ASSIGNMENT ADVANCED MACROECONOMETRICS

Replication of “How to Construct Monthly VAR Proxies Based on Daily Futures Market Surprises” by Lutz Kilian (2024)

Marina Ferrari and Wiktor Grocholewski

Introduction

In recent years, oil market literature has grown at a rapid pace, generating many alternative approaches to modeling the real price of oil and its macroeconomic consequences. Kilian, among the others, is the author that has most contributed to this field since 2000s (e.g. with Kilian (2006, 2008a, 2009), Lippi and Nobili (2012), Kilian and Murphy (2012) and Inoue and Kilian (2013), Kilian and Murphy (2014), etc.). A particular important example is Kilian and Murphy (2014) which emphasizes the role of oil price expectations on the real price of oil and introduces inventories as a key variable for distinguishing between supply, demand, and speculative shocks.

In this paper, we replicate and extend a part of a recent application by Kilian (2024), which builds on Kilian–Murphy (2014) and Kanzig (2021) framework using different identification strategies for identifying a real oil price shock: (i) local projection instrumental variables (LP-IV), and (ii) SVAR models identified through heteroskedasticity. Despite relying on different identification assumptions, both approaches yield highly similar results.

Critical Assessment of Kilian (2024)

Kilian (2024) adopts a common approach in the Proxy SVAR literature: estimating macroeconomic responses to news shocks derived from surprise changes in daily oil futures prices around OPEC announcements. These price changes are used as an external instrument to identify exogenous variation in the real price of oil. The identifying assumption is that, as long as the risk premium remains unchanged on the announcement day, the change in the oil futures price is triggered only by the information from the OPEC announcement.

The author builds on the 6-variable-VAR specification first used by Kanzig (2021), seeking to improve its identification strategy and robustness of results in three key ways. First, addresses the temporal aggregation bias that arises when daily instruments are used to estimate monthly VARs. He introduces a generalizable framework that maps high-frequency shocks into a lower-frequency data, taking into account the timing of the announcement within the month. We believe this correction plays a crucial role, as even a theoretically valid instrument may fail to isolate the correct monthly structural shock if it is not properly aligned with the frequency of the data.

Second, the most critical point of the paper is to understand what the proxy is actually identifying. Earlier studies, such as Kanzig (2021), treated the OPEC surprise as an oil supply shocks. However, this paper shows that it no longer holds when using the "corrected" proxy. Thus, he adopts the structural VAR framework of Kilian and Murphy (2014) for the core oil market variables, finding that the expectations shock identified by the proxy VAR behaves almost entirely like a flow demand shock, i.e., a surprise increase in demand for oil for current use. The most reliable empirical evidence (Fig. 2) supports this view: On impact, the real price of oil rises, and world oil production increases gradually over the first year, consistently with low short-run supply elasticity (Killian & Murphy 2014). More importantly, global industrial production shows a strong and persistent increase, reflecting the demand-driven nature of the shock.

However, we believe that there are still some inconsistencies with this interpretation. First, the response of U.S. industrial production is zero on impact and negative later, which conflicts with the expectation that the U.S. should benefit from a global demand expansion through trade and exports. If energy costs go up, the demand effect should still dominate in a large open economy like the U.S. In both specifications (Figures 2 and 4), the U.S. responses resemble more those of a supply shock - falling output and rising inflation - making it difficult to reconcile with a clean flow demand narrative. Moreover, oil inventories do not show any decline on impact, even though a sharp rise in the demand would typically lead to a temporary drawdown to smooth consumption. Overall, these results raise some concerns about what exactly is being identified. A plausible explanation is that the instrument may not be strong enough to cleanly isolate the shock of interest. This brings us to the third way in which Kilian (2024) improves upon the existing literature: he accounts for the possibility of weak instruments by implementing Anderson–Rubin robust inference (AR-C.S.). While this improves the reliability of the inference, it also highlights the fragility of the identification and so the need of caution when interpreting the results.

Model Specification and Alternative Identification Strategies

In this work, we replicate Fig. 8a and 8b of Killian (2024), reported below for convenience as Fig. 2 and 4, respectively. We adopt the same six-variable VAR with 24 lags (monthly frequency), including: the Real price of oil, World oil production, World oil inventories, Global real activity, U.S. industrial production, and U.S. consumer price index.

We considered different identification strategies commonly used in the oil market literature, but found most of them unsuitable given the characteristics of our data and the model structure.

Our first consideration was identification via short-run restrictions, following the approach of Kilian (2009), who used a trivariate model to distinguish oil supply, aggregate demand, and oil-specific demand shocks. This implies our shock of interest to be ordered third. However, in the extended model, we could not justify the assumption that U.S. macro variables are unaffected contemporaneously by other structural shocks. As a result, we were not able to put enough short-run restrictions to make it possible to identify shocks. Another method which we considered was long run restrictions; however, we quickly realized that we don't have enough solid arguments to impose such assumptions as changes in supply, demand, and real price levels do impact each other even in the long run.

Another candidate approach was sign restrictions. While we could sign some responses based on theory, we did not have suitable empirical estimates of elasticities or other structural parameter that would allow us to identify the shock of interest. As a result, we used it only to align the signs of responses under our second identification strategy based on heteroskedasticity. Finally, since our data was monthly, we did not consider using by non-Gaussianity given lack the high-frequency variation typically needed in this approach.

First alternative strategy: Local Projection-IV

As a first identification strategy, we implemented a local projection instrumental variable (LP-IV) approach. The first reason is that the results in Kilian (2024) appear somewhat sensitive to model specification, e.g. the impulse responses differ notably when using Kanzig's (2021) VAR (12) specification compared to Kilian and Murphy's (2014) VAR(24). Indeed, local projections are more robust to dynamic misspecification, so it is a valuable robustness check: by estimating each horizon independently, the LPs allows us to assess whether the economic interpretation of the identified shock holds outside the structural VAR setting. Second, LP-IV allows us to use the same external instrument as in Kilian (2024). This ensures that both approaches identify the same underlying structural shock, preserving comparability in economic interpretation.

We replicate the setup of Figure 8b in Kilian (2023), reported below as Figure 2, using the six-variable specification described earlier, but restricting the sample to the period 1989M04 to 2017M12. This restriction is necessary because oil futures markets was limited to selected dates and maturities before April 1989, so future prices were not fluctuating daily.

We follow Kilian (2024) in normalizing all impulse responses to a 10% increase in the real price of oil on impact and in estimating cumulative impulse responses for growth rate variables (Oil production and Inventories) and level responses for the other variables.

For estimation, we use HAC (heteroskedasticity and autocorrelation consistent) standard errors, which are crucial in the LP framework to obtain valid inference. Although the Newey-West automatic bandwidth selection suggested a lag of 7, we adopt a more conservative bandwidth of 12 to account for business cycle persistence.

To assess instrument relevance, we compute the Kleibergen–Paap Wald F-statistic in the first stage of each regression. This test is robust to both heteroskedasticity and autocorrelation. In this application, the F-statistic exceeds 10 across all horizons and variables, indicating it may be a sufficiently strong instrument. So, we report conventional 68% and 90% confidence intervals and not the weak-instrument robust ones.

The most important part of this exercise is to assess whether the LP-IV method recovers a shock with similar economic interpretation as in the proxy SVAR. Comparing the two figures, the LP-IV estimates broadly confirm the identification of a flow demand shock.

The real price of oil rises sharply on impact, and global real activity increases significantly during the first 10 months. While Kilian's SVAR shows a much stronger global activity response, our LP-IV estimates suggest a more moderate effect, peaking around 0.3 to 0.5 percent. World oil production remains flat on impact in our estimates but rises more persistently and with greater magnitude over time compared to the SVAR. Both approaches show a temporary dip in production around the 10th month. Oil inventories show a slightly negative but statistically insignificant response on impact in both specifications, providing no evidence of a speculative demand component. Thus, overall these dynamics are consistent with a positive current demand shock under conditions of sluggish short-run supply elasticity.

Turning to U.S. macroeconomic variables, Kilian's SVAR reports a persistent and statistically significant contraction in U.S. industrial production. In contrast, our LP-IV estimates show a response that is statistically indistinguishable from zero; for U.S. CPI, both models yield a modest and not persistent increase on impact, consistent with pass-through from global oil prices to domestic prices.

In general, one limitation of the LP-IV framework that is clear also is this exercise, is the reliability of estimates at longer horizons. Beyond 30-40 periods they are not only less precise, but also economically implausible. This issue has been already noted in many empirical studied including Ramey and Zubairy (2018), who emphasize that they may suffer from accumulated noise and instability.

Figure 1: IRFs of each variable to a 10% shock of real oil price: LP-IV replication

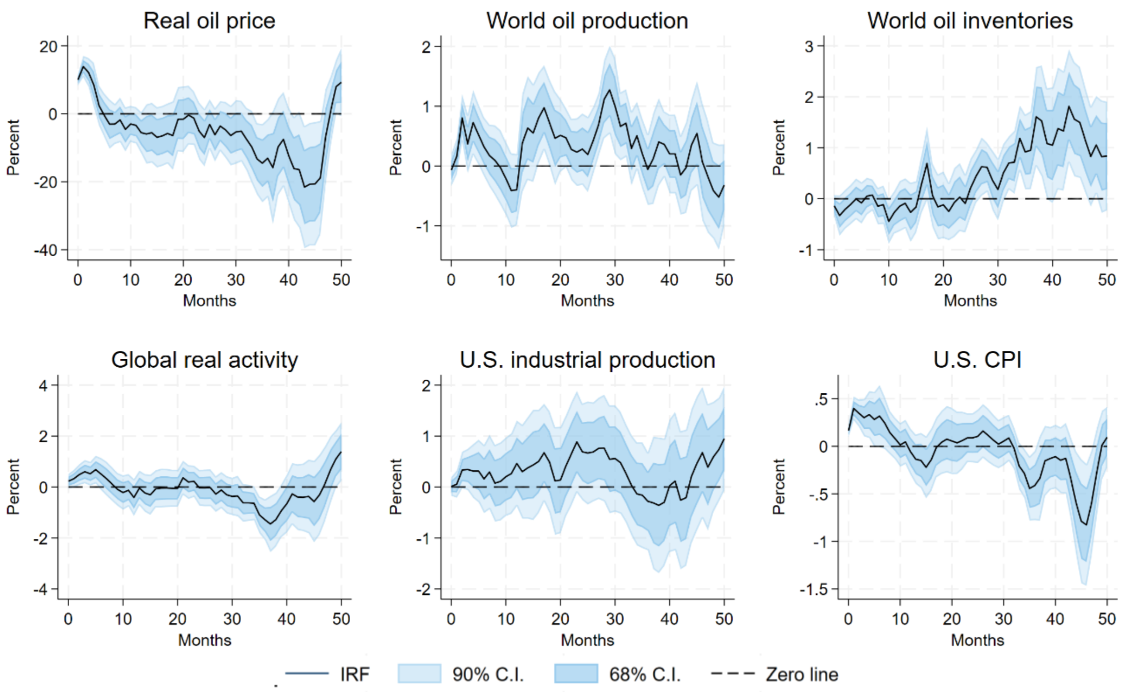
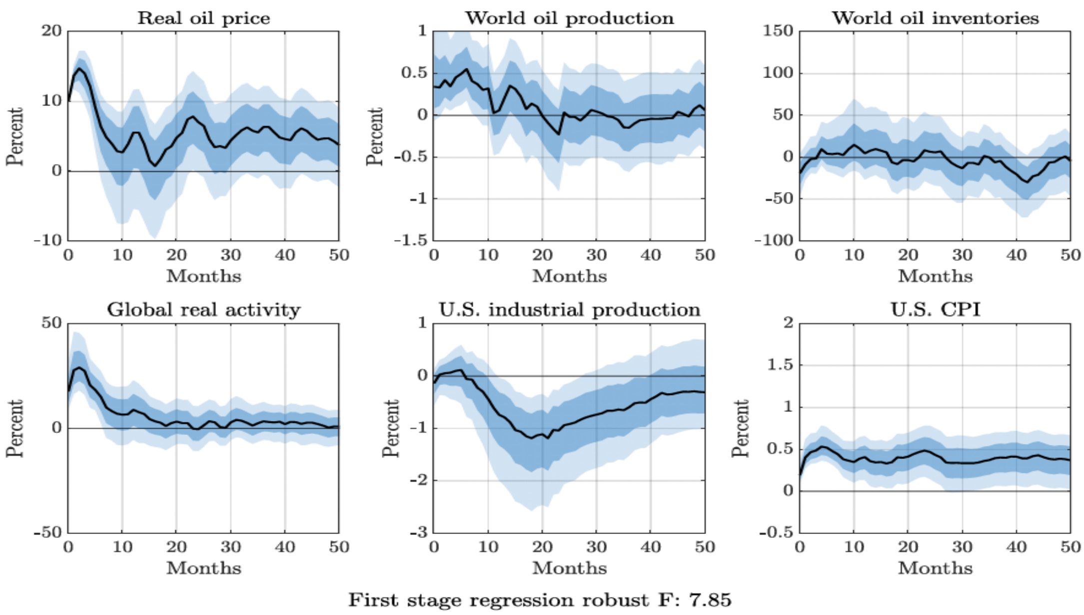


Figure 2: IRFs of each variable to 10% shock of real price: Kilian (2023) comparable results (Fig. 8b of the original paper)



Second alternative strategy: Identification by Heteroskedasticity

The second identification strategy we used is the identification by heteroskedasticity, which was popularized by Rigobon (2003). In its simplest form, it implies that the volatility of shocks changed at one point in time, while the coefficients themselves did not. Therefore, one can estimate separately the covariance matrix before and after the breakpoint, which gives $K(K+1)$ distinct equations, that can be decomposed into a $K * K$ matrix, which is the inverse of B_0, i.e. the matrix of contemporaneous effects and a diagonal matrix Σ, which represents the change in variance after the break. Important to note is that this decomposition is only valid up to a sign flip of each row in ${B^{-1}}_0$ and only for distinct entries of matrix Σ.

$$\Sigma_u = {B^{-1}}_0 {B^{-1}}'_0 \quad or \;\; \Sigma_u = {B^{-1}}_0 \Sigma_w {B^{-1}}_0{}'$$

The reason why we decided to use this strategy is that, as indicated in Kilian (2024), before April 1989, trading in futures markets was limited to selected dates and maturities. These limitations were removed in 1989, which stimulated more trading, and as a result, we expected that the price volatility of oil would similarly increase, as there would be more signals from futures markets.

In order to check the potential break, we used a test by Qu and Perron (2005). As the dataset is not too large, we decided only to consider one break. The test indicated that 1 break is better than 0 at 1% significance and suggested that the right break point is 206th month, i.e., February 1991. We found surprising as it is not exactly April 1989 (that would be 183rd month), but it is still very close, so we assumed that maybe this event only impacted the price volatility with a two-year lag and proceeded with one break at 206th observation. To make sure that this does not influence our results, we also ran the same analysis using the 183rd month as a break, but the IRFs we obtained were exactly identical, so these two years didn't make enough of a difference to raise concerns. Therefore, we decided to continue with the 206th month as a break, which is consistent with the test.

Implementation involved first estimating the reduced form VAR and the residuals, then decomposing the estimated covariance matrix of the shocks before and after the break into the matrix ${B^{-1}}_0$ and the diagonal matrix. The main issue is that we still only decompose it up to a sign flip and we have to assume that the diagonal matrix Σ has distinct entries, which is not an unreasonable assumption, but still a required one. We used an i.i.d. bootstrap with fixed regressors, as for the identification method by heteroskedasticity, it is a valid and the most straightforward approach. We computed percentile-t C.I. as they typically provide asymptotic refinements (Efron, 1981).

A significant downside of this identification strategy is that the structural shocks we estimate are not the same as the shock estimated by SVAR-IV that was used by Killian (2024), as without additional economic interpretation one cannot say which structural shock is which. Therefore, we had to first examine the IRFs for each shock based on economic theory to establish what structural shock it represents. It turned out that the oil price shock, which we wanted to uncover, was most likely the sixth in our matrix representation, after the sign flip. The reasoning was that this shock had a clear positive impact on the Real oil price, as well as U.S. CPI. It also had a positive short-run and negative long-run impact on World industrial production and a positive long-run impact on World oil inventories, consistent with the results found by Killian (2024).

Some minor differences are that in Figure 4, Inventories are affected by a small yet significant drawdown in the short run, consistent with the small positive shock in the World oil production. Moreover, although the shape of the IRFs in Fig. 3 and 4 is the same for U.S. industrial production, Killian (2024) reports a negative effect only in the long run, whereas we also found a negative effect in the short run. Finally, we have also found a negative long run impact on Real global activity.

With that being said, the results match the ones obtained by Killian almost perfectly. On one hand, indicates that there is a chance that both identification methods could be correct, as such a result in highly unlikely under normal circumstances. On the other hand, it raises some concerns about the economic interpretation of the shock, as several variables (like World oil production, Global real activity, U.S. IP) exhibit a significant persistent decline after 15-20 months.

Figure 3: IRFs of each variable to a 10% shock of real oil price: Id. by Heteroskedasticity replication (Percentile t-bootstrap C.I.)

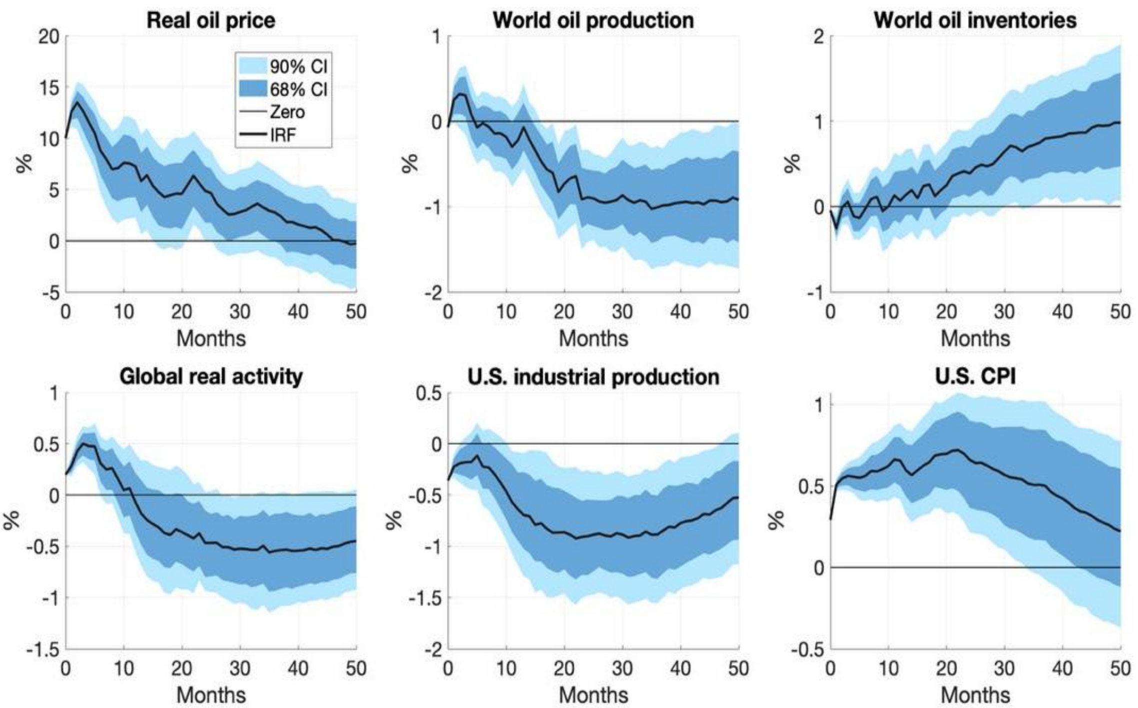
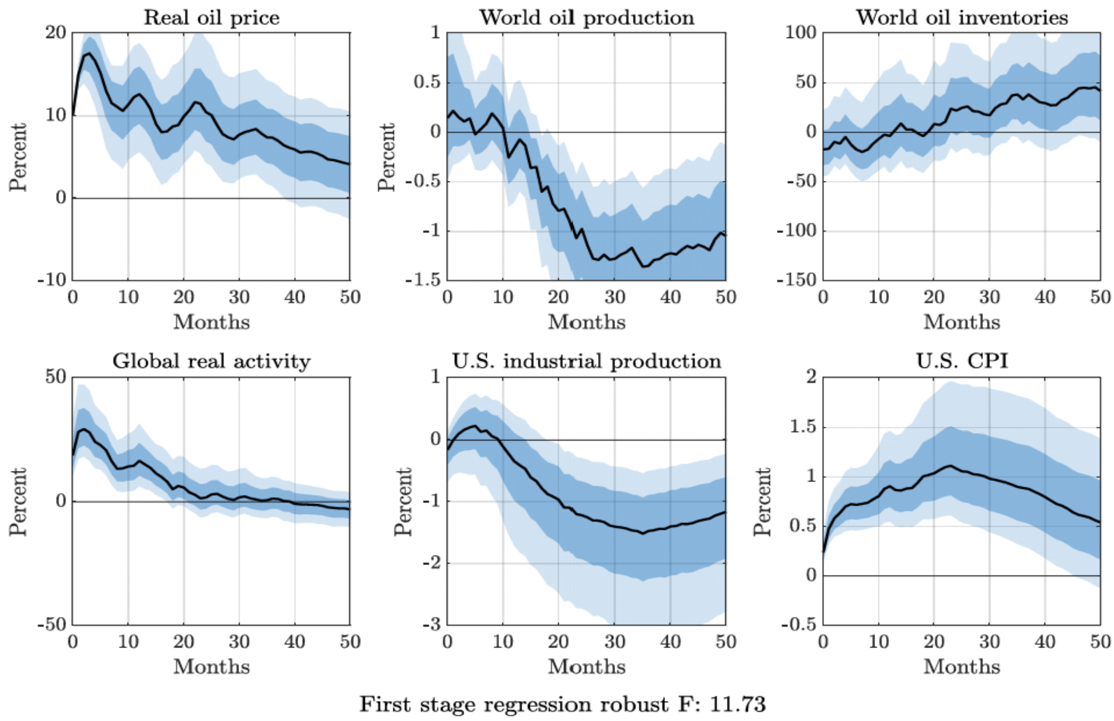


Figure 4: IRFs of each variable to 10% shock of real price: Kilian (2024) comparable results (Fig. 8b of the original paper)



Conclusion

Overall, both the alternative identification strategies yield similar results as Kilian (2024). The LP-IV method appears to be a valid alternative, though it is better suited for evaluating short- to medium-run dynamics. While Fig. 2 should be more reliable than Fig. 4 from Kilian (2024) (given the reliability of the proxy from 1989), the fact that an independent identification approach confirms similar patterns gives additional support for using the extended sample used in Fig. 4. The economic interpretation is broadly coherent as Global real activity consistently rises on impact. However, certain responses remain puzzling, particularly the behavior of U.S. industrial production.

References

Känzig, Diego R. 2021. "The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements." *American Economic Review* 111 (4): 1092–1125.DOI: 10.1257/aer.20190964

Kilian, L. and Murphy, D. (2014) The Role of Inventories and Speculative Trading in the Global Market for Crude Oil. Journal of Applied Econometrics, 29, 454-478. http://dx.doi.org/10.1002/jae.2322

Lutz Kilian, How to construct monthly VAR proxies based on daily surprises in futures markets, Journal of Economic Dynamics and Control, Volume 168, 2024, 104966, ISSN 0165-1889, https://doi.org/10.1016/j.jedc.2024.104966.

Kilian, Lutz and Zhou, Xiaoqing, The Econometrics of Oil Market VAR Models (2020). CESifo Working Paper No. 8153, Available at SSRN: https://ssrn.com/abstract=3552393 or http://dx.doi.org/10.2139/ssrn.3552393

Rigobon, R. (2003). Identification through Heteroskedasticity. *The Review of Economics and Statistics*, *85*(4), 777–792. http://www.jstor.org/stable/3211805

Zhongjun Qu & Pierre Perron, 2005. "Estimating and testing structural changes in multivariate regressions," Boston University - Department of Economics - Working Papers Series WP2005-012, Boston University - Department of Economics.