Discussion of "Local Projections or VARs? A Primer for Macroeconomists" by José Luis Montiel Olea, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf*

Christiane Baumeister
University of Notre Dame
NBER and CEPR

June 11, 2025

In this paper, Montiel Olea, Plagborg-Møller, Qian, and Wolf provide a very useful summary of their recent research agenda that compares the benefits of vector autoregressions (VARs) and local projections (LPs) as alternative ways for conducting impulse response analysis, the main tool for quantifying the dynamic causal effects of one-time shocks on macroeconomic outcomes. While VARs have long been the workhorse model in empirical macroeconomics to represent the complex interrelationships between macroeconomic and financial variables, LPs have emerged as a popular alternative to VARs. The authors frame the question of which estimation method to choose in practice through the lens of a finite-sample bias-variance tradeoff between VARs and LPs of equal order that emphasizes different qualities of both estimators based on canonical statistical criteria.

Given the increasingly widespread use (and abuse) of LPs in applied work, now seems a good time to take stock and systematically evaluate both approaches and this is what Montiel Olea et al. do in the form of "lessons learnt" drawing on theoretical properties and running large-scale simulation exercises. Based on this evidence, they formulate a set of "best-practice" rules, basically a recipe for applied researchers intended to assist them in their selection between the two estimation methods paired with some practical advice on the implementation of various modeling options and specification choices.

Before diving into details, let me begin by calling attention to a key conclusion of the paper that I think bears repeating. Researchers may have two different objectives in choosing the best way to estimate impulse response functions (IRFs). The first objective is accuracy, typically measured by the expected squared difference between the estimate and the truth. The evidence in this

*I thank Jooyoung Cha, Ferre De Graeve, Pascal Frank, Jim Hamilton, Florian Huber, Julian Ludwig, and Christian Matthes for stimulating conversations on the topic.

paper reinforces the conclusion of much earlier research, which is that if the goal is accuracy, IRFs obtained from iterated VARs are often preferable to those from LPs. The second possible objective is statistical reliability of hypothesis tests about IRFs, typically measured by coverage probabilities. The evidence in the paper suggests that if the goal is accurate coverage, LPs are often preferred.

There is no one more qualified than the paper's second discussant to endorse the LP framework, so I view my role as advocating for VARs in this debate. With this in mind, I will scrutinize some of the authors' proposed recommendations and offer a set of modified lessons based on a more multifaceted analysis that supports a strong showing of VARs.

1 The Bias-Variance Tradeoff: It's All in the Lags

A key question of interest in empirical macroeconomics is how does a surprise change in variable x at date t cause us to revise our forecast of variable y for date $t + h$. For illustration, suppose that x is the federal funds rate and that the Fed unexpectedly lowers it by 25 basis points at its June 18, 2025 meeting, then how should that cause us to update our forecast of real GDP growth for, say, 2030Q2? The authors' suggestion is to regress real GDP growth on x and a number of other variables dated five years earlier and use the coefficient on x as the answer. My take is that our forecast of real GDP growth for 2030Q2 should not change at all as a result of today's policy intervention. The reason is that I have strong prior information from two sources: I have statistical information that real GDP growth is stationary, meaning that the long-run forecast equals the unconditional mean, and I have economic information that the real effects of monetary policy die out over time, meaning that monetary policy is neutral in the long run. What general insights can be gained from this simple example that can inform our modeling choice?

1.1 Horizon vs Lag Length: Empirical and Simulation-Based Evidence

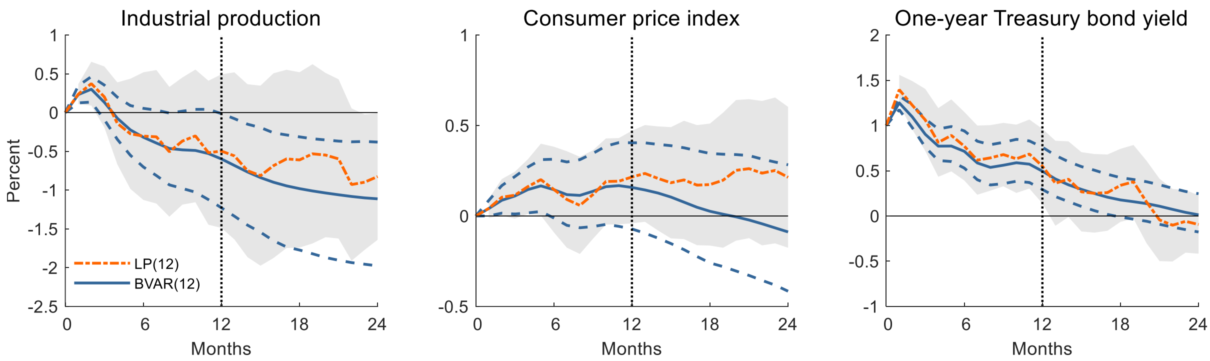
The example highlights that a fundamental question is how long the impulse response horizon h should be. This question is intimately tied to the choice of the lag structure in view of the bias-variance tradeoff between LPs and VARs of equal order summarized in Lessons 4, 5, and 8 in the paper. Given that VARs with lag length p extrapolate impulse responses for horizons $h > p$ from the first p sample autocovariances, they have low variance but are more prone to bias, whereas LPs with the same number of lags attain low bias but at the expense of higher variance at intermediate and long horizons. This suggests that there are two options for practitioners on how to proceed which I will discuss in turn.

Option 1. Don't try to report any estimate for the effects for large h but rather focus on short-run dynamics.

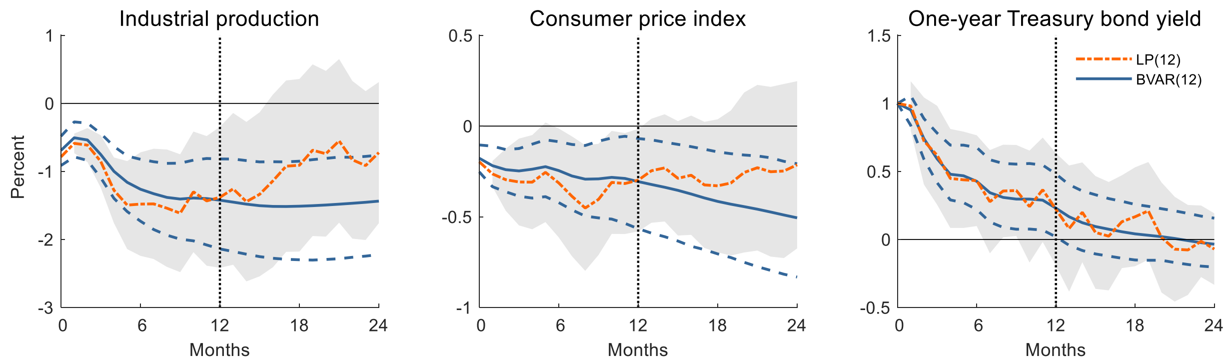
Figure 1 provides an empirical illustration using three standard variables to study the effects

Figure 1. Comparison of IRFs from Monthly Monetary Bayesian VAR(12) and LP(12)

Panel A: Reduced-form IRFs to a 100 Basis Point Increase in the Interest Rate



Panel B: Structural IRFs to a 100 Basis Point Monetary Policy Shock



Notes. Impulse responses from LP(12) and VAR(12) for three-variable monthly model that includes the log of industrial production, the log of consumer price index, and the one-year Treasury bond yield. The VAR is estimated with Bayesian methods where the tightness of the prior is determined in the data-driven way of Giannone, Lenza, and Primiceri (2015). In panel B, the monetary policy shock is identified with the informationally-robust instrument of Miranda-Agrippino and Ricco (2021). The solid blue lines are the median estimates from the Bayesian VAR and the dashed-dotted lines are the point estimates from the LP. Dashed blue lines and shaded areas indicate 90 percent error bands where the former have been obtained based on the 5^{th} and 95^{th} percentiles of the posterior distribution and the latter based on heteroskedasticity-consistent standard errors. The vertical dotted line indicates the lag length. Sample period: 1979.1-2014.12.

of an unexpected change in the interest rate.[1] Both the VAR and the LP are estimated with $p = 12$ which is the conventional choice for monthly data. Panel A reports the reduced-form IRFs for forecasts of each of the three variables after a 100 basis point increase in the one-year rate where the horizontal axis indicates the number of months in advance that the forecast is made for a

[1]The dataset is taken from Miranda-Agrippino and Ricco (2021) and contains monthly observations for the log of industrial production, the log of the consumer price index, and the one-year Treasury bond yield for the period 1979M1 to 2014M12.

maximum of $h_{\max} = 24$.[2] Panel B displays the structural IRFs to a monetary policy shock identified with the informationally-robust instrument proposed in Miranda-Agrippino and Ricco (2021) and normalized to raise the one-year rate by 100 basis points. Both estimation methods, the iterated VAR and the LP, deliver approximately the same IRFs for horizons up to $h = p$, as first shown by Plagborg-Møller and Wolf (2021), but start to diverge thereafter. Hamilton (2025) points out that in practice differences can also arise as a result of using $(h-1)$ fewer observations to construct the LP estimates as the horizon h lengthens. Some other well-known features of VAR-based and LP-based IRFs emerge from the plots: (a) the error bands of VARs are typically tighter yielding more conclusive inference, which suggests that if we care about accuracy, the VAR is the obvious choice, and (b) the IRFs from LPs are noisier, whereas those from VARs are smoother, which is often more appealing for economic interpretation than sudden jumps inherent in LP-based IRFs.

The takeaway is that the IRF horizon should coincide with the lag length used in estimation or at least not exceed it by much. Put differently, the choice of the lag length should be guided by the length of the IRF horizon which in turn should be guided by the empirical question at hand and the reasonableness from an economic point of view. In the illustrative example above, reporting an estimate for the five-year effects of monetary policy is probably not sensible. Given that the differences between IRFs from VARs and LPs tend to be small as long as $p = h$, the debate about which estimation method to choose becomes moot; if anything, applied researchers should favor VARs in light of their smoothness property.

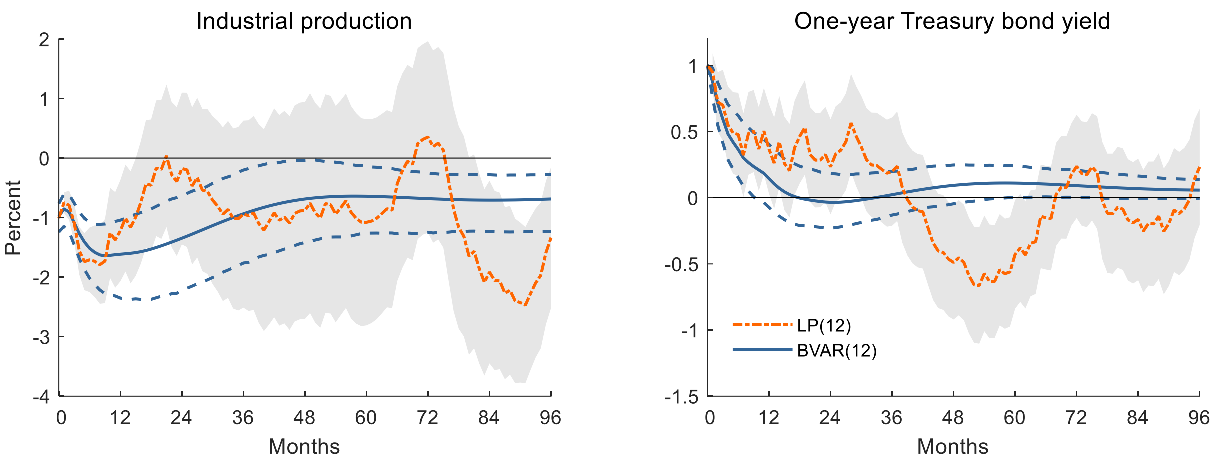
MODIFIED LESSON 1. *Choose the lag length p according to the horizon h of economic interest setting $p \geq h$. In doing so, be mindful that there is only limited information about the long run in a finite sample of data.*

The qualifier 'of economic interest' might pose a problem if the economic question does focus on the long-run effects. For example, there is an ongoing debate about the 'long and variable' lags in the transmission of monetary policy shocks to the real economy. A case in point is the recent contribution by Jordà, Singh, and Taylor (2024), where they show that monetary policy has very persistent and long-lasting effects on economic activity; specifically, using a panel LP-IV framework with annual data covering over a century, they find that a 100 basis point increase in the policy rate leads to a 4% decline in real GDP after 12 years. This is in sharp contrast to the picture I painted earlier about what economists 'know' about the effects of monetary policy at long horizons. To cross-validate their striking finding, Jordà et al. (2024) turn to prominent studies in the literature that use a more standard setup for the U.S. economy with monthly data and simply extend the IRF horizon. What is puzzling is that they switch from their baseline LP framework to Bayesian VAR models even though there exist many applications that use LPs for analyzing the transmission of monetary policy; even the paper by Miranda-Agrippino and Ricco (2021) that they picked offers evidence based on both estimation methods. Extending the IRF horizon to 96 months as Jordà et

[2] Given that contemporaneous observations of industrial production and consumer prices are included as controls, this is equivalent to assuming a recursive structure for identification, commonly known as Cholesky decomposition.

al. (2024) do but now not only for the VAR(12) but also for LP(12) using the codes of Miranda-Agrippino and Ricco (2021), Figure 2 reveals that LP(12) produces IRFs for industrial production that are quite volatile and estimated with substantial uncertainty.[3] It turns out that the effect of monetary policy on industrial production is not statistically significantly different from zero 16 months after the shock and that for the subsequent five years it can be anywhere between -3 and +2 percent; only after 82 months, the response of industrial production becomes significantly negative again, pointing to long-run non-neutrality of the initial 100 basis point increase in interest rates seven years earlier. For comparison, the Bayesian VAR(12) that Jordà et al. (2024) chose for cross validation corroborates their finding of persistent negative effects that are statistically significant throughout the 8-year impulse response horizon.

Figure 2. The Long-Run Real Effect of Monetary Policy Shocks



Notes. Long-horizon impulse responses from LP(12) and Bayesian VAR(12) for baseline six-variable monthly model of Miranda-Agrippino and Ricco (2021) that includes log of industrial production, log of consumer price index, one-year Treasury bond yield, unemployment rate, log of commodity price index, and the excess bond premium with monetary policy shock identified using their informationally-robust instrument. See Figure 1.

A concern discussed extensively by Montiel Olea et al. is that short-lag VARs are more accurate but yield seriously distorted IRFs when extrapolated well beyond their lag length due to dynamic misspecification. Based on their main criterion – confidence intervals with accurate coverage probability – the evidence obtained with LP(12) would be deemed reliable, whereas the VAR(12) results would be deemed too fragile to be trustworthy. This leads me to consider the second option.

Option 2. Increase p as the impulse response horizon lengthens to mitigate concerns of dynamic misspecification but possibly supplement with additional prior information

[3]Relative to Figure 12 in Miranda-Agrippino and Ricco (2021) where they compare IRFs from a VAR with LP, I made two modifications: first, I included the excess bond premium which is part of the VAR model that Jordà et al. (2024) used for replication, and second, when computing error bands, I follow Montiel Olea et al.'s suggestion to not correct for serial correlation in line with Lesson 9. Using HAC standard errors leads to even wider error bands.

to obtain an accurate estimate since information in the data about the long run is limited.

The (obvious) solution to dynamic misspecification inherent in VARs – also hinted at by the authors – would be to dramatically increase the lag length. However, in applied work, researchers are typically reluctant to allow for a generous lag structure for fear of estimation uncertainty being so dominant that no definitive conclusions on economic questions can be reached. De Graeve and Westermark (2013) show that this concern is unfounded and that considerably increasing the number of lags in VARs to better approximate the DGP underlying standard DSGE models is not only feasible but also beneficial. In fact, they point out that increased uncertainty due to parameter proliferation is counteracted by two forces: a reduction in bias *and* a reduction in variance, both resulting from mitigating dynamic misspecifiation thanks to a richer lag structure. In a series of Monte Carlo simulations with data generated from a diverse set of popular DSGE models that do not admit a finite-order VAR representation, they demonstrate that long-lag VARs simultaneously achieve smaller bias, greater accuracy, and better coverage.

Montiel Olea et al. also marshal a variety of simulation support for their conclusion that the longer the impulse response horizon of interest, the more the choice of estimation method matters. While repeatedly alluded to in a parenthetical remark – "LPs or *VARs with very long lag lengths*" – as the only way to robustly achieve low bias and satisfactory coverage, the case of longer lags in VARs is not included in the simulation exercises of Montiel Olea et al. or in the previous work of Li, Plagborg-Møller, and Wolf (2024), and Montiel Olea, Plagborg-Møller, Qian, and Wolf (2024).[4] It seems a grave omission not to consider this possibility as part of a simulation study. I am trying to fill this gap here using the codes provided by Montiel Olea et al. (2024) where they use the Smets-Wouters model as the DGP, which is also covered by De Graeve and Westermark (2013), to draw a direct comparison with the performance of LPs of the same order along the same criteria they applied. All I do is to increase the lag length from $p = 4$, considered in Montiel Olea et al. (2024) as a typical choice in applied work for quarterly data, to $p = 16$, $p = 20$, and $p = 24$.[5] Figure 3 reports the coverage probabilities and median confidence interval length for all four cases for their cost-push shock application. The first row reproduces the result of Montiel Olea et al. (2024) that confidence intervals of a VAR(4) massively undercover for horizons exceeding four quarters, whereas LP(4) coverage is close to the nominal level all the way up to horizon 40 but at the cost of wider confidence bands. Increasing the lag order to capture dynamics of the past four, five, and six years, changes the picture substantially. VAR coverage outperforms LP coverage across the board, while both methods yield comparable bandwidth up to horizon p with VAR estimates being more accurate thereafter.

A recent empirical paper that validates the practical feasibility of long-lag VARs is Antolín-Díaz and Surico (2025). They estimate a quarterly Bayesian VAR with 60 lags using standard

[4] Li et al. (2024) show that increasing the lag length from $p = 4$ to $p = 8$ makes the Bayesian VAR relatively more attractive, as the prior reduces the effective dimensionality of the otherwise highly parameterized VAR system.

[5] For simplicity, I focus on the VAR and LP models where confidence intervals were computed via the delta method.

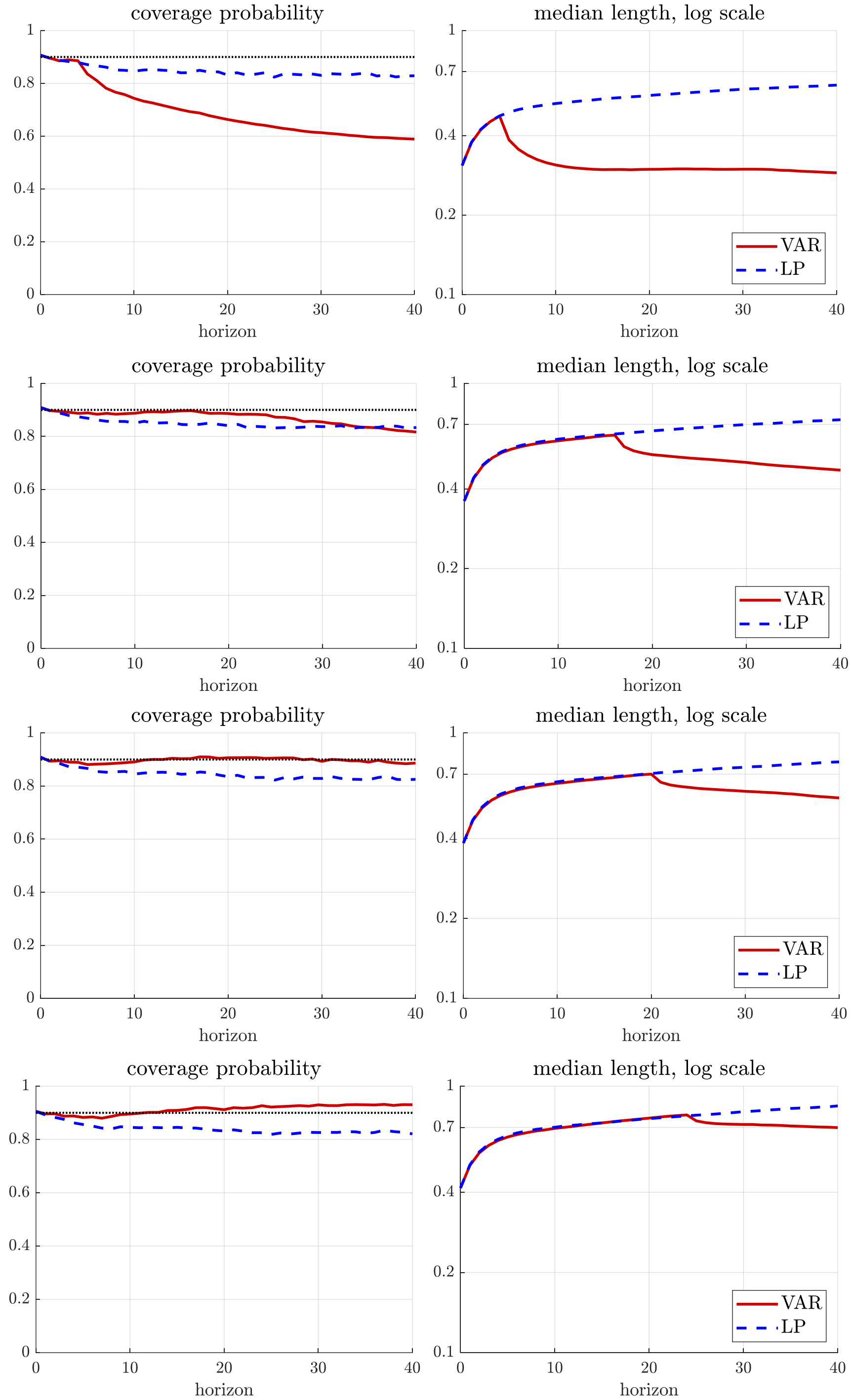
Figure 3. Simulation Evidence

Lag length p = 4

Lag length p = 16

Lag length p = 20

Lag length p = 24

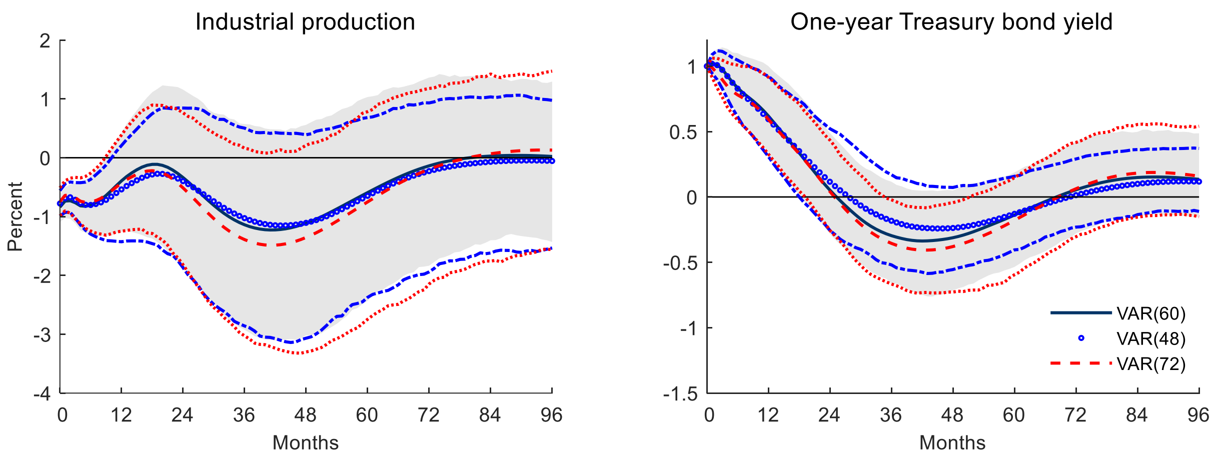


Notes. Simulations based on codes of Montiel Olea et al. (2024) with data generated from Smets and Wouters (2007) DSGE model for 5,000 samples of size T = 240 quarterly observations. Response of inflation to cost-push shock estimated with VARs of different lag order. Confidence intervals are computed via the delta method. The target confidence level is 90 percent.

shrinkage priors to investigate whether U.S. government spending stimulates long-run growth and study IRFs for horizons of up to 15 years, thus effectively setting $p = h$ as suggested earlier, given their interest in the role of long-lasting diffusion of technological innovations. They find a significant and persistent increase in output and productivity that extends beyond business-cycle frequencies. They corroborate the reliability of their inference about the long-run effects with a Monte Carlo analysis and a comparison with frequentist LPs without regularization.

Let's apply the insight that a longer IRF horizon calls for longer lags to revisit Jordà et al.'s (2024) supporting VAR evidence discussed above. Figure 4 shows that re-estimating the monthly Bayesian VAR with a much richer lag structure of $p = 60$ makes their result go away: there is no long-run effect of a 100 basis point monetary policy shock on industrial production. The median response is pretty persistent and reaches a trough only four years after the shock, but the effect becomes statistically insignificant after 9 months. The same holds true when setting the lag length to $p = 48$ and $p = 72$. This is a powerful illustration that the lag choice can exert a major influence on economic conclusions!

Figure 4. Long-lag Bayesian VARs



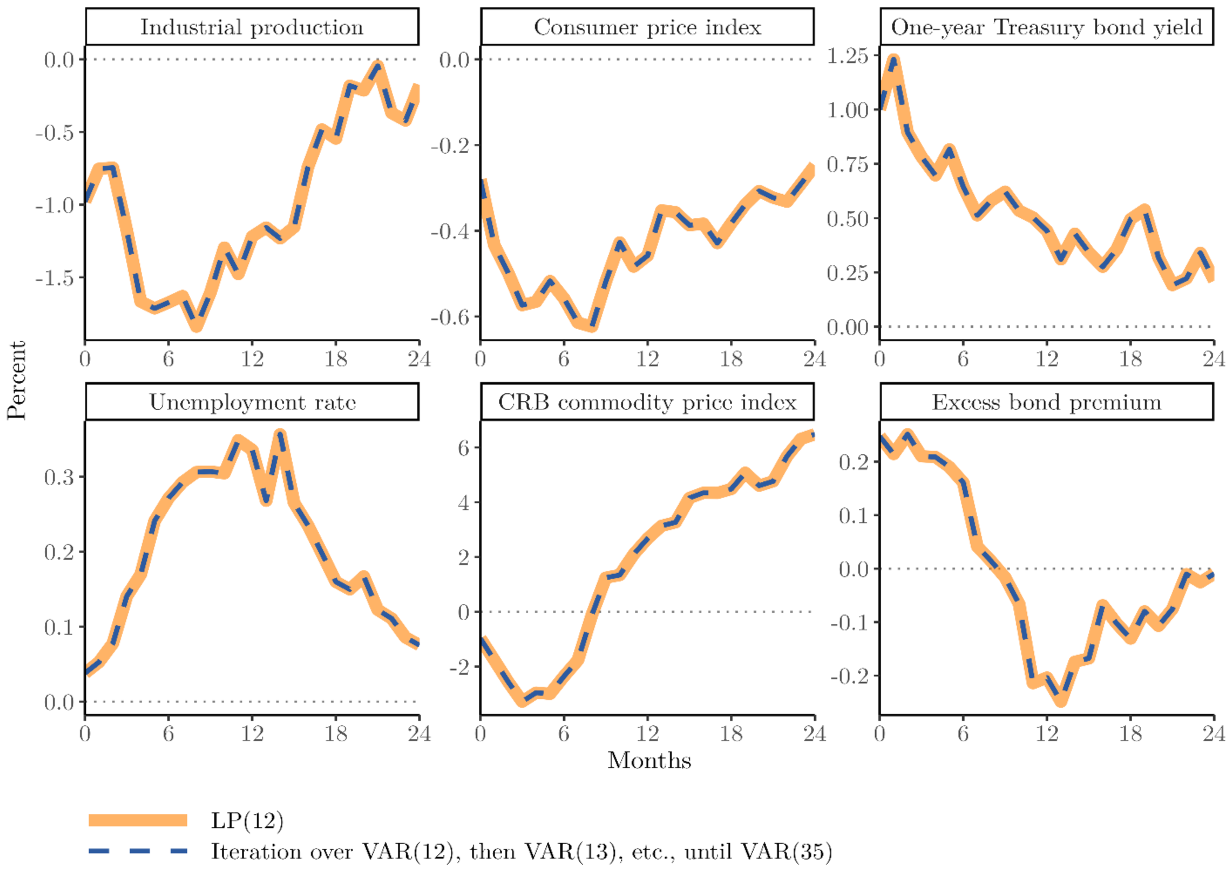
Notes. Impulse responses from the six-variable monthly Bayesian VAR model described in Figure 2 with different lag lengths. Shaded areas, dash-dotted blue lines, and dotted red lines indicate 90 percent posterior coverage bands for the VAR with $p = 60$, $p = 48$, and $p = 72$, respectively.

MODIFIED LESSON 2. *If the focus is on causal effects at long(er) horizons, extend the lag length of the VAR to ensure that richer dynamics do not change your economic results and to strike a good balance between bias, coverage, and accuracy. To deal with the high dimensionality of the VAR, rely on additional information in the form of Bayesian priors for estimation. Prior information can take the form of shrinkage of reduced-form dynamics and economic knowledge about the probability of long-run (non-)neutrality of structural shocks on particular outcome variables.*

1.2 The Anatomy of LP(p) IRFs: A Novel Theoretical Result

On the theoretical front, Ludwig (2024) offers a breakthrough in understanding the link between LPs and VARs in finite samples that resolves the bias-variance tradeoff encountered in applied work. He demonstrates that LPs consist of a series of one-step-ahead predictions made by VARs with an increasing number of lags as the horizon lengthens, implying that there exists an exact mapping between the two estimation methods in finite samples. That is, it is possible to reproduce any LP impulse response using a sequence of VAR forecasts (and vice versa). Figure 5 illustrates this equivalence for the six-variable monetary policy example discussed above where impulse responses from an LP(12) are replicated by iterating over sequentially expanding VAR(p) models with p increasing from 12 to 35 lags over a two-year horizon.

Figure 5. Ludwig (2024) Equivalence of LP and VAR of Increasing Order



Given that LPs are VARs of increasing order, an LP(p) is more highly parameterized relative to a VAR(p) which means that effectively LPs estimate more VAR parameters than a VAR with the same lag length. This brilliant discovery solves the bias-variance tradeoff riddle in a very straightforward and intuitive way: an LP(p) is less parsimonious than a VAR(p) which is why the variance is greater; at the same time, an LP(p) is less restrictive than a VAR(p) which is why it is less biased. This implies that the comparison of an LP(p) and a VAR(p) really is a comparison of a large versus a small model and not the estimation method per se. In other words, LPs are not a fundamentally different class of estimators; rather, they are built from VARs— just of different lag lengths for each horizon. Understanding this explicit relationship sheds light on some of the puzzling features of LPs. For example, the jaggedness of LP-based IRFs can be traced back to LPs being mongrels of VARs of different order.

Ludwig's (2024) theorem has important implications for model selection and empirical practice. Most notably, it invalidates the standard comparison between LP(p) and VAR(p), since these models differ substantially in complexity. This asymmetric treatment explains much of the observed bias-variance tradeoff. Thus, for a proper evaluation of the econometric properties of both model frameworks, it is necessary to put them on "equal footing" first. Ludwig (2024) accomplishes this by introducing the concept of model "size", defined as the average number of estimated coefficients per equation across impulse response horizons, since what matters is not the number of estimated parameters in each model framework or the "order" but rather the number of parameters it takes to rewrite one model into the other. This idea is based on a simple, yet very powerful counting exercise. Once LPs and VARs are aligned by model size, the supposed tradeoff between bias and variance disappears. Differences in performance across estimation methods can then be attributed to genuine model features, such as misspecification or prior assumptions, rather than an artifact of a misleading comparison.

MODIFIED LESSON 3. *There is no bias-variance tradeoff even in finite samples as established by Ludwig (2024). You can always map a low-order LP into a VAR of increasing order (meaning that, as the horizon extends by one period, you add one more lag to the VAR) which yields the exact same impulse responses.*

2 Applied Practice: Why Things are Not as Simple as They Seem

After having clarified empirically and theoretically that the difference between the two estimation methods essentially boils down to the number of lags, let me turn to reviewing some of Montiel Olea et al.'s practical recommendations for the implementation of LPs and VARs through the lens of the preceding discussion and other recent developments in the literature.

2.1 Selecting the Number of Lags

In their summative advice, Montiel Olea et al. caution against relying on the Akaike information criterion (AIC) or other conventional selection criteria to determine the VAR lag length, while proposing that for LPs the choice of the lag order can be guided by the AIC using an 'auxiliary' VAR model.[6] It is well known that the AIC has a tendency to indicate a number of lags smaller

[6]There is some tension here given that in their simulation exercise the authors also select the lag length for the reduced-form VAR using the AIC. While I understand that they are mimicking applied work to highlight shortcomings, it would be more useful to showcase how to improve upon established practice.

than the standard choice based on the data frequency despite being considered the ‘most generous’ among popular lag selection procedures. The reason is that the AIC is geared toward parsimony which is a useful property for forecasting but harmful for structural analysis since it introduces truncation bias. In light of Ludwig’s (2024) finite-sample equivalence result, it is clear now why a restrictive lag structure favored by the AIC is less of a concern for the performance of LPs given that they are more richly parameterized than a VAR of equal order.

While practical, this recommendation is not entirely satisfactory, not only because it doesn’t provide any guidance to VAR practitioners but, most importantly, because the lag selection criterion is not designed for the purpose of dynamic causal analysis. That’s a point forcefully made in a recent contribution by González-Casasús and Schorfheide (2025) where they stress that model selection criteria intended for optimizing forecasting performance are not useful for selecting the model specification for IRF analysis which focuses on different (structural) dynamics. They devise a novel information criterion tailored to IRF estimation that jointly determines the lag length, Bayesian shrinkage, and estimation method (LP vs VAR) in a data-driven way and is robust to different types of dynamic misspecification. They conclude that their "findings discredit the widespread idea that LPs are *always* preferred under misspecification" (p. 41) and they emphasize that the choice between VAR-based and LP-based IRF estimates should be sample dependent. The latter point echoes Kuersteiner (2005) who proposes data-dependent selection rules for the specification of VAR(p) approximations to VAR(∞) models where p is determined based on information in the sample.

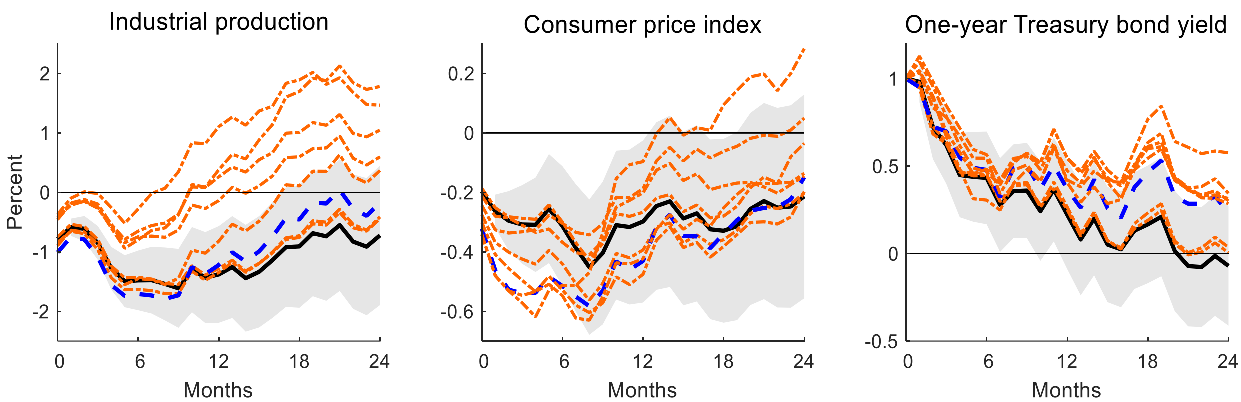
MODIFIED LESSON 4. *Model selection criteria should be targeted at the relevant object of interest. It is advisable to complement statistical procedures for lag length selection with an economic perspective about what a sensible IRF horizon is for the question at hand.*

2.2 Selecting the Predictor Variables

To increase efficiency and robustify LP estimation against dynamic misspecification, Montiel Olea et al. suggest controlling for *strong* predictors of either the outcome variable or the impulse variable (observed shock measure) as part of their recipe for applied researchers. However, they are silent about what qualifies as a ‘strong’ predictor. They once again recommend that practitioners using LPs avail themselves of an ‘auxiliary’ VAR model to decide which variables to include as controls based on the AIC. It is important to note that while the AIC is occasionally used in VAR studies to support the choice of the lag length, it strikes me as extremely rare that VAR modelers rely on the AIC for variable selection. The standard approach in structural VAR analysis is to determine the set of variables such that they accurately capture the basic characteristics of an economic system or market that a researcher wishes to model. Thus, the choice is driven by considerations such as possible transmission channels of shocks, policy reaction functions, institutional features etc that describe the economic environment of interest; explanatory power is typically not a consideration.

The authors also advise to use economic knowledge to select relevant covariates but contend that LP is relatively insensitive to the omission of controls with moderate predictive power. Figure 6 provides evidence to the contrary. Returning to the monetary policy example, Figure 6 reports as the baseline the point estimates of the responses of industrial production, consumer prices, and the one-year interest rate to a 100-basis-point monetary policy shock obtained with an LP(12) together with 90 percent confidence intervals. I then gradually add other controls arguably related to the propagation of monetary policy shocks, such as the unemployment rate, commodity prices, the excess bond premium, business inventories, and the money stock, to the three-variable specification. As can be seen, increasing the number of covariates substantially changes the qualitative and quantitative effects of monetary policy on output and prices, with many of the point estimates falling outside the error bands of the baseline model, especially for industrial production. This illustrates that the inclusion of a different set of control variables can lead to very different results, highlighting the fragility of LPs to the selection of covariates.

Figure 6. The Role of the Selection of Covariates



Notes. Impulse responses obtained with LP(12) with increasing number of covariates. The black solid line refers to the specification with only the three variables shown with the shaded areas being the corresponding 90 percent error bands. The blue dashed lines are the responses from the baseline six-variable specification described in Figure 2. The orange dash-dotted lines are responses from specifications that gradually augment the three-variable specification with one additional variable from four to 11 variables. These variables include: unemployment rate, log of commodity price index, excess bond premium, average weekly hours in manufacturing, term spread (between one-year and ten-year bond yields), business inventories, average earnings in manufacturing, and M2 money stock.

What is often observed in applied work is that controls vary across LPs for different outcome variables without any economic *or* statistical rationale for why this should be the case. As a matter of fact, control variables are too scarcely discussed in LP applications despite the considerable influence they can exert on IRF dynamics. The procedure outlined by Montiel Olea et al. based on the auxiliary VAR also does not seem to guarantee that the same set of controls is selected for different outcome variables. Absent a reliable criterion, there is the temptation to pick and choose what 'works' or what 'looks reasonable'; while VARs are not immune to this criticism either, the issue seems more acute in single-equation LPs.

Given this, a more systematic and robust approach to covariate selection seems warranted. Instead of relying on an auxiliary VAR model, a recent paper by Cha (2025) proposes a statistical algorithm for the selection of covariates based on their predictive power for both the outcome and shock variables in high-dimensional settings directly applied within the LP framework. In particular, Cha (2025) introduces an orthogonal greedy algorithm which follows an iterative procedure that optimally chooses covariates and orders them based on the strength of their explanatory power. This ordering affords the approach additional economic interpretability about the most influential covariates for structural impulse response dynamics, going beyond purely statistical advantages. A comparison with conventional LP reveals that, as the number of controls (and lags) increases, the standard LP estimator displays erratic patterns and large standard errors due to the inclusion of a large set of covariates, while Cha's novel selection technique yields more robust and reliable results as well as narrower confidence bands.

MODIFIED LESSON 5. *LP estimation is less robust to the inclusion of different sets of control variables than commonly believed. Starting from a comprehensive list of variables compiled based on economic theory, the selection of relevant covariates should be done in a systematic way within the LP framework. If explanatory power is the guiding principle, the high-dimensional method proposed by Cha (2025) seems promising.*

2.3 Alternative Strategies for Robustifying IRFs

Much of the discussion about the choice between estimation methods revolves around robustness in the form of accurate coverage in the presence of misspecification. Montiel Olea et al. give the impression that the only way to robustify IRFs and conduct proper uncertainty assessment is to rely on LPs (or equivalently VARs with very long lags which they consider pretty much infeasible, however). But there are alternative options available. I will briefly discuss two of them that I find particularly attractive.

Huber and Marcellino (2024) develop a straightforward method for dealing with potential model misspecification of unknown type in the context of Bayesian VARs by replacing the exact likelihood with a coarsened likelihood. In practical terms this amounts to downweighting the information in the standard likelihood by a parameter that captures the degree of misspecification and is determined by recursively minimizing a predictive loss function. They demonstrate in a simulation exercise that coarsening alleviates biases in estimated IRFs that result from misspecification of different forms leading to more robust structural inference.

Ho, Lubik, and Matthes (2024) propose a flexible approach to combine IRF estimators obtained with LPs and VARs based on optimal linear prediction pools. The idea is to exonerate the applied researcher from having to settle on one particular method for estimating IRFs but rather to draw on the relative strength of each estimator by deriving optimal weights that depend on the entire predictive distribution which are then used to construct horizon- and variable-specific responses. This inclusive, data-driven approach delivers a robust assessment of the propagation of structural shocks by reducing bias due to averaging and accounting for uncertainty across estimation methods.

2.4 Identifying Causal Effects

Montiel Olea et al. sustain that the choice between LPs and VARs is entirely independent of considerations of structural identification. They base this view on the bold assertion that "when it comes to identification, anything you can do with VARs, you can do with LPs" (Lesson 3) which strikes me as overly strong, especially when most of the discussion is framed around the availability of valid external instruments or observed shock proxies. Granted, in earlier work, Plagborg-Møller and Wolf (2021) propose clever ways to map some popular VAR identification strategies into the LP framework including timing assumptions, sign restrictions, and narrative events[7] — once again we have to turn to VARs for guidance! — but this mapping can only go so far and misses two recent developments on the identification front that are tied to thinking about the economy as a system of dynamic interactions which is best represented by structural VAR models. The first is the Bayesian approach to identification that allows researchers to incorporate doubts about the underlying structure of the economy into the estimation of causal effects and the study of the dynamic propagation of shocks (see, e.g., Baumeister and Hamilton, 2015, 2024a; Hamilton, 2025). The second is the increasing desire to combine information from various sources to sharpen identification. Here I'm thinking, for example, of combining zero and sign restrictions with proxy variables (Braun and Brüggemann, 2023; Nguyen, 2025) and sign restrictions with priors on the long run (Baumeister and Hamilton, 2015), augmenting sign restrictions with identification by heteroskedasticity and non-Gaussianity (Drautzburg and Wright, 2023; Carriero, Marcellino, and Tornese, 2024), using priors on elasticities and policy rule coefficients together with impact effects of shocks (Baumeister and Hamilton, 2018, 2019; Belongia and Ireland, 2021; Lukmanova and Rabitsch, 2023), exploiting the heterogeneity of micro units for identification (De Graeve and Karas, 2014; Baumeister and Hamilton, 2024b), and jointly drawing on prior knowledge about elasticities, historical events, shapes of IRFs, among other economic objects (Baumeister, Loria, and Maih, 2025). Thus, there exists a wealth of economic identifying information that we can rely on flexibly within VARs that cannot be (easily) accommodated in LPs.

Taken together, if the goal is to quantify the effects of structural shocks based on observed data, the choice of the estimation method should also be guided by the economic information available for identification. Given Montiel Olea et al.'s emphasis on the need for an accurate characterization of uncertainty as an integral part of dynamic causal analysis, it would seem key to take the uncertainty deriving from identifying assumptions into account when deciding on the econometric framework. VARs offer a consistent and versatile framework for the exact and inexact identification of multiple fundamental drivers of the economy, whereas LPs are more limited in that regard.

[7]The key insight is whether to treat the LP coefficients as (a) structural impulse response coefficients as is the case, for example, when a shock variable or narrative proxy is directly included in the regression or contemporaneous values of covariates are included in line with a recursive identification approach, or (b) reduced-form multipliers which are then combined with VAR-based identification strategies such as sign or long-run restrictions.

MODIFIED LESSON 6. *When it comes to identification, anything you can do with LPs, you can do with VARs — and more! The system approach inherent in structural VAR models offers greater flexibility to incorporate (imperfect) knowledge about economic relationships along multiple dimensions that yields more credible identification and acknowledges uncertainty about the underlying structure. Empirical macroeconomists should put structural identification front and center in the choice of the econometric framework.*

3 The Way Forward (as I see it)

Where does that leave us? I think it is fair to say that the authors and I agree that short-lag VARs give rise to concern when the economic question centers on medium- to long-horizon IRFs but our recommendations for empirical practice differ markedly. Montiel Olea et al.'s suggestion is simply to 'use LPs always and everywhere' — as I hope I have demonstrated, while LPs might seem the easy way out, their use in applied work is not as simple and unproblematic as the authors make it sound. While LPs are hailed as more general and more resilient to misspecification of different nature, they are just as susceptible to modeling choices as alternative approaches albeit along different dimensions. LPs are also a pretty needy estimation method as they are heavily reliant on VARs for specification choices (both lag length and variable selection), inferential procedures (for example, the bootstrap implementation to construct percentile-t confidence intervals; see Appendix B.1), and identification strategies, often in an ad-hoc way. In contrast, VARs offer a coherent framework where all these things can be accomplished simultaneously. Even more, VARs are closer to theoretical models that researchers like to use as organizing frameworks for how the economy works as a whole and how macroeconomic variables interact to produce equilibrium outcomes. Stand-alone LPs are a crude device for studying macroeconomic dynamics and fall short of helping us understand the economy as a system of structural relationships.

If we are serious about dynamic causal analysis and the uncertainty that arises from the limited information contained in the data, then the choice of estimation framework cannot be separated from the identification strategy. Given the fundamental role that identifying assumptions play in IRF analysis, I conclude that VARs, particularly when combined with Bayesian methods to handle incomplete identification, remain the superior framework. My hope is that the six modified lessons I put forward are of some help to applied researchers who are willing to weigh the appeal of simplicity more carefully against economic considerations which are key to good empirical work. The brilliant insights by Ludwig (2024) that make the finite-sample bias-variance tradeoff between LPs and VARs obsolete, paired with a Bayesian approach to estimation and identification in combination with the model selection criteria recently proposed by González-Casasús and Schorfheide (2025) provide the practical tools for putting VARs "back in business".

References

Antolín-Díaz, Juan, and Paolo Surico (2025). "The Long-Run Effects of Government Spending," *American Economic Review*, forthcoming.

Baumeister, Christiane, and James D. Hamilton (2015). "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica* 83(5): 1963-1999.

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 Identification: Revisiting the Role of Oil Supply and Demand Shocks," *American Economic Review*, 109(5): 1873-1910.

Baumeister, Christiane, and James D. Hamilton (2024a). "Advances in Using Vector Autoregressions to Estimate Structural Magnitudes," *Econometric Theory* 40(3): 472-510.

Baumeister, Christiane, and James D. Hamilton (2024b). " Uncovering Disaggregated Oil Market Dynamics: A Full-Information Approach to Granular Instrumental Variables," UCSD and Notre Dame.

Baumeister, Christiane, Francesca Loria, and Junior Maih (2025). "Flexible Priors and Restrictions for Structural Vector Autoregressions," Notre Dame, *mimeo*.

Belongia, Michael T., and Peter N. Ireland (2021). "A Classical View of the Business Cycle," *Journal of Money, Credit and Banking* 53(2-3): 333-366.

Braun, Robin, and Ralf Brüggemann (2023). "Identification of SVAR Models by Combining Sign Restrictions With External Instruments," *Journal of Business and Economic Statistics* 41(4): 1077-1089.

Carriero, Andrea, Massimiliano Marcellino, and Tommaso Tornese (2024). "Blended Identification in Structural VARs," *Journal of Monetary Economics* 146: 103581.

Cha, Jooyoung (2025). "Local Projections Inference with High-dimensional Covariates without Sparsity," Vanderbilt University.

De Graeve, Ferre, and Alexei Karas (2014). "Evaluating Theories of Bank Runs with Heterogeneity Restrictions," *Journal of the European Economic Association* 12(4): 969-996.

De Graeve, Ferre, and Andreas Westermark (2013). "Un-truncating VARs," Sveriges Riksbank Working Paper 271.

Drautzburg, Thorsten, and Jonathan H. Wright (2023). "Refining Set-Identification in VARs through Independence," *Journal of Econometrics* 235(2): 1827-1847.

Giannone, Domenico, Michele Lenza, and Giorgio Primiceri (2015). "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics* 97(2): 436-451.

González-Casasús, Oriol, and Frank Schorfheide (2025). "Misspecification-Robust Shrinkage and Selection for VAR Forecasts and IRFs," NBER Working Paper 33474.

Hamilton, James D. (2025). *Vector Autoregressions.* Cambrigde University Press, forthcoming.

Ho, Paul, Thomas A. Lubik, and Christian Matthes (2024). "Averaging Impulse Responses Using Prediction Pools," *Journal of Monetary Economics*, 146: 103571.

Huber, Florian, and Massimiliano Marcellino (2024). "Coarsened Bayesian VARs," Bocconi University and University of Salzburg.

Jordà, Òscar, Sanjay R. Singh, and Alan M. Taylor (2024). "The Long-Run Effects of Monetary Policy," *Review of Economics and Statistics*, forthcoming.

Kuersteiner, Guido (2005). "Automatic Inference for Infinite Order Vector Autoregressions," *Econometric Theory* 21(1): 85-115.

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf (2024). "Local Projections vs. VARs: Lessons from Thousands of DGPs," *Journal of Econometrics* 244: 105722.

Ludwig, Julian F. (2024). "Local Projections are VAR Predictions of Increasing Order," Texas Tech University.

Lukmanova, Elizaveta, and Katrin Rabitsch (2023). "Evidence on Monetary Transmission and the Role of Imperfect Information: Interest Rate versus Inflation Target Shocks," *European Economic Review* 158: 104557.

Miranda-Agrippino, Silvia, and Giovanni Ricco (2021). "The Transmission of Monetary Policy Shocks," *American Economic Journal: Macroeconomics* 13(3): 74-107.

Montiel Olea, José Luis, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf (2024). "Doubly Robust Local Projections and Some Unpleasant VARithmetic," Princeton University.

Nguyen, Lam (2025). "Bayesian Inference in Proxy SVARs with Incomplete Identification: Reevaluating the Validity of Monetary Policy Instruments," Citigroup Inc.

Plagborg-Møller, Mikkel, and Christian K. Wolf (2021). "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica* 89(2): 955-980.

Smets, Frank, and Rafael Wouters (2007). "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review* 97(3): 586-606.