choose these papers, in part, to provide metrics for researchers trying to decide which to use for empirical applications. For each restriction scheme, or *baseline*

Figure 2

Simulated Results for D_{KL} versus Other Measures of Distance from the DGP

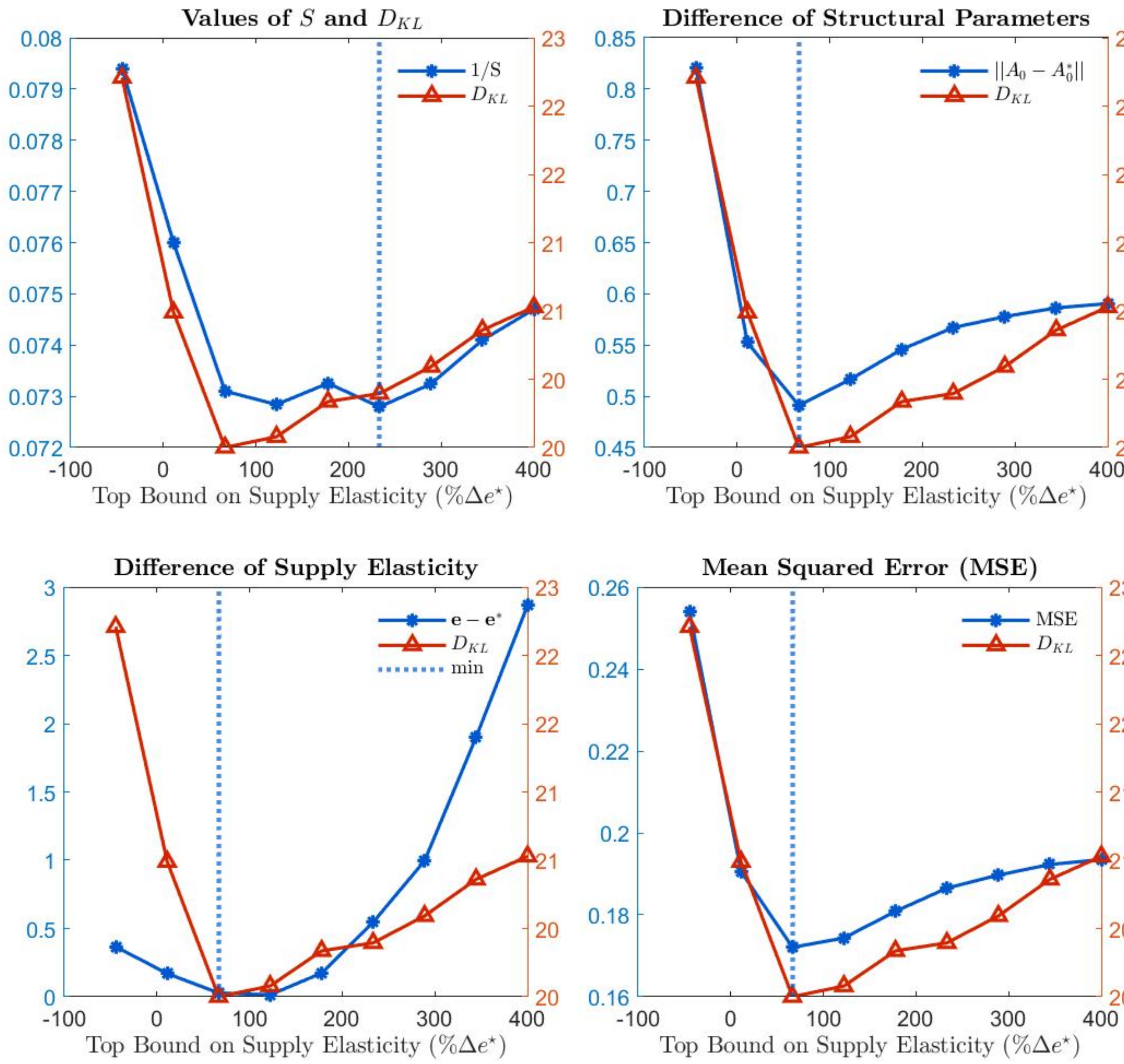


Table 1: **Simulated Results of Unobservable and Observable Measures**

Model + Restrictions	Elasticity Bound	Unobservable Measures MSE^{*}	$e - e^{*}$	$\|\|A - A^{*}\|\|$	Observable Measures S	D_{KL}
Type 1: Supply	.	1.19	5.84	1.92	11.37	27.47
Type 2: Supply + Elasticity Bounds	Tight	1.29	7.05	2.22	9.26	39.69
	Medium	1.19	4.22	1.94	11.41	27.3
	Loose	1.19	4.84	1.92	11.37	27.44
Type 3: Supply + Elasticity Bounds + Demand	Tight	0.47	0.17	1.29	10.68	32.89
	Medium	0.27	0.8	0.76	12.53	23.69
	Loose	0.27	1.96	0.78	12.41	24.07
Type 4: Supply + Elasticity Bounds + Demand + IRFs	Tight	0.39	0.23	1.15	8.97	36.42
	Medium	0.17	1.12	0.52	13.73	20.08
	Loose	0.19	2.69	0.59	13.39	21.03

model, we apply the same target set of narrative restrictions against which we test the baseline model's performance. In this paper, we set nine narrative restrictions that span three distinct events across fourteen years. The first event is the Lehman Brother's collapse in September of 2008 that marks the beginning of the Global Financial Crisis. The second event is the start of the Libyan Civil War that started in late March of 2011. Finally, we include a set of restrictions marking the start of the COVID19 pandemic in February of 2020. The Appendix gives a full accounting of the target restrictions.

There are other narrative restrictions that reasonable people may feel are more important to include, and there is no theoretical reason that nine is the right number of restrictions. In the practical application for this paper, however, nine seems to provide enough bite that it avoids having too many 0-pass rates or perfect 1-pass rates. For this application, we feel our target restrictions accomplish the goal; they bite enough to showcase how informative our measures can be.

To give a sense of how much our measures can change depending on the underlying data set, the baseline SVARs are estimated on four samples of data. All data sets begin in 1973.01, but the samples vary by length and end in one of four different periods: August of 2008, January of 2011, January of 2020 and January of 2022. Breaking the data sets up into smaller samples affords us an opportunity to see the models' relative performance with targets that are outside of the sample on which the baseline model was estimated. The idea is to mimic as closely as possible out of sample SVAR performance. Given how closely our measures are correlated with the $MSE^{\star}$ from the simulations in Section 6, we believe this exercise is a significant contribution.

Of the five SVARs that we investigate, Antolín-Díaz and Rubio-Ramírez (2018) seems to perform the best by our measures. We therefore dig deeper in the results and perform a grid search for the upper bound on the price elasticity of supply that minimizes $1/\bar{S}$ and $\bar{D}_{KL}$ using our set of nine restrictions as the target. The exercise is informative. The current upper bound imposed in the paper, 0.0258 seems too restrictive, so we re-estimate AR18 with the new upper bound. We then apply the re-estimated model to performing a hypothetical scenario: *what happens if we have another COVID19 Pandemic outbreak like we did in February of 2020?*. We compare the results for the original and revised AR18 model to show that the relaxed bound has a modest but reasonable impact on the model's structural scenario analysis and conditional forecasting.

- **Kilian (2009) (K9):** The authors use a recursive identification scheme on a three variable VAR, $y_t = [q_t, \text{rea}_t, \text{rpo}_t, \text{inv}_t]$. This restriction scheme imposes a vertical short-run supply curve, that is, demand cannot influence supply in the short run.

- **Kilian and Murphy (2012) (KM12)**: We impose the sign restrictions as presented in Table 3, together with an upper bound on the price elasticity of oil supply of 0.0258, and a lower bound on element (2,3) of Table 3. This paper is often cited and updated (e.g. Antolín-Díaz and Rubio-Ramírez (2018),Baumeister and Hamilton (2019),Zhou (2020)), so replication code is readily available.

- **Kilian and Murphy (2014) (KM14):** We impose the sign restrictions following the original paper as laid out in Table 4 below. The authors present several sets of restrictions that can be imposed on the model. We impose only the restrictions that coincide with Kilian and Murphy (2012), namely that the price elasticity of oil supply of 0.0258 and a lower bound on element (2,3) of 4 of -1.5.

- **Baumeister and Hamilton (2019) (BH19)** impose priors directly on the elements of A_0 rather than restrictions on A_0^{-1}. We use their data set and the directions to update it are available publically: https://sites.google.com/site/cjsbaumeister/research

- **Antolín-Díaz and Rubio-Ramírez (2018) (AR18):** Their baseline model has four variables: $y_t = [q_t, \text{rea}_t, \text{rpo}_t, \text{inv}_t]$, and follows the design of (Kilian & Murphy, 2012) except for one difference: the authors require that aggregate demand was the least important contributing factor to the spike in the real oil price for August of 1990. Similar to (Kilian, 2009) and (Kilian & Murphy, 2012), we start the model with flat priors (i.e. Jeffrey's) and update their data set so that it runs from January of 1973 until February 2022.

Table 2: **Measures Applied to Five Identification Schemes***

Measure	Sample	Model	Lehman Jan73-Aug08	Libya Jan73-Jan11	Covid Jan73-Jan20	All Events Jan73-Jan22
S	In	K9	100	0	47.4	0.2
		KM12	100	96.1	100	96.1
		KM14	71.9	31.3	90.6	28.1
		AR18	100	98.7	97.3	96.8
		BH19	99.2	100	2	2.2
	Out	K9	100	0	75	0
		KM12	100	84.6	100	44.2
		KM14	64.5	25.8	93.5	6.5
		AR18	100	88	100	52.4
		BH19	70.8	100	17.8	31.4
DK	In	K9	0.91	.	1.77	.
		KM12	0.87	0.84	0.86	2.55
		KM14	1.97	4.27	1.79	13.67
		AR18	0.92	0.95	0.92	2.79
		BH19	0.85	1.2	0.87	2.97
	Out	K9	0.82	.	1	.
		KM12	0.86	1.01	0.81	2.74
		KM14	2.14	7.66	1.84	50.09
		AR18	0.85	0.96	0.83	2.56
		BH19	0.81	1.41	0.82	2.98

*K9 = Killian (2009); AR18 = Antolin-Diaz et al (2018); KM12 = Kilian & Murphey (2012); KM14 = Kilian and Murphey (2013); BH19 = Baumeister and Hamilton (2019).

There are a few results to highlight from Table 2. First, there are clearly drawbacks to using $\bar{S}$, as the measure is more often 100 or 0, whereas $\bar{D}_{KL}$ is never identical across models. Second, the recursive restriction used in $K9$ performs exceptionally poorly. The recursive identification scheme used in that paper are no longer common for oil market vars, and the measures we calculate here provide support for that. Finally, the model with the lowest $\bar{S}$ and $\bar{D}_{KL}$ measures is AR18. The next section will perform more analysis on this model to see if our measures can be further reduced by adjusting the bounds on the oil supply elasticity.

7.1 On the Upper Elasticity Bound

For AR18, varying the elasticity bound from 20% below to 100% percent above the current threshold revealed that at a higher bound, the model performs better by our measures. Relaxing the upper bound to 80% percent above its current level, or 0.0405, seems to improve the model the most. In Figure 3 we plot the $\bar{D}_{KL}$ and $\bar{S}$ measures for each elasticity bound on our grid and it reaches local minimum of $D_{KL} = 2.78$ at the baseline value of 0.0258, but a much lower minimum at another point with a $\bar{D}_{KL} = 2.31$, roughly a 20% improvement.

The bound of 0.0258 was first proposed by Kilian and Murphy (2012), and recently tested for robustness by Zhou (2020). Remarkably, Zhou (2020) also concludes that the one-month elasticity for oil supply should be increased to 0.04 but for very different reasons. Citing data from oil producers in North Dakota in Bjørnland et al. (2021), Zhou (2020) argues that the largest credible micro-econometric estimate using reduced form regressions for this elasticity is 0.035. According to Zhou (2020), the round number of 0.04 seems like a reasonable upper limit. The results in this paper support that conclusion although they were achieved in a very different fashion.

We then perform an informal experiment to test how different the model is under 0.0405 versus 0.02058. We run a hypothetical scenario: *what happens if we have another Covid19 outbreak like we did in February and March of 2020?* We use the approach in (Baumeister & Kilian, 2014) and pull the aggregate demand shock from that period in 2020 and apply it to an unconditional forecast of the oil price going forward 18 months. The results of the scenario's impact on the level of the oil price are presented in the top of Figure 4. The impact is modest. With our optimized AR18 model, the scenario predicts an extra five point drop in the oil price relative to the baseline. This is not dramatic difference, but we see the change as reasonable.

8 Conclusions

When running hypothetical scenarios, the possible outcomes can vary widely by identification scheme and assumption about the reasonable ranges of parameters. In this paper we develop two measures to guide the implementation of SVAR restrictions. We then run a simple simulation exercise to show that in fact both measures move us closer to the true parameters of the underlying data process. In simulations, maximizing the value of statistic will (1) minimizes the distance between the model's fitted parameters and the true ones driving the data and (2) minimize the squared distance between the model's structural shocks and the true ones. We then apply our new measures to several well known SVAR models of the

Figure 3

Optimizing Elasticity Bounds for AR18

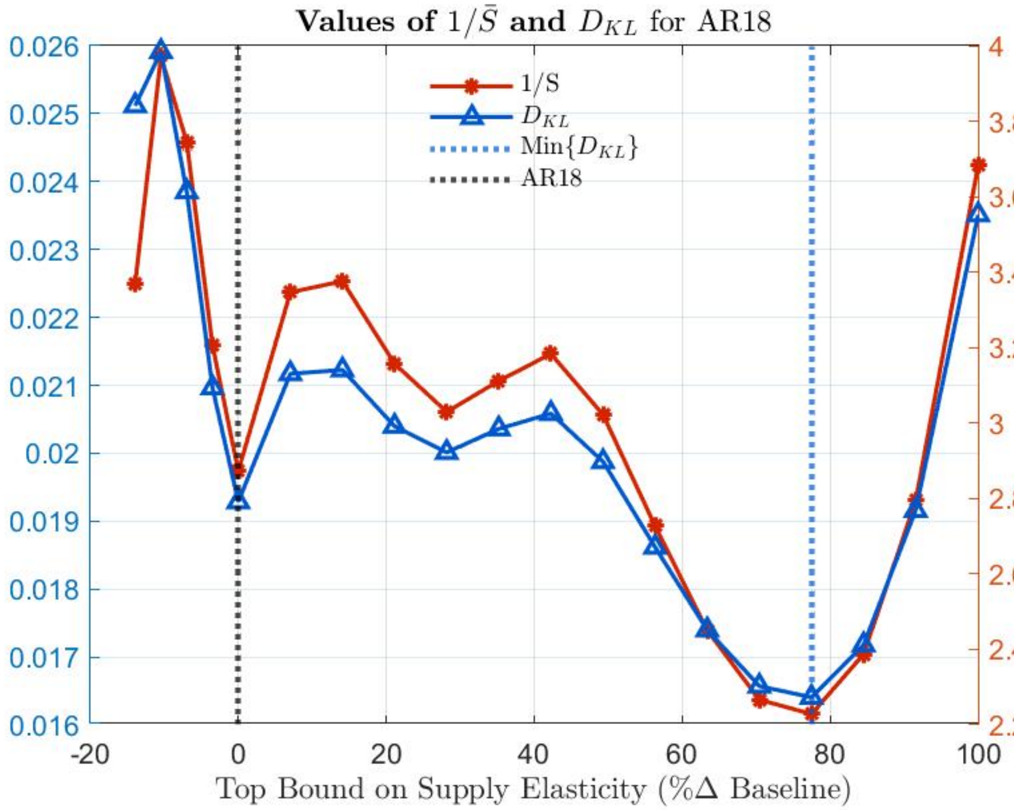
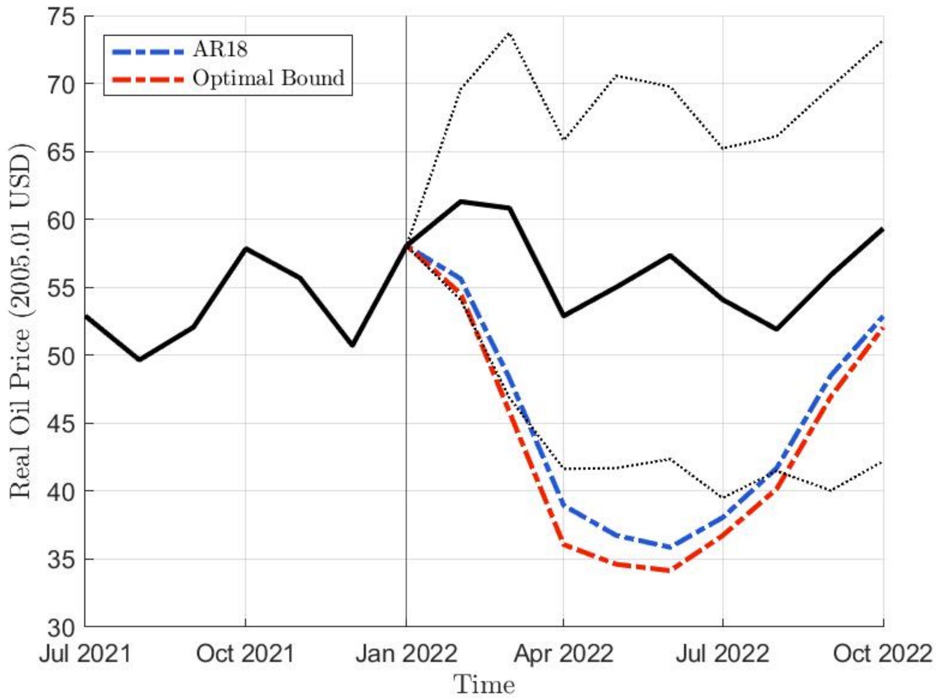


Figure 4

Covid19 Scenario: Applying New Elasticity Bounds in AR18

Optimizing Elasticity Bounds for AR18



Optimizing Elasticity Bounds for AR18

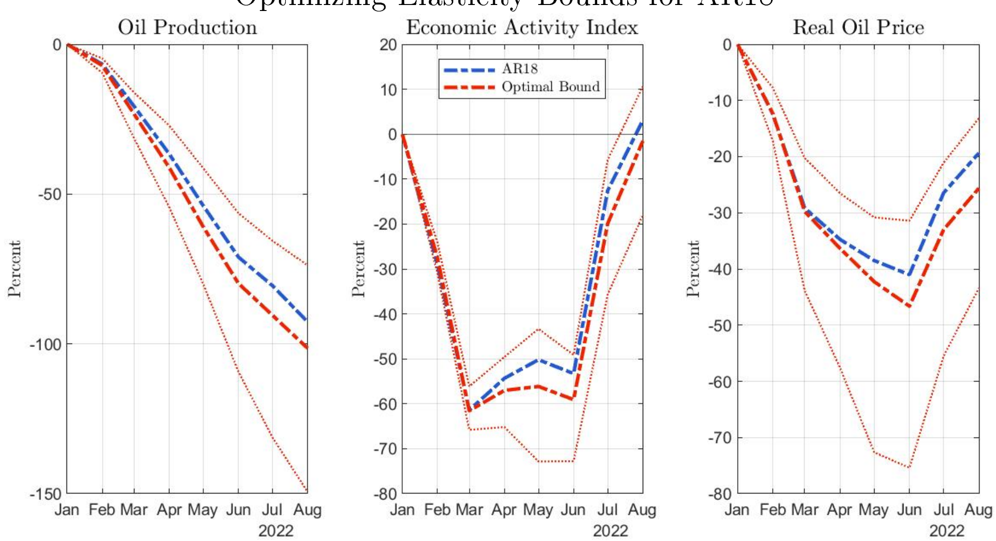


Table 3: **Sign Restrictions on 3-Variable VAR**

	Structural Shocks		
Variables	Oil Supply	Aggregate Demand	Oil-Specific Demand
Oil Production	-	+	+
Real Economic Activity	-	+	-
Real Oil Price	+	+	+

oil market and we optimize over the bounds on the price elasticity of oil supply for one of the models, Antolín-Díaz and Rubio-Ramírez (2018) (AR18). The optimal bound implied by our new measures coincides almost exactly with recent microeconomics estimates from North Dakota. We re-run AR18 to investigate how much the model's predictions have changed as a result of the increased bound, if at all. Specifically, we generate a hypothetical scenario: *what if COVID19 breaks out again?*, and we compare the hypothetical forecasts generated by the original AR18, and the one with the increased bound. The results are close to each other although our revised model predicts an extra 5% drop in the oil price beyond the original model.

A Appendix

A.1 Model Replications

Description of standard approach to applying sign restrictions, e.g. Uhlig (2017) Decompose the reduced form covariance matrix Σ into the pair of matrices LL' using Cholesky factorization. If structural parameters $A_0 = LR$, then as long as $RR' = I$, a cloud of acceptable A-matrices can be generated and filtered until a suitable number is reached that satisfy the all the sign restrictions. Imposing restrictions on the elements in the A_0-matrix will greatly sharpen the identification around the 'true' data generating process by discarding the models that are counterfactual. There is some debate about the best way to draw elements for the R matrix. In most applications, the elements of R are drawn from a normal distribution and then divided by the sum of the draws to normalized R to have unit length. This method is known as drawing from the Haar prior (Haar, 1933).

A.2 Description of the Dataset

- Brent Oil Price: Daily Spot Oil Price data averaged over the month from Feb1989 until Jan2022. Dates before Feb1989 until Jan1973 were backwards-forecasted using the

Table 4: **Sign Restrictions on 4-Variable VAR**

	Structural Shocks			
Variables	Oil Supply	Aggregate Demand	Oil Demand	Speculative Demand
Oil Production	-	.	+	+
Real Economic Activity	-	.	+	-
Real Oil Price	+	.	+	+
Inventories	-	.	-	+

WTI oil price and other correlates. More details are available at https://www.eia.gov/.

- CPI: Monthly Consumer Price Inflation, Jan1953 until Jan2022. Provided by FRED of the US Federal Reserve.
- Real Acquisition Cost for Domestic Fuel Importers. This data can be downloaded from https://www.eia.gov/.
- Real Economic Activity: We use the index developed in Kilian (2009) and updated in Kilian (2019) as an indicator for global economic activity. Data is available here: https://www.dallasfed.org/research/igrea.
- Fuel Inventories: data on this series is available from https://www.eia.gov/.
- Global Oil Production: data on this series is available from https://www.eia.gov/.

A.3 Connection to Likelihood Function

The numerator and denominator of (11) are not common objects in Bayesian VARs, but we show here that they are proportional to the likelihood of the data. To see this, for example, start with the denominator in Equation (11), $\pi\left(\varepsilon^T|x^T,\theta\right)$ and substitute out ε^T using $g(y_t;x_t,\theta)$

$$\begin{aligned}\pi\left(\varepsilon^T|x^T,\theta\right) &= \prod_{s=1}^{s=T}\pi\left(\varepsilon_s|x^T,\theta\right)\\ &= \prod_{s=1}^{s=T}\pi\left(g(y_t;x_t,\theta)|x_t,\theta\right)\\ &= \prod_{s=1}^{s=T}\pi\left(y_t|x_t,\theta\right)\frac{d}{dy}g(y_t;x_t,\theta)\end{aligned}$$

and we arrive at the last line from the chain rule. We can then use (3) to show that $\frac{d}{dy}g(y_t; x_t, \theta) = |\Sigma|^{1/2}$. In total we have

$$\pi\left(\varepsilon^T | x^T, \theta\right) = \pi\left(y^T | \theta\right) |\Sigma|^{1/2}$$

The right hand side is the likelihood function of the data multiplied by a constant.

A.4 Target Narrative Restrictions for Section 7 Application

Narrative Test Restriction 1: "Lehman"
Lehman Brothers Collapse from September 2008 - December 2008

1. Economic Activity Shock is the largest contributor to Aggregate Demand.
2. Economic Activity Shock is negative: $u_{t,rea} < 0$
3. Oil Supply Shock is not the largest contributor to Real Oil Price.
4. Oil Supply Shock is not the largest contributor to Aggregate Demand.

Narrative Test Restriction 2: "Libya"
Start of the Libyan Civil War from February 2011 - March 2011

1. Oil Supply Shock is the largest contributor to Oil Supply.
2. Oil Supply Shock is positive: $u_{t,q} > 0$
3. Economic Activity Shock is not the largest contributor to Real Oil Price.
4. Economic Activity Shock is not the largest contributor to Oil Supply.

Narrative Test Restriction 3: "Covid19"
Covid19 Pandemic Outbreak from February 2020 and March 2020

1. Economic Activity Shock is the largest contributor to Aggregate Demand.
2. Economic Activity Shock is negative: $u_{t,rea} < 0$
3. Oil Supply Shock is not the largest contributor to Real Oil Price.
4. Oil Supply Shock is not the largest contributor to Aggregate Demand.

References

Alquist, R., Kilian, L., & Vigfusson, R. J. (2011). Forecasting the price of oil.

Antolin-Diaz, J., Petrella, I., & Rubio-Ramírez, J. F. (2021). Structural scenario analysis with svars. *Journal of Monetary Economics*, *117*, 798–815.

Antolín-Díaz, J., & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for svars. *American Economic Review*, *108*(10), 2802–29.

Baumeister, C., & Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review*, *109*(5), 1873–1910.

Baumeister, C., & Kilian, L. (2014). Real-time analysis of oil price risks using forecast scenarios. *IMF Economic Review*, *62*(1), 119–145.

Bjørnland, H. C., Nordvik, F. M., & Rohrer, M. (2021). Supply flexibility in the shale patch: Evidence from north dakota. *Journal of Applied Econometrics*, *36*(3), 273–292.

Clark, T. E., & McCracken, M. W. (2014). Evaluating conditional forecasts from vector autoregressions. *Available at SSRN 2504526*.

Haar, A. (1933). Der massbegriff in der theorie der kontinuierlichen gruppen. *Annals of Mathematics*, *34*(1), 147–169. Retrieved from `http://www.jstor.org/stable/1968346`

Inoue, A., & Kilian, L. (2013). Inference on impulse response functions in structural var models. *Journal of Econometrics*, *177*(1), 1–13.

Keynes, J. M. (1939). *Treatise on money and the general theory of employment, interest and money.* NY.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, *99*(3), 1053–69.

Kilian, L. (2019). Measuring global real economic activity: Do recent critiques hold up to scrutiny? *Economics Letters*, *178*, 106–110.

Kilian, L., & Murphy, D. P. (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, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied econometrics*, *29*(3), 454–478.

Lucas, R. E. (1980). Methods and problems in business cycle theory. *Journal of Money, Credit and banking*, *12*(4), 696–715.

McCloskey, D. N. (1998). *The rhetoric of economics.* Univ of Wisconsin Press.

Rubio-Ramirez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, *77*(2), 665–696.

Sims, C. A., et al. (1986). Are forecasting models usable for policy analysis? *Quarterly Review*, *10*(Win), 2–16.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, *52*(2), 381–419.

Uhlig, H. (2017). Shocks, sign restrictions, and identification. *Advances in economics and econometrics*, *2*, 95.

Zhou, X. (2011). *Bayesian model selection in terms of kullback-leibler discrepancy* (Unpublished doctoral dissertation). Columbia University.

Zhou, X. (2020). Refining the workhorse oil market model. *Journal of Applied Econometrics*, *35*(1), 130–140.