Double Robustness of Local Projections and Some Unpleasant VARithmetic*

José Luis Montiel Olea
Cornell University

Mikkel Plagborg-Møller
Princeton University

Eric Qian
Princeton University

Christian K. Wolf
MIT & NBER

November 11, 2025

Abstract: We consider impulse response inference in a locally misspecified vector autoregression (VAR) model. The conventional local projection (LP) confidence interval has correct coverage even when the misspecification is so large that it can be detected with probability approaching 1. This result follows from a "double robustness" property analogous to that of popular partially linear regression estimators. By contrast, the conventional VAR confidence interval with short-to-moderate lag length can severely undercover for misspecification that is small, difficult to detect statistically, and cannot be ruled out based on economic theory. The VAR confidence interval has robust coverage if, and only if, the lag length is so large that the interval is as wide as the LP interval.

Keywords: bias-aware inference, double robustness, local projection, misspecification, structural vector autoregression. *JEL codes:* C22, C32.

*Email: montiel.olea@gmail.com, mikkelpm@princeton.edu, ericqian@princeton.edu, and ckwolf@mit.edu. We received helpful comments from three anonymous referees, Isaiah Andrews, Tim Armstrong, Joel Flynn, Jim Hamilton, Lars Peter Hansen, Òscar Jordà, Lutz Kilian, Michal Kolesár, Ulrich Müller, Pablo Ottonello, Frank Schorfheide, Harald Uhlig, Mark Watson, Ke-Li Xu, and seminar participants at several venues. Nelson Matthew Tan provided excellent research assistance. Plagborg-Møller acknowledges that this material is based upon work supported by the NSF under Grant #2238049 and by the Alfred P. Sloan Foundation, and Wolf does the same for NSF Grant #2314736.

1 Introduction

In recent years, local projection (LP) estimators of impulse response functions have become a very popular alternative to structural vector autoregressions (henceforth interchangeably referred to as VAR or SVAR, Sims, 1980). In addition to their simplicity, one potential explanation for the popularity of LPs is their perceived robustness to misspecification, as claimed by Jordà (2005) in his seminal article that proposed the estimation method:

> "*[T]hese projections are local to each forecast horizon and therefore more robust [than VARs] to misspecification of the unknown DGP.*"

While this sentiment has been echoed in influential reviews (e.g., Ramey, 2016; Nakamura and Steinsson, 2018; Jordà, 2023), there so far exist essentially no theoretical results on the relative robustness of LP and VAR inference procedures to misspecification. Plagborg-Møller and Wolf (2021) and Xu (2023) show that the two estimators are in fact asymptotically equivalent—and thus equally robust to misspecification—in a general VAR(∞) model if the estimation lag length diverges to infinity with the sample size. However, this result does not directly speak to the empirically relevant case where researchers employ small-to-moderate lag lengths to preserve degrees of freedom. Applied researchers must therefore base their choice of inference procedure on empirically calibrated simulation studies (Kilian and Kim, 2011; Li, Plagborg-Møller, and Wolf, 2024).

In this paper we provide a formal proof of Jordà's claim that conventional LP confidence intervals for impulse responses are surprisingly robust to misspecification. On the other hand, VAR confidence intervals are robust if, *and only if*, they are as wide as LP intervals asymptotically, as is the case when they control for a large number of lags. If the confidence interval is shorter, then it is *necessarily* unreliable.

We consider a large class of stationary data generating processes (DGPs) that are well approximated by a finite-order SVAR model, but subject to local misspecification in the form of an asymptotically vanishing moving average (MA) process, of potentially infinite order. This class is consistent with essentially all linearized structural macroeconomic models and covers many types of dynamic misspecification, such as under-specification of the lag length, failure to include relevant control variables, inappropriate aggregation, and measurement error. Intuitively, with this set-up we capture the idea that finite-order VAR models provide a good but imperfect approximation of reality.

In this setting, we prove that the conventional LP confidence interval has correct (point-wise) asymptotic coverage even for local misspecification that is of such a large magnitude that it can be detected with probability 1 in large samples. This robustness property requires that we control for those lags of the data that are strong predictors of the outcome or impulse variables, but—crucially for applied work—the omission of lags with small-to-moderate predictive power does not threaten coverage. We argue that our result can be interpreted as a consequence of the *double robustness* of the LP estimator, which is analogous to the double robustness of modern partially linear regression estimators in the literature on debiased machine learning.[1]

In stark contrast to LP, even small amounts of misspecification can cause conventional VAR confidence intervals for impulse responses to suffer from severe undercoverage. We first derive analytically the worst-case bias and coverage of VARs over all possible misspecification processes, subject to a constraint on the overall magnitude of the misspecification. The worst-case bias and coverage distortion are small if, *and only if*, the asymptotic variance is close to that of LP. In general, the only way to guarantee robustness of conventional VAR inference is thus to include so many lags that the VAR estimator is asymptotically equivalent with LP. If instead the VAR confidence interval is much shorter (as is typically the case in applied practice), then it will severely undercover even for a misspecification term that: (i) is small in magnitude; (ii) has dynamic properties that cannot be ruled out *ex ante* based on economic theory; and (iii) is difficult to detect *ex post* with model specification tests. Instead of increasing the lag length, coverage can also be restored by using a larger bias-aware critical value (Armstrong and Kolesár, 2021), but we show that the resulting confidence intervals are so wide that one may as well report the LP interval.

We demonstrate the practical relevance of our theoretical results through a comprehensive review of current practice in the applied VAR literature, together with a simulation study. In papers published in top economics journals, researchers tend to select small to moderate lag lengths, and often report impulse responses at horizons far exceeding the lag length. Our theory suggests this practice is likely to render inference vulnerable to misspecification. To substantiate this conclusion, we simulate data calibrated to the oil shock application in Känzig (2021). The DGP is taken to be a VAR estimated on the paper's actual data, but with 18 lags rather than 12. The VAR confidence interval materially undercovers—particularly at medium and long horizons—if the lag length is set to 12 or selected by AIC, in line with applied practice, while the LP interval attains close to nominal coverage. Increasing the

[1]Important contributions include Robins, Mark, and Newey (1992), Robins and Rotnitzky (1995), Robins, Rotnitzky, and van der Laan (2000), Robins and Rotnitzky (2001), Bang and Robins (2005), Ai and Chen (2007), Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018), and Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2022).

estimation lag length beyond the conventional choices ameliorates the VAR undercoverage, at the cost of delivering confidence intervals as wide as those of LP.

Literature. Relative to the previously cited simulation studies of LPs and VARs, we here derive *analytical* results on the worst-case asymptotic properties of these two inference procedures that hold for a wide range of stationary, locally misspecified VAR models. The simulations in Li, Plagborg-Møller, and Wolf (2024) suggest a stark bias-variance trade-off between LP (low bias, high variance) and moderate-lag VAR estimators (moderate bias, low variance). The reason behind the theoretical superiority of LP proved in this paper is that, if the objective is to construct confidence intervals with robust coverage for a wide range of DGPs, then even a moderate amount of VAR bias cannot be tolerated, as it causes the VAR confidence interval to be poorly centered. A concern for correct confidence interval coverage thus effectively induces a large weight on bias in the researcher's objective function, justifying the use of LP despite its higher variance.

The robustness of LPs to misspecification discussed here—with stationary data and at fixed horizons—is conceptually and theoretically distinct from the robustness of LPs to the persistence in the data and the length of the impulse response horizon shown by Montiel Olea and Plagborg-Møller (2021). Nevertheless, it turns out that controlling for lags ("lag augmentation") is key to all the robustness properties established in Montiel Olea and Plagborg-Møller (2021) and in the present paper.

We also build upon previous research into misspecified VAR models, uncovering novel results about the robustness of LPs and the worst-case properties of VAR procedures. Braun and Mittnik (1993) derive expressions for the probability limits of VAR estimators under global MA misspecification; however, since bias always dominates variance asymptotically in their framework, they do not characterize the properties of LP and VAR inference procedures, which is the focus of our paper. Schorfheide (2005) characterizes the asymptotic mean squared errors of iterated and direct multi-step forecasts in a reduced-form VAR model with MA terms of order $T^{-1/2}$, and González-Casasús and Schorfheide (2025) use this framework to select hyperparameters for VAR forecasts. Müller and Stock (2011) construct Bayesian forecast intervals in a locally misspecified univariate AR model. Relative to these papers, we here contribute by: (i) focusing on structural impulse responses rather than forecasting; (ii) allowing for more general rates of local misspecification, which is key to uncovering the double robustness of LP; and (iii) deriving simple analytical formulae for worst-case bias and coverage of VARs. As such, our results formalize concerns by applied practitioners about the lack of VAR robustness and sensitivity to lag length (Chari, Kehoe, and McGrattan, 2008; Nakamura and Steinsson, 2018; see also Inoue and Kilian, 2002, and Kilian and Lütkepohl, 2017, Chapters 2.6.5 and 6.2).

Whereas our paper deals with bias imparted by dynamic misspecification, the analysis does not capture other familiar sources of small-sample bias. In particular, our asymptotics abstract from the order-T^{-1} biases that arise from (i) persistence in the data (Pope, 1990; Kilian, 1998; Herbst and Johannsen, 2024) and (ii) the nonlinearity of the impulse response transformation of the VAR parameters (Jensen's inequality).

Outline. Section 2 defines the local-to-SVAR model as well as the LP and VAR estimators. Section 3 proves the robustness of LP and the fragility of VAR confidence intervals. Section 4 derives analytically the worst-case bias and coverage of VARs, and shows that bias-aware VAR confidence intervals tend to be wider than the LP interval. Section 5 demonstrates the practical relevance of our theoretical results through a review of the applied literature and a simulation study. Section 6 concludes. Replication materials are available online.[2]

Notation. All asymptotic limits are taken as the sample size $T \to \infty$ and are *pointwise* in the sense of fixing the true model parameters and the impulse response horizon. A sum $\sum_{\ell=a}^{b} c_\ell$ is defined to equal 0 when $a > b$.

2 Framework

We start out by defining the model and estimators.

2.1 Model and assumptions

Extending the forecasting model of Schorfheide (2005), we consider a multivariate, stationary structural VARMA$(1, \infty)$ model that is local to an SVAR(1) model:

$$y_t = Ay_{t-1} + H[I + T^{-\zeta}\alpha(L)]\varepsilon_t, \quad \text{for all } t, \tag{2.1}$$

where the data vector $y_t = (y_{1,t}, \ldots, y_{n,t})'$ is n-dimensional, the shock vector $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{m,t})'$ is m-dimensional, A is an $n \times n$ matrix, H is an $n \times m$ matrix, $\alpha(L) = \sum_{\ell=1}^{\infty} \alpha_\ell L^\ell$ is an $m \times m$

[2] https://github.com/ckwolf92/lp_var_inference

lag polynomial, and T denotes the sample size. We allow the number of shocks m to potentially exceed the number of variables n, and *vice versa*. We show below that equation (2.1) encompasses local-to-SVAR models with $p > 1$ lags by writing them in companion form.

The model (2.1) captures the idea that the time series dynamics of the data are well approximated by an autoregressive model driven by unobserved white noise shocks ε_t, but with a small amount of misspecification in the form of an MA process $T^{-\zeta}\alpha(L)\varepsilon_t$. The misspecification is asymptotically small in the sense that the MA coefficients converge to zero at the rate $T^{-\zeta}$, though the misspecification may still affect the properties of estimators, as shown by Schorfheide (2005) and as demonstrated below. We argue below that MA misspecification of this form can capture many empirically relevant types of dynamic misspecification. We consider local rather than global misspecification in the spirit of local power analysis (e.g., Rothenberg, 1984), since this makes the bias-variance trade-off between the VAR and LP estimators matter even asymptotically as the sample size T diverges, allowing us to make tractable analytical comparisons between these two procedures.

The parameter of interest is the response at horizon h of the variable $y_{i^*,t}$ with respect to the shock $\varepsilon_{j^*,t}$ for some indices $i^* \in \{1,\dots,n\}$ and $j^* \in \{1,\dots,m\}$, to be defined below.

Assumption 2.1. *For each T, $\{y_t\}_{t\in\mathbb{Z}}$ is the stationary solution to equation (2.1), given the following restrictions on parameters and shocks:*

i) $\varepsilon_t \overset{i.i.d.}{\sim} (0_{m\times 1}, D)$, *where* $D \equiv \operatorname{diag}(\sigma_1^2,\dots,\sigma_m^2)$, *and the elements of ε_t are mutually independent. For all $j = 1,\dots,m$, $\sigma_j^2 > 0$ and $E(\varepsilon_{j,t}^4) < \infty$.*

ii) All eigenvalues of A are strictly below 1 in absolute value.

iii) The first j^* rows of H are of the form $(\tilde{H}, 0_{j^*\times(m-j^*)})$, where $\tilde{H}$ is a $j^* \times j^*$ lower triangular matrix with 1's on the diagonal. In particular, we require $j^* \le n$.

iv) $S \equiv \operatorname{Var}(\tilde{y}_t)$ *is non-singular, where* $\tilde{y}_t \equiv (I - AL)^{-1}H\varepsilon_t$ *is the stationary solution to* (2.1) *when* $\alpha(L) = 0$. *Specifically,* $\operatorname{vec}(S) = (I - A\otimes A)^{-1}\operatorname{vec}(\Sigma)$, *where* $\Sigma \equiv HDH'$.

v) $\alpha(L)$ *is absolutely summable.*

vi) $\zeta > 0$.

The assumptions that the shocks are mutually and serially independent are made to simplify the exposition; we prove formally in Supplemental Appendix C.1 that our results on the robustness of LP and on the asymptotic bias of VAR go through for a large class of conditionally heteroskedastic shock processes. The assumptions on H correspond to recursive (also known as Cholesky) identification of the shock of interest $\varepsilon_{j^*,t}$, with a unit effect normalization $H_{j^*,j^*} = 1$. A special case is when the shock is directly observed, which corresponds to ordering it first (i.e., $j^* = 1$). Supplemental Appendix C.2 shows that our results extend also to identification via external instruments or proxies (Stock and Watson, 2018). Absolute summability of $\alpha(L)$ is a weak regularity condition ensuring the vector MA(∞) process $\alpha(L)\varepsilon_t$ is well-defined (Brockwell and Davis, 1991, Proposition 3.1.1). Finally, the assumption that $\zeta > 0$ restricts attention to local misspecification, as discussed earlier.

The impulse response of interest is defined as

$$\theta_{h,T} \equiv e'_{i^*,n} \left(A^h H + T^{-\zeta} \sum_{\ell=1}^{h} A^{h-\ell} H \alpha_\ell \right) e_{j^*,m} = E[y_{i^*,t+h} \mid \varepsilon_{j^*,t} = 1] - E[y_{i^*,t+h} \mid \varepsilon_{j^*,t} = 0],$$

where $e_{i,n}$ denotes the n-dimensional unit vector with a 1 in position i. The first term in the parenthesis is the usual VAR impulse response formula, while the second term arises from the MA component. Importantly, and consistent with our focus on the consequences of dynamic misspecification, we do not treat the VAR misspecification as non-classical measurement error that should be ignored for structural analysis; instead, the true causal model has a VARMA form (with small but potentially non-zero MA terms), and we care about the full transmission mechanism of shocks in this model.

Additional lags. Our framework covers local-to-SVAR(p) models of the form

$$\check{y}_t = \sum_{\ell=1}^{p} \check{A}_\ell \check{y}_{t-\ell} + \check{H}[I + T^{-\zeta}\alpha(L)]\varepsilon_t, \tag{2.2}$$

where $\check{y}_t$ is $\check{n}$-dimensional, the $\check{A}_\ell$ matrices are $\check{n} \times \check{n}$, and $\check{H}$ is $\check{n} \times m$ and satisfies Assumption 2.1(iii). This fits into the original model (2.1) if we set $n = \check{n}p$ and define the companion form representation

$$y_t = \begin{pmatrix} \check{y}_t \\ \check{y}_{t-1} \\ \check{y}_{t-2} \\ \vdots \\ \check{y}_{t-p+1} \end{pmatrix}, \quad A = \begin{pmatrix} \check{A}_1 & \check{A}_2 & \dots & \check{A}_{p-1} & \check{A}_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}, \quad H = \begin{pmatrix} \check{H} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In particular, we can allow the estimation lag length p to exceed the true minimal lag length p_0 of the model by setting $\check{A}_\ell = 0$ for $\ell > p_0$. This fact will prove useful when we consider what happens as the lag length of the estimated VAR is increased.

Types of misspecification. Our local-to-SVAR model (2.1) with MA misspecification covers several empirically relevant types of model misspecification. While essentially all modern discrete-time, linearized macro models have VARMA representations, they usually cannot be represented exactly as finite-order VAR models (e.g., Kilian and Lütkepohl, 2017, Chapter 6.2). Even if the true DGP were a finite-order VAR, dynamic misspecification of the estimation model can give rise to MA terms, for example due to under-specification of the lag length or failing to control for some of the variables in the true system. Relatedly, MA terms may appear because of a failure of invertibility of the shocks (Alessi, Barigozzi, and Capasso, 2011). VARMA representations can also arise from temporal or cross-sectional aggregation of finite-order VAR models, including contamination by classical measurement error (Granger and Morris, 1976; Lütkepohl, 1984). In all of these cases, if the number of lags used for estimating the VAR is chosen to be sufficiently large, then the MA remainder will be small, consistent with the spirit of our locally misspecified model (2.1).

In terms of structural shock identification, our framework accommodates both the case of a well-identified shock (or instrument/proxy, see Supplemental Appendix C.2) but misspecification in other parts of the model, as well as misspecification in the structural shock identification itself. Key to this generality is that we allow the $m \times m$ MA polynomial $\alpha(L)$ to be arbitrary. To see this, consider the case $j^* = 1$, so interest centers on the dynamic causal effects of the first shock $\varepsilon_{1,t}$. If the first row of $\alpha(L)$ is zero, then $\varepsilon_{1,t}$ is well-identified as the reduced-form residual in the first equation of the VAR. If the first row of $\alpha(L)$ is non-zero, then the reduced-form residual will be contaminated by lagged shocks, thus allowing for the possibility that shock identification itself is not entirely accurate.

2.2 Estimators

We consider two estimators of the impulse response $\theta_{h,T}$ using the data $\{y_t\}_{t=1}^T$:

1. The *LP* estimator is the coefficient $\hat{\beta}_h$ in a regression of $y_{i^*,t+h}$ on $y_{j^*,t}$, controlling for $\underline{y}_{j^*,t} \equiv (y_{1,t}, \dots, y_{j^*-1,t})'$ (i.e., the variables ordered before $y_{j^*,t}$, if any) and lagged data:

$$y_{i^*,t+h} = \hat{\beta}_h y_{j^*,t} + \hat{\omega}_h' \underline{y}_{j^*,t} + \hat{\gamma}_h' y_{t-1} + \hat{\xi}_{i^*,h,t}, \tag{2.3}$$

where $\hat{\xi}_{i^*,h,t}$ is the least-squares residual. Recall from the previous subsection that if we are estimating an SVAR(p) specification in the data $\check{y}_t$, then the vector y_{t-1} actually contains p lags $\check{y}_{t-1}, \dots, \check{y}_{t-p}$.

2. The *VAR* estimator is defined as the response of $y_{i^*,t+h}$ with respect to the j^*-th recursively orthogonalized innovation, where the magnitude of the innovation is normalized such that $y_{j^*,t}$ increases by one unit on impact:

$$\hat{\delta}_h \equiv e'_{i^*,n} \hat{A}^h \hat{\nu},$$

where

$$\hat{A} \equiv \left(\sum_{t=2}^{T} y_t y'_{t-1} \right) \left(\sum_{t=2}^{T} y_{t-1} y'_{t-1} \right)^{-1}, \quad \hat{\nu} \equiv \hat{C}^{-1}_{j^*,j^*} \hat{C}_{\bullet,j^*},$$

and $\hat{C}_{\bullet,j^*}$ is the j^*-th column of the lower triangular Cholesky factor $\hat{C}$ of the covariance matrix $\hat{\Sigma} \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}'_t = \hat{C}\hat{C}'$ of the residuals $\hat{u}_t \equiv y_t - \hat{A} y_{t-1}$. Again, in the case of an SVAR(p) specification, the above formulae operate on the companion form.

Note that the two estimators coincide at the impact horizon: $\hat{\beta}_0 = \hat{\delta}_0$ (see Lemma E.5 in Supplemental Appendix E).

It is well known that conventional confidence intervals based on both these estimators would have correct asymptotic coverage in a well-specified VAR model. However, the presence of the additional MA term in the model (2.1) means that, in principle, both the LP and VAR estimators ought to control for infinitely many lags of the data, rather than just one. Nevertheless, as we will now establish, this dynamic misspecification has much more serious consequences for the VAR procedure than for LP.

3 Robust local projections, fragile VARs

This section shows that the conventional LP confidence interval is robust to large amounts of misspecification. In contrast, the conventional VAR confidence interval has fragile coverage, except when it is asymptotically as wide as the LP interval, as will be the case with sufficiently large lag length.

3.1 Large-sample distributions and confidence interval coverage

We begin by characterizing the large-sample distributions of the LP and VAR estimators.

The robustness of LPs. Our first main result establishes that the large-sample distribution of the LP estimator is unaffected by large amounts of misspecification.

Proposition 3.1. *Under Assumption 2.1,*

$$\hat{\beta}_h - \theta_{h,T} = \frac{1}{\sigma_{j^*}^2} \frac{1}{T} \sum_{t=1}^{T} \xi_{i^*,h,,t} \varepsilon_{j^*,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2}),$$

where

$$\xi_{h,t} = (\xi_{1,h,t}, \dots, \xi_{n,h,t})' \equiv A^h \overline{H}_{j^*} \overline{\varepsilon}_{j^*,t} + \sum_{\ell=1}^{h} A^{h-\ell} H \varepsilon_{t+\ell},$$

with $\overline{H}_{j^*} \equiv (H_{\bullet,j^*+1}, \dots, H_{\bullet,m})$ *and* $\overline{\varepsilon}_{j^*,t} \equiv (\varepsilon_{j^*+1,t}, \dots, \varepsilon_{m,t})'$.

Proof. See Appendix B.1. □

The result implies that the first-order asymptotic behavior of LP does not depend on the misspecification parameter $\alpha(L)$, provided $\zeta > 1/4$ so that $O_p(T^{-2\zeta}) = o_p(T^{-1/2})$. Though this robustness property of LP is with respect to local (i.e., asymptotically vanishing) misspecification, it is still quantitatively meaningful, given that MA terms of order $T^{-\zeta}$ with $\zeta \in (1/4, 1/2)$ in the model (2.1) can be detected with probability 1 asymptotically by conventional VAR model specification tests, such as the Hausman test considered in Section 3.2.

Why is LP robust to misspecification of such large magnitude? We will offer two mathematically equivalent pieces of intuition, with our discussion throughout deliberately heuristic. The classic omitted variable bias (OVB) formula suggests that the bias of the LP impulse response estimator $\hat{\beta}_h$ in the regression (2.3) is proportional to the product of two factors: (i) the direct effect of omitted lags on $y_{i^*,t+h}$, and (ii) the covariance of the residualized regressor of interest $y_{j^*,t} - E[y_{j^*,t} \mid \underline{y}_{j^*,t}, y_{t-1}]$ with the omitted lags. The factor (i) is of order $T^{-\zeta}$ in our local-to-SVAR model (2.1). The factor (ii) is also of order $T^{-\zeta}$, since the residualized regressor equals $\varepsilon_{j^*,t} + O_p(T^{-\zeta})$ under Assumption 2.1(iii), and the shock $\varepsilon_{j^*,t}$ is uncorrelated with any lagged data. Hence, the OVB is of order $T^{-2\zeta}$, so when $\zeta > 1/4$, the bias of the estimator is negligible relative to the standard deviation (which is of order $T^{-1/2}$, as in the correctly specified case). This argument relies on the LP regression controlling for the most important lags of the data (i.e., y_{t-1}); without lagged controls, one or both factors in the OVB formula may not be small (González-Casasús and Schorfheide, 2025).

The preceding intuition is a special case of the *double robustness* property of partially linear regressions, see Example 1.1 in Chernozhukov et al. (2018) and Example 1 in Chernozhukov et al. (2022). We will now argue that this property applies also to LP, again settling for a heuristic argument. For notational simplicity, set $j^* = 1$ so $\underline{y}_{j^*,t} = 0$. Consider any dynamic model (for example a VARMA(p, q)) that implies the following LP representation:

$$y_{i^*,t+h} = \theta_{0,h} y_{1,t} + \gamma_0(y^{t-1}) + \xi_{i^*,h,t}, \quad \text{where} \quad \xi_{i^*,h,t} \perp\!\!\!\perp y^t \equiv (y_t, y_{t-1}, \dots).$$

Here $\theta_{0,h}$ is the true impulse response, $\gamma_0(\cdot)$ is a function of lagged data, and "$\perp\!\!\!\perp$" signifies independence. Define $\nu_0(y^{t-1}) \equiv E[y_{1,t} \mid y^{t-1}]$. By applying the Frisch-Waugh lemma to the regression (2.3), we see that the LP estimator $\hat{\beta}_h$ is the sample analogue of the solution $\theta_{0,h}$ to the moment condition

$$E[\{y_{i^*,t+h} - \theta_{0,h} y_{1,t} - \gamma_0(y^{t-1})\}\{y_{1,t} - \nu_0(y^{t-1})\}] = 0.$$

If we evaluate the moment on the left-hand side at arbitrary functions $\gamma(\cdot)$ and $\nu(\cdot)$ rather than at the true ones $\gamma_0(\cdot)$ and $\nu_0(\cdot)$, a simple calculation shows that it equals $E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{\nu_0(y^{t-1}) - \nu(y^{t-1})\}]$.[3] Hence, the moment condition is satisfied at the true impulse response parameter $\theta_{0,h}$ as long as *either* $\gamma = \gamma_0$ or $\nu = \nu_0$, making the LP estimator *doubly robust*: it is consistent if we correctly specify either the controls $\gamma(y^{t-1})$ in the outcome equation or the controls $\nu(y^{t-1})$ in the implicit first-stage regression that isolates the shock $\varepsilon_{j^*,t} = y_{j^*,t} - \nu(y^{t-1})$. Because of double robustness, and as argued more generally by Chernozhukov et al. (2018) (and confirmed by our proof), it turns out that estimation error in γ_0 and ν_0 only affects the asymptotic distribution of $\hat{\beta}_h$ through the *product* of the estimation errors $\|\hat{\gamma} - \gamma_0\| \times \|\hat{\nu} - \nu_0\|$. In our local-to-SVAR model (2.1), both terms in this product are of order $T^{-\zeta}$ due to the omitted lags. The product is then of order $T^{-2\zeta}$ and thus asymptotically negligible when $\zeta > 1/4$, consistent with our earlier intuition.

The fragility of VARs. In contrast to LP, the VAR estimator is fragile.

Proposition 3.2. *Under Assumption 2.1,*

$$\hat{\delta}_h - \theta_{h,T} = \operatorname{trace}\left\{S^{-1}\Psi_h H T^{-1} \sum_{t=1}^{T} \varepsilon_t \tilde{y}_{t-1}'\right\} + \frac{1}{\sigma_{j^*}^2} e_{i^*,n}' A^h T^{-1} \sum_{t=1}^{T} \xi_{0,t}\varepsilon_{j^*,t}$$
$$+ T^{-\zeta}\,\mathrm{aBias}(\hat{\delta}_h) + o_p(T^{-1/2} + T^{-\zeta}),$$

[3] We can write the moment as $E[\{y_{i^*,t+h} - \theta_{0,h}y_{1,t} - \gamma_0(y^{t-1}) + \gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{y_{1,t} - \nu(y^{t-1})\}] = E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{y_{1,t} - \nu_0(y^{t-1}) + \nu_0(y^{t-1}) - \nu(y^{t-1})\}]$, since $y_{i^*,t+h} - \theta_{0,h}y_{1,t} - \gamma_0(y^{t-1}) = \xi_{i^*,h,t}$ is independent of y^t (orthogonality would suffice if $\nu(\cdot)$ were linear). The claim now follows from $E[y_{1,t} - \nu_0(y^{t-1}) \mid y^{t-1}] = 0$ by definition of $\nu_0(\cdot)$.

where

$$\text{aBias}(\hat{\delta}_h) \equiv \text{trace}\left\{S^{-1}\Psi_h H \sum_{\ell=1}^{\infty} \alpha_\ell D H'(A')^{\ell-1}\right\} - e'_{i^*,n} \sum_{\ell=1}^{h} A^{h-\ell} H \alpha_\ell e_{j^*,m},$$

$$\Psi_h \equiv \sum_{\ell=1}^{h} A^{h-\ell} H_{\bullet,j^*} e'_{i^*,n} A^{\ell-1},$$

and $\{\tilde{y}_t\}$ *and* S *are defined in Assumption 2.1.*

Proof. See Appendix B.2. $\square$

The convergence rate $T^{-\min\{1/2,\zeta\}}$ of the VAR estimator is weakly slower than the $T^{-\min\{1/2,2\zeta\}}$ rate achieved by LP. This is because the VAR estimator suffers from bias of order $T^{-\zeta}$, while the stochastic terms of order $T^{-1/2}$ are the same as they would be in a correctly specified SVAR(p) model.[4] The VAR bias is only asymptotically negligible if $\zeta > 1/2$, a much smaller degree of robustness than shown above for LP. The case $\zeta = 1/2$ is of particular interest, as then the bias and standard deviation are of the same asymptotic order (see also Schorfheide, 2005). MA terms of order $T^{-1/2}$ can be detected with asymptotic probability strictly between 0 and 1 by specification tests, as will be shown in Section 3.2.

The asymptotic bias is due to two forces: first, the coefficient matrix $\hat{A}$ is biased due to the endogeneity caused by the MA terms, and second, the VAR estimator extrapolates the horizon-h impulse response based on a parametric formula $\hat{A}^h$ that does not hold exactly in the true VARMA model (2.1). This is more easily seen in the special case of a univariate model $y_t = \rho y_{t-1} + [1 + T^{-\zeta}\alpha(L)]\varepsilon_t$ with $n = m = 1$, in which case

$$\text{aBias}(\hat{\delta}_h) \equiv \underbrace{h\rho^{h-1}}_{\frac{\partial(\rho^h)}{\partial\rho}} \underbrace{(1-\rho^2)\sum_{\ell=1}^{\infty}\rho^{\ell-1}\alpha_\ell}_{\text{aBias}(\hat{\rho}) = \frac{\text{Cov}(\alpha(L)\varepsilon_t, \tilde{y}_{t-1})}{\text{Var}(\tilde{y}_{t-1})}} \underbrace{-\sum_{\ell=1}^{h}\rho^{h-\ell}\alpha_\ell}_{\theta_{h,T}-\rho^h},$$

where $\hat{\rho} = \hat{A}$ is the AR(1) coefficient from an OLS regression of y_t on y_{t-1}.[5] While the dynamic responses estimated by the VAR are prone to bias, the shock identification *per se* is not, in the sense that the VAR's estimated impact (horizon-0) response is identical to the doubly robust LP estimate (see Section 2.2).

[4] The first stochastic term captures sampling uncertainty in the reduced-form impulse responses $\hat{A}^h$, while the second term captures uncertainty in the structural impact response vector $\hat{\nu}$.

[5] Lag augmentation of the VAR impulse response estimator as in Inoue and Kilian (2020) may reduce the first term in the bias formula, but it does not affect the second term.

Confidence intervals. The preceding results imply that the conventional LP confidence interval is robust to misspecification while the conventional VAR interval is not. We define the level-$(1-a)$ LP and VAR confidence intervals using the standard formulae:

$$\mathrm{CI}(\hat{\beta}_h) \equiv \left[\hat{\beta}_h \pm z_{1-a/2}\sqrt{\mathrm{aVar}(\hat{\beta}_h)/T}\right], \quad \mathrm{CI}(\hat{\delta}_h) \equiv \left[\hat{\delta}_h \pm z_{1-a/2}\sqrt{\mathrm{aVar}(\hat{\delta}_h)/T}\right]. \tag{3.1}$$

Here $z_{1-a/2}$ is the $1-a/2$ quantile of the standard normal distribution, and $\mathrm{aVar}(\hat{\beta}_h)$ and $\mathrm{aVar}(\hat{\delta}_h)$ are the asymptotic variances of the leading (order-$T^{-1/2}$) stochastic terms in the representations of the LP and VAR estimators in Propositions 3.1 and 3.2; explicit formulae for the asymptotic variances are given in Corollary A.2 in Appendix A.3, which also implies that $\mathrm{aVar}(\hat{\beta}_h) \geq \mathrm{aVar}(\hat{\delta}_h)$. None of the results below would change if we replaced the asymptotic variances with the conventional consistent estimates of these (that assume correct specification, as implemented in standard econometric software packages).[6]

Corollary 3.1. *Under Assumption 2.1 and* $\zeta > 1/4$, $\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = 1-a$. *If moreover* $\mathrm{aVar}(\hat{\delta}_h) > 0$ *and* $\mathrm{aBias}(\hat{\delta}_h) \neq 0$, *then* $\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\delta}_h)) = \lim_{T\to\infty}\{1 - r\left(T^{1/2-\zeta}b_h; z_{1-a/2}\right)\}$, *where* $b_h \equiv \mathrm{aBias}(\hat{\delta}_h)/\sqrt{\mathrm{aVar}(\hat{\delta}_h)}$, $r(b;c) \equiv P_{Z\sim N(0,1)}(|Z+b| > c) = \Phi(-c-b) + \Phi(-c+b)$, *and* $\Phi(\cdot)$ *is the standard normal distribution function.*

Proof. Considering separately the three cases $\zeta \in (1/4, 1/2)$, $\zeta = 1/2$, and $\zeta > 1/2$, the result is an immediate consequence of Propositions 3.1 and 3.2. □

LP robustly controls coverage when $\zeta > 1/4$, while the VAR confidence interval generically has coverage converging to zero for $\zeta \in (1/4, 1/2)$, and strictly below the nominal level $1-a$ for $\zeta = 1/2$. Intuitively, the VAR confidence interval has the right width (the same as in the correctly specified case) but the wrong location due to the bias.

3.2 Hausman misspecification test

To aid in interpreting the magnitude of the local misspecification in our set-up, we consider a Hausman (1978) test of correct specification of the VAR model that compares the VAR and LP impulse response estimates. This test rejects for large values of $\sqrt{T}|\hat{\beta}_h -$

[6] Under Assumption 2.1, homoskedastic standard errors suffice. For LP, Heteroskedasticity and Autocorrelation Robust inference would generally be required under the weaker assumptions in Supplemental Appendix C.1, though simple heteroskedasticity-robust standard errors suffice under the assumptions discussed by Montiel Olea and Plagborg-Møller (2021) and Xu (2023).

$\hat{\delta}_h|/\sqrt{\operatorname{aVar}(\hat{\beta}_h) - \operatorname{aVar}(\hat{\delta}_h)}$. A test of this kind was proposed by Stock and Watson (2018) in the context of testing for invertibility.

Proposition 3.3. *Impose Assumption 2.1, $\zeta > 1/4$, and $\operatorname{aVar}(\hat{\beta}_h) > \operatorname{aVar}(\hat{\delta}_h) > 0$. Then the asymptotic rejection probability of the Hausman test equals*

$$\lim_{T\to\infty} P\left(\frac{\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|}{\sqrt{\operatorname{aVar}(\hat{\beta}_h) - \operatorname{aVar}(\hat{\delta}_h)}} > z_{1-a/2}\right) = \lim_{T\to\infty} r\left(\frac{T^{1/2-\zeta}b_h}{\sqrt{\operatorname{aVar}(\hat{\beta}_h)/\operatorname{aVar}(\hat{\delta}_h) - 1}}; z_{1-a/2}\right),$$

where b_h and $r(\cdot,\cdot)$ were defined in Corollary 3.1.

Proof. Considering separately the three cases $\zeta \in (1/4, 1/2)$, $\zeta = 1/2$, and $\zeta > 1/2$, the result follows from Propositions 3.1 and 3.2 as well as Corollary A.2 in Appendix A.3. $\square$

As claimed previously, the Hausman test is consistent against MA misspecification of order $T^{-\zeta}$ with $\zeta \in (1/4, 1/2)$, except in the knife-edge case where $\operatorname{aBias}(\hat{\delta}_h) = 0$. When $\zeta = 1/2$ and $\operatorname{aBias}(\hat{\delta}_h) \neq 0$, the asymptotic rejection probability is strictly between the significance level a and 1. In Section 4 we will use the Hausman test to quantify the difficulty of detecting especially pernicious types of model misspecification.

3.3 Lag length selection

We now elaborate further on the role of the estimation lag length. We first discuss the properties of LP when the lag length is selected using standard information criteria. We then show that increasing the lag length robustifies VAR inference by rendering it equivalent with LP inference.

Lag length selection for LP. Supplemental Appendix C.3 shows that the LP confidence interval maintains correct asymptotic coverage when the lag length p is selected via the Bayesian Information Criterion (BIC), provided that $\zeta \geq 1/2$. Specifically, the BIC should be applied to an auxiliary VAR in the observed data series $\check{y}_t$; the selected lag length then determines the number p of lags to control for in subsequent LP inference (which otherwise discards the auxiliary VAR). The resulting LP inference is robust to model selection errors that are known to cause difficulties for VAR inference (Leeb and Pötscher, 2005; Kilian and Lütkepohl, 2017, chapter 2.6.5). At a high level, the greater reliability of LP inference with data-dependent lag length is a consequence of the double robustness discussed earlier, see Belloni and Chernozhukov (2013) and Chernozhukov et al. (2018).

While the BIC suffices in theory for valid local projection inference, we follow Kilian and Lütkepohl (2017) and recommend that researchers employ the more conservative Akaike Information Criterion (AIC) in practice. The reason is that, in finite samples, the downsides of under-specifying the lag length outweigh the slight inefficiency associated with over-selecting the lag length. Section 5 demonstrates that this lag length selection procedure delivers LP confidence intervals with accurate coverage in realistic DGPs.

LONG-LAG VARS. One simple way to remove the asymptotic bias of the VAR estimator is to control for sufficiently many lags—typically many more lags than indicated by conventional information criteria. This is because in this case the estimator is asymptotically equivalent with the LP estimator. See Plagborg-Møller and Wolf (2021) and Xu (2023) for related results in models without explicit MA misspecification.

Corollary 3.2. *Suppose the model (2.2) written in companion form (2.1) satisfies Assumption 2.1 and $\zeta > 1/4$. Let $\tilde{\tilde{y}}_t$ denote the stationary solution to equation (2.2) when $\alpha(L) = 0$. If $\varepsilon_{j^*,t-\ell} \in \text{span}(\tilde{\tilde{y}}_{t-1}, \dots, \tilde{\tilde{y}}_{t-p})$ for all $\ell = 1, \dots, h$, then $\text{aBias}(\hat{\delta}_h) = 0$ and $\text{aVar}(\hat{\delta}_h) = \text{aVar}(\hat{\beta}_h)$. In particular, these results obtain if either of the following two sufficient conditions hold:*

i) The model is a local-to-SVAR(p_0) model (i.e., $\check{A}_\ell = 0$ for $p_0 < \ell \leq p$) and $h \leq p - p_0$, where p is the estimation lag length.

ii) The shock of interest is directly observed and ordered first (i.e., $j^* = 1$ and $\check{A}_{1,j,\ell} = 0$ for all j, ℓ), and $h \leq p$.

Proof. See Appendix B.3. □

We see that, the larger the impulse horizon h of interest, the larger is the estimation lag length p required for bias reduction. In fact, Section 4 shows that the *only* way to guarantee that the asymptotic bias of the VAR estimator is zero is to control for so many lags that LP and VAR are asymptotically equivalent.

4 VAR inference under bounded misspecification

To show that the fragility of VARs is likely to matter in practice, we now investigate the worst-case properties of VAR procedures under a tight constraint on the amount of misspecification. We prove that the conventional VAR confidence interval is robust if, *and only if*, LP and VAR intervals coincide asymptotically. VARs with short-to-moderate lag lengths suffer from severe coverage distortions even for small amounts of misspecification that are hard to rule out either economically or statistically. Beyond increasing the lag length, an alternative strategy to fix VAR undercoverage is to use a larger bias-aware critical value; however, we show that the resulting confidence interval is usually wider than the LP interval. Finally, we show that all conclusions extend to the case of joint inference on multiple impulse responses.

Throughout this section we set $\zeta = 1/2$ so that the asymptotic bias-variance trade-off between LP and VAR is non-trivial.

4.1 Worst-case bias and mean-squared error

Building towards our main results on VAR coverage distortions, we begin by deriving the worst-case bias and mean-squared error of the VAR estimator.

Misspecification bound. To quantify the amount of misspecification in the local-to-SVAR model (2.1) with $\zeta = 1/2$, we define the *noise-to-signal ratio*

$$\operatorname{trace}\left\{\operatorname{Var}(T^{-1/2}\alpha(L)\varepsilon_t)\operatorname{Var}(\varepsilon_t)^{-1}\right\} = \operatorname{trace}\left\{\left(T^{-1}\sum_{\ell=1}^{\infty}\alpha_\ell D\alpha_\ell'\right)D^{-1}\right\} = T^{-1}\|\alpha(L)\|^2,$$

where we define the norm $\|\alpha(L)\| \equiv \sqrt{\sum_{\ell=1}^{\infty}\operatorname{trace}\{D\alpha_\ell' D^{-1}\alpha_\ell\}}$. Suppose we are willing to impose *a priori* that the noise-to-signal ratio is at most M^2/T for some constant $M \in (0,\infty)$. For small M^2/T, this roughly means that a fraction M^2/T of the variance of the model's error term is due to the misspecification. This corresponds to restricting the parameter space for $\alpha(L)$ to all absolutely summable lag polynomials that satisfy $\|\alpha(L)\| \leq M$. In the following we will consider the worst-case properties of the VAR estimator over this parameter space, treating the other (consistently estimable) parameters (A, H, D) as fixed.

Worst-case bias.

Proposition 4.1. *Impose Assumption 2.1, $\zeta = 1/2$, and $\operatorname{aVar}(\hat{\delta}_h) > 0$. Then*

$$\max_{\alpha(L)\colon \|\alpha(L)\|\leq M} |b_h| = M\sqrt{\frac{\operatorname{aVar}(\hat{\beta}_h)}{\operatorname{aVar}(\hat{\delta}_h)} - 1},$$

where we recall the definition $b_h = \operatorname{aBias}(\hat{\delta}_h)/\sqrt{\operatorname{aVar}(\hat{\delta}_h)}$. Recall also that $\operatorname{aVar}(\hat{\beta}_h)$ and $\operatorname{aVar}(\hat{\delta}_h)$ do not depend on $\alpha(L)$.

Proof. The claim is a special case of Proposition C.2 in Supplemental Appendix C.4. □

Under our bound M^2/T on the noise-to-signal ratio, the worst-case (scaled) VAR bias is a simple function of M and of the relative asymptotic precision $\text{aVar}(\hat{\beta}_h)/\,\text{aVar}(\hat{\delta}_h)$ of the VAR estimator vs. LP. These two quantities are "sufficient statistics" for the worst-case bias regardless of the number n of variables in the VAR, the lag length p, the specific VAR parameters (A, H, D), and the horizon h. Hence, our subsequent analysis of the worst-case properties of VAR procedures depends only on M and on the relative precision, allowing us to concisely present analytical results that cover a wide range of local-to-SVAR models without having to resort to simulations that inevitably only cover a finite number of DGPs.

Proposition 4.1 shows that VAR estimators must trade off efficiency and robustness: the worst-case VAR bias is small precisely when the VAR estimator has nearly the same variance as LP. While the worst-case bias can be reduced by increasing the VAR estimation lag length p, the proposition shows that this can *only* happen at the expense of increasing the variance. If we include so many lags that the worst-case bias is zero (cf. Corollary 3.2), then the VAR estimator must *necessarily* be asymptotically equivalent with LP.

Worst-case mean squared error. For future reference we briefly discuss how the worst-case mean squared error (MSE) of the VAR estimator depends on the imposed bound on misspecification. Based on Propositions 3.1 and 3.2 as well as Corollary A.2, we define the asymptotic MSE of the VAR and LP estimators as follows:

$$\text{aMSE}(\hat{\beta}_h) \equiv \text{aVar}(\hat{\beta}_h), \quad \text{aMSE}(\hat{\delta}_h) \equiv \text{aBias}(\hat{\delta}_h)^2 + \text{aVar}(\hat{\delta}_h).$$

Corollary 4.1. *Impose Assumption 2.1 and $\zeta = 1/2$. Then*

$$\sup_{\alpha(L)\colon \|\alpha(L)\| \le M} \{\text{aMSE}(\hat{\delta}_h) - \text{aMSE}(\hat{\beta}_h)\} = (M^2 - 1)\{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}.$$

Proof. See Appendix B.4. □

In words, the worst-case MSE regret of VAR relative to LP is proportional to the variance reduction of VAR relative to LP, with a proportionality constant of $M^2 - 1$. If $M > 1$ (corresponding to a noise-to-signal ratio greater than $1/T$), the worst-case MSE of VAR thus strictly exceeds the MSE of LP. From here it is also straightforward to recover the minimax optimal way to average LP and VAR estimates.

Corollary 4.2. *Impose Assumption 2.1, $\zeta = 1/2$, and $\text{aVar}(\hat{\beta}_h) > \text{aVar}(\hat{\delta}_h)$. Consider the model-averaging estimator $\hat{\theta}_h(\omega) \equiv \omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h$, and denote its asymptotic MSE by $\text{aMSE}(\hat{\theta}_h(\omega))$. Then*

$$\operatorname*{argmin}_{\omega\in\mathbb{R}} \sup_{\alpha(L)\colon \|\alpha(L)\|\le M} \text{aMSE}(\hat{\theta}_h(\omega)) = \frac{M^2}{1+M^2}.$$

Proof. See Appendix B.5. □

If $M = 1$, it is minimax optimal to weight the LP and VAR estimates equally. If $M = 2$ (corresponding to a noise-to-signal ratio of $4/T$), the LP estimator receives 80% weight.

4.2 Worst-case coverage

We now turn to our main area of interest: the worst-case asymptotic coverage of the conventional VAR confidence interval under our bound on the amount of misspecification. This turns out to take a very simple form.

Corollary 4.3. *Impose Assumption 2.1, $\zeta = 1/2$, and $\text{aVar}(\hat{\delta}_h) > 0$. Then*

$$\inf_{\alpha(L)\colon \|\alpha(L)\|\le M} \lim_{T\to\infty} P(\theta_{h,T} \in \text{CI}(\hat{\delta}_h)) = 1 - r\left(M\sqrt{\text{aVar}(\hat{\beta}_h)/\,\text{aVar}(\hat{\delta}_h) - 1};\, z_{1-a/2}\right).$$

Proof. This is an immediate consequence of Corollary 3.1 and Proposition 4.1. □

Based on this corollary, Figure 4.1 provides a complete characterization of the robustness-efficiency trade-off for VAR confidence intervals. It plots the worst-case coverage probability as a function of the ratio of standard errors for VAR and LP, given significance level $a = 10\%$ and different values of M. The shaded area depicts an empirically relevant range of standard error ratios obtained in four empirical applications from Ramey (2016).[7] We see that, even for $M = 1$ (corresponding to a noise-to-signal ratio of $1/T$), the worst-case coverage probability is below 48% whenever the asymptotic standard deviation of the VAR estimator is less than half that of LP—a value that is typical in applied work. Further, at the bottom end of the empirically relevant range, the worst-case coverage probability is essentially zero as soon as $M \ge 1$. It is only at the very right side of the figure—when the VAR includes enough lags

[7] We replicate Ramey's identification schemes for monetary, tax, government spending, and technology shocks. The shaded area shows the 10th to 90th percentiles of standard error ratios at horizons exceeding 1 year. See the online replication materials for details.

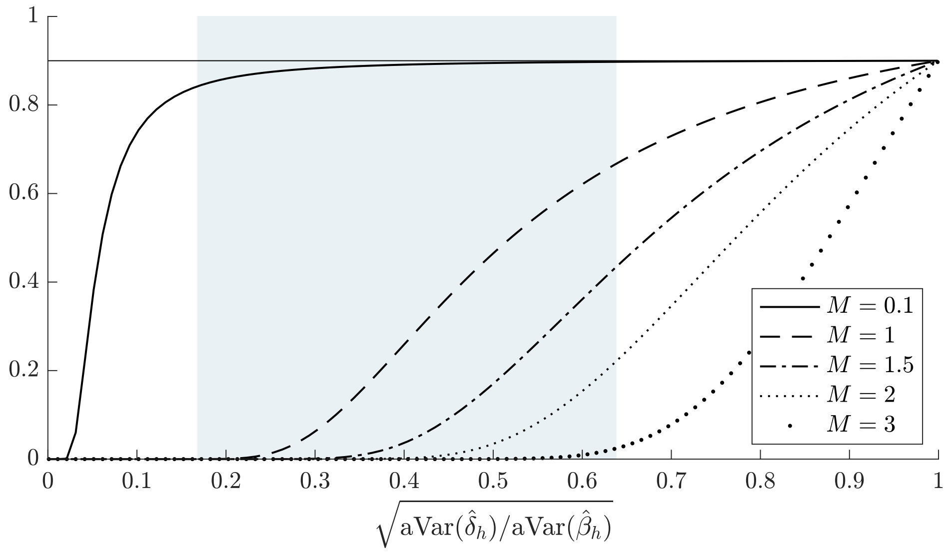

Figure 4.1: Worst-case asymptotic coverage probability of the conventional 90% VAR confidence interval. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. Different lines: different bounds M on $\|\alpha(L)\|$. Shaded area: empirical 10th–90th percentile range of relative standard errors based on Ramey (2016), see the online replication materials for details. The solid horizontal line marks the nominal coverage probability $1 - a = 90\%$.

to remove nearly all bias, thus increasing the standard error almost to that of LP—that the VAR confidence interval has coverage close to the nominal level.

The potential for VAR undercoverage documented here may not be so concerning if the worst-case misspecification can be ruled out on economic theory grounds, or if it is easily detectable statistically. We now argue that neither appears to be the case.

Economic theory. The shape and magnitude of the least favorable misspecification is difficult to rule out generally based on economic theory. The least favorable MA polynomial $\alpha^\dagger(L; h, M) = \sum_{\ell=1}^{\infty} \alpha^\dagger_{\ell,h,M} L^\ell$ for VAR coverage is the same as the least favorable one for bias (i.e., the $\alpha(L)$ that achieves the maximum in Proposition 4.1). Since $\text{aBias}(\hat{\delta}_h)$ is linear in $\alpha(L)$, the least favorable choice given the constraint $\|\alpha(L)\| \leq M$ follows from the Cauchy-Schwarz inequality (see the proof of Proposition C.2 in Supplemental Appendix C.4):

$$\alpha^\dagger_{\ell,h,M} \propto D^{1/2} H' \Psi_h' S^{-1} A^{\ell-1} H D^{1/2} - \mathbb{1}(\ell \leq h) \sigma_{j^*}^{-1} D^{1/2} H' (A')^{h-\ell} e_{i^*,n} e'_{j^*,m}, \quad \ell \geq 1, \tag{4.1}$$

where the constant of proportionality (which does not depend on the lag ℓ) is chosen so that $\|\alpha^\dagger(L; h, M)\| = M$. Note that the shape of the least favorable MA polynomial depends on the particular horizon h of interest but not on M; i.e., the bound M^2/T on the noise-to-signal ratio only scales the polynomial up or down.

We note two main properties of the least favorable misspecification. First, the magnitude of the MA coefficients $\alpha^{\dagger}_{\ell,h,M}$ decays exponentially as $\ell \to \infty$. In other words, not only is the overall magnitude of the least favorable model misspecification small (as imposed in the noise-to-signal bound), the MA coefficients at long lags are in fact particularly small. Second, numerical examples shown in Appendix A.1 suggest that the MA coefficients tend to be largest in magnitude at horizon h, displaying either a hump-shaped pattern as a function of ℓ—consistent with economic theories of adjustment costs or learning—or a single zig-zag pattern—consistent with theories of overshooting or lumpy adjustment. We thus view MA dynamics of the worst-case form as empirically and theoretically relevant.[8]

Statistical tests. The least favorable misspecification is also difficult to detect statistically. Propositions 3.3 and 4.1 imply that, for $\alpha(L) = \alpha^{\dagger}(L; h, M)$, the asymptotic rejection probability of the Hausman test of correct VAR specification equals $r(M; z_{1-a/2})$. When $M = 1$ (corresponding to a noise-to-signal ratio of $1/T$), the odds of the Hausman test *failing* to reject the misspecification are nearly 3-to-1 at significance level $a = 10\%$, since $r(1; z_{0.95}) = 26\%$. At significance level $a = 5\%$, the odds are nearly 5-to-1, since $r(1; z_{0.975}) = 17\%$. Standard *ex post* model misspecification tests are thus unlikely to indicate a problem even if the potential for undercoverage is severe.

Rather than committing *a priori* to a parameter space for $\alpha(L)$ through choice of M, we can also ask a different question: across all possible types and magnitudes of misspecification, what is the worst-case probability that the conventional VAR confidence interval fails to cover the true impulse response, yet we fail to reject correct specification of the VAR model?

Corollary 4.4. *Impose Assumption 2.1, $\zeta = 1/2$, and $\operatorname{aVar}(\hat{\beta}_h) > \operatorname{aVar}(\hat{\delta}_h) > 0$. Consider the joint event $\mathcal{A}_T$ that $\theta_{h,T} \notin \operatorname{CI}(\hat{\delta}_h)$ and the Hausman test in Proposition 3.3 fails to reject misspecification. Then*

$$\sup_{\alpha(L)} \lim_{T\to\infty} P(\mathcal{A}_T) = \sup_{b\geq 0}\, r(b; z_{1-a/2}) \left\{ 1 - r\left(\frac{b}{\sqrt{\operatorname{aVar}(\hat{\beta}_h)/\operatorname{aVar}(\hat{\delta}_h) - 1}}; z_{1-a/2} \right) \right\},$$

where the supremum on the left-hand side is taken over all absolutely summable lag polynomials $\alpha(L)$.

[8]However, the least favorable MA polynomial derived above need not be of interest to researchers who trust that some equations in their SVAR specification are exactly correctly specified, as this imposes the additional restrictions that some linear combinations of the rows of the MA polynomial $\alpha(L)$ equal zero.

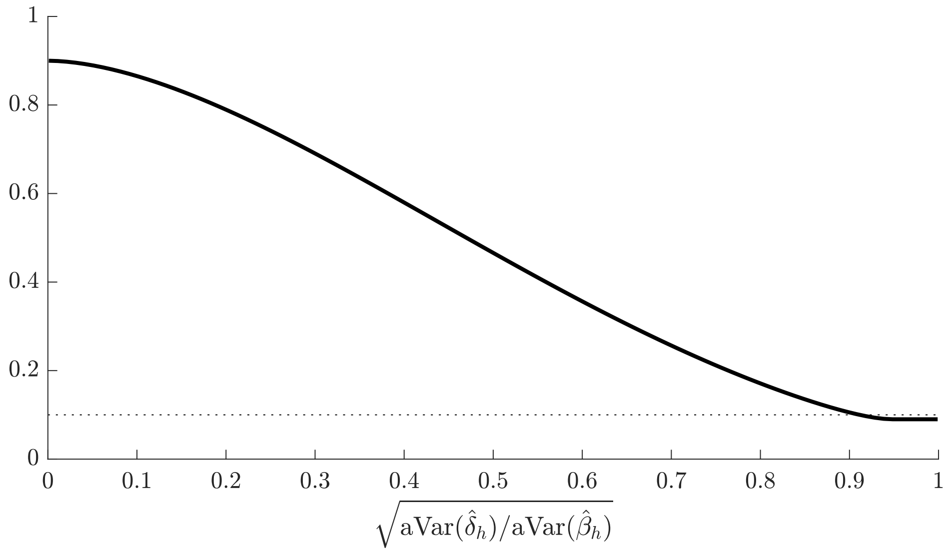

Figure 4.2: Worst-case asymptotic probability of the joint event that the conventional VAR confidence interval fails to cover the true impulse response and yet the Hausman test fails to reject misspecification. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. The dotted horizontal line marks the nominal significance level $a = 10\%$.

Proof. See Appendix B.6. $\square$

Figure 4.2 plots this worst-case probability for a significance level of $a = 10\%$, which by Corollary 4.4 depends only on the ratio $\mathrm{aVar}(\hat{\delta}_h)/\,\mathrm{aVar}(\hat{\beta}_h)$. Under correct specification, the probability of the joint event is equal to $a(1-a)$ ($= 9\%$ when $a = 10\%$). With misspecification, the joint probability instead exceeds 46% when the asymptotic standard deviation of the VAR estimator is less than half that of the LP estimator. As $\mathrm{aVar}(\hat{\delta}_h)/\,\mathrm{aVar}(\hat{\beta}_h) \to 0$, the worst-case joint probability approaches $1 - a$. We thus again see that statistical tests may fail to warn against the potential for severe VAR coverage distortions.

4.3 Bias-aware inference

Rather than removing bias by increasing the lag length (thus ensuring equivalence with LP), an alternative way to fix the undercoverage of the conventional VAR confidence interval is to adjust the critical value upward to compensate for the bias, as suggested in a general setting by Armstrong and Kolesár (2021). Suppose again that we restrict the misspecification $\alpha(L)$ to satisfy $\|\alpha(L)\| \leq M$. Then we define the *bias-aware* VAR confidence interval

$$\mathrm{CI}_B(\hat{\delta}_h; M) \equiv \left[\hat{\delta}_h \pm \mathrm{cv}_{1-a}\left(M\sqrt{\frac{\mathrm{aVar}(\hat{\beta}_h)}{\mathrm{aVar}(\hat{\delta}_h)} - 1}\right)\sqrt{\mathrm{aVar}(\hat{\delta}_h)/T}\right],$$

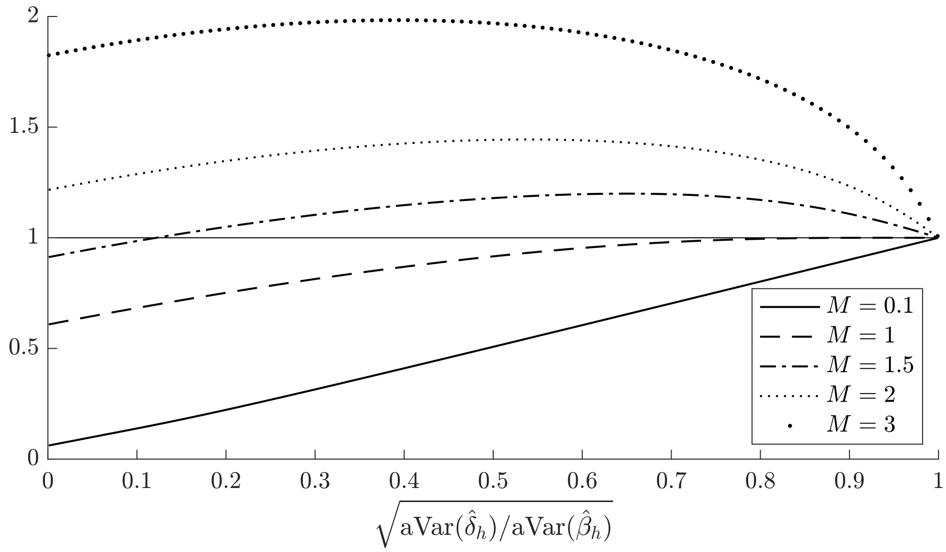

Figure 4.3: Relative length of bias-aware VAR confidence interval vs. conventional LP interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. Different lines: different bounds M on $\|\alpha(L)\|$. The solid horizontal line marks the value 1.

where the bias-aware critical value $\mathrm{cv}_{1-a}(b)$ is given by the number such that $r(b; \mathrm{cv}_{1-a}(b)) = a$, and $r(\cdot, \cdot)$ is defined in Corollary 3.1. By construction, this bias-aware confidence interval has correct (but potentially conservative) asymptotic coverage.

Corollary 4.5. *Impose Assumption 2.1, $\zeta = 1/2$, and $\mathrm{aVar}(\hat{\delta}_h) > 0$. Then*

$$\inf_{\alpha(L)\colon \|\alpha(L)\| \le M} \lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}_B(\hat{\delta}_h; M)) = 1 - a.$$

Proof. The result follows immediately from Propositions 3.2 and 4.1. □

It turns out, however, that a very tight bound M on the signal-to-noise ratio is required for the bias-aware VAR interval to be shorter than the LP interval. Figure 4.3 plots the relative interval length as a function of the relative asymptotic standard deviation of VAR and LP, for a significance level of $a = 10\%$ and for different misspecification bounds M. The figure shows that M has to be quite small—apparently below 1—for the bias-aware VAR length to dominate the LP length regardless of the DGP and horizon. Even for $M = 1.5$, bias-aware VAR is at best only moderately shorter than LP. Finally, for values of M above 2 (corresponding to a noise-to-signal ratio above $4/T$), bias-aware VAR is dominated by LP.

In Appendix A.2 we furthermore show that the conventional LP confidence interval is at worst slightly wider than a more efficient bias-aware confidence interval centered at the model averaging estimator $\hat{\theta}_h(\omega) = \omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h$, introduced in Corollary 4.2 above. Even if the weight ω is chosen to optimize confidence interval length, the gains relative to the LP interval are very small when $M \geq 2$ (corresponding to a noise-to-signal ratio above $4/T$).

We thus conclude that, while bias-aware VAR inference is possible in theory, in practice the gains relative to the simpler LP interval are small at best, unless we put an extremely tight bound on the noise-to-signal ratio.

4.4 Inference on multiple impulse responses

Since the least favorable MA polynomial derived in Section 4.2 depends on the horizon h of interest, one might hope that VARs would not be as prone to bias and thereby undercoverage if interest centers on *multiple* impulse responses. Unfortunately, Supplemental Appendix C.4 shows that this is not the case. There we consider inference on a vector of impulse responses for any combination of response variables i, shocks j, and horizons h. Generalizing Proposition 4.1, we show that the worst-case norm of the bias is non-negligible if the VAR offers efficiency gains for *any* linear combination of the parameters of interest. This implies that the conventional VAR confidence interval has fragile coverage even if the target parameter is a linear combination of impulse responses (such as the integral or sum across multiple horizons, as in the fiscal multiplier applications reviewed in Ramey, 2016). The conventional Wald confidence ellipsoid centered at the VAR estimator is similarly fragile.

5 Practical relevance

This section establishes the practical relevance of our theoretical conclusions by comprehensively reviewing current practice for lag length selection in the applied VAR literature, coupled with a simulation study calibrated to the application in Känzig (2021).

5.1 Review of current practice for VAR lag length selection

To evaluate lag length selection practice in the applied VAR literature, we created a comprehensive list of articles published between January 2015 and June 2025 in six top economics journals: *American Economic Review*, *American Economic Journal: Macroeconomics*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*. We restrict attention to papers that use VARs on time series data (not panel data) to estimate *structural* impulse response functions. This yields

81 papers total. For each paper, we picked a single specification as the "main" one.[9] To be conservative, if multiple specifications received equal attention in the main analysis, we picked the one with the largest lag length. The full list of papers, together with the recorded information about the VAR specifications, is provided in the online replication materials.

Our findings suggest that the theoretical results of the preceding sections are likely to have bite in practice, as typical VAR estimation lag lengths in applied papers are short or moderate. The modal lag length is 4 in quarterly data and 12 in monthly data, the mean is slightly below the mode, and less than 10% of papers employ lag lengths greater than or equal to twice the modal values.[10] 20% of papers select the lag length in a data-dependent way, often using information criteria. On average across papers, the estimation lag length is only 28% as large as the the longest reported impulse horizon. We conclude that few applied papers follow the recommendation of Kilian and Lütkepohl (2017, pp. 58–66) to use long lag lengths and avoid information criteria. This suggests that VAR inference results reported in much of the applied literature could be subject to the fragility we highlighted in the preceding sections. In fact, since around 40% of papers employ Bayesian shrinkage, the estimation bias could be even larger than indicated by the lag length alone.

5.2 Empirically calibrated simulation study

We now show through simulations that our asymptotic results are informative about the performance of LPs and VARs in an empirically relevant finite-sample setting when the lag length is selected as in current applied practice. Our DGP is calibrated to the oil news shock application in Känzig (2021).

Set-up. The DGP is a VAR estimated on the dataset of Känzig (2021), using a somewhat longer lag length than that used in the paper. The data series are that paper's oil shock proxy, the real price of oil, world oil production, world oil inventories, world industrial production, U.S. industrial production, and the U.S. consumer price index (CPI). Whereas Känzig employs 12 lags, we estimate a recursively identified VAR(18) by OLS on his data and use this as the simulation DGP, with i.i.d. Gaussian shocks. We do not claim that the VAR(18) DGP is more "realistic" than a VAR(12) estimated on the same data, but we contend that it is desirable that confidence intervals should have reliable coverage in both

[9] In a few cases, the main use of VARs in a paper was only reported in the appendix.

[10] Additionally, all the various empirical VAR specifications reported in the handbook chapters of Ramey (2016) and Stock and Watson (2016) are consistent with these patterns.

these DGPs.

The parameters of interest are the impulse responses of CPI to the observed oil shock. To be consistent with our theory, we use an "internal instruments" specification that orders the proxy first in the VAR; this differs from the "external instruments" specification used by Känzig. The sample size is $T = 720$ months. We will entertain different choices of the estimation lag length p for the LP and VAR estimators. We report results for both delta method and bootstrap confidence intervals. Results are based on 10,000 Monte Carlo simulations. See Supplemental Appendix D for implementation details.

Results. Figure 5.1 shows that our theoretical results on LP robustness and VAR fragility are practically relevant. The figure depicts the coverage probabilities (left panel) and median confidence interval length (right panel) for VARs (in red, solid and dashed) and LPs (in blue, solid and dashed). The top panel fixes the estimation lag length at $p = 12$, while the bottom panel selects the lag length by AIC. Both these choices are frequently encountered in the applied literature, as documented earlier. Given these conventional lag lengths, VAR confidence intervals tend to be shorter than LP intervals, but quite materially undercover, with coverage falling below 60% at medium and long horizons. LP instead attains close to the nominal coverage level of 90% throughout, as expected. The mean lag length selected by AIC is 9.7, evidently insufficient to guard against dynamic misspecification.

Supplemental Appendix D illustrates that the VAR under-coverage can be ameliorated by increasing the lag length beyond what is typically used in current applied practice, at the expense of higher variance, consistent with Section 3.3. The supplement also reports that the VAR estimator achieves lower MSE than LP, suggesting that—despite the poor performance of the VAR confidence interval—the larger bias of the VAR estimator may not compromise its usefulness as a *point estimator* (see also Li, Plagborg-Møller, and Wolf, 2024).

Montiel Olea, Plagborg-Møller, Qian, and Wolf (2025) find that the qualitative conclusions above extend to a wide range of empirically calibrated simulation DGPs based on richly specified dynamic factor models: VAR confidence intervals fail to adequately control coverage in a sizable fraction of DGPs, while LP confidence intervals robustly maintain accurate coverage. They document that LP is not only more robust to lag length selection, but also to the choice of control variables, consistent with the theory in this paper.

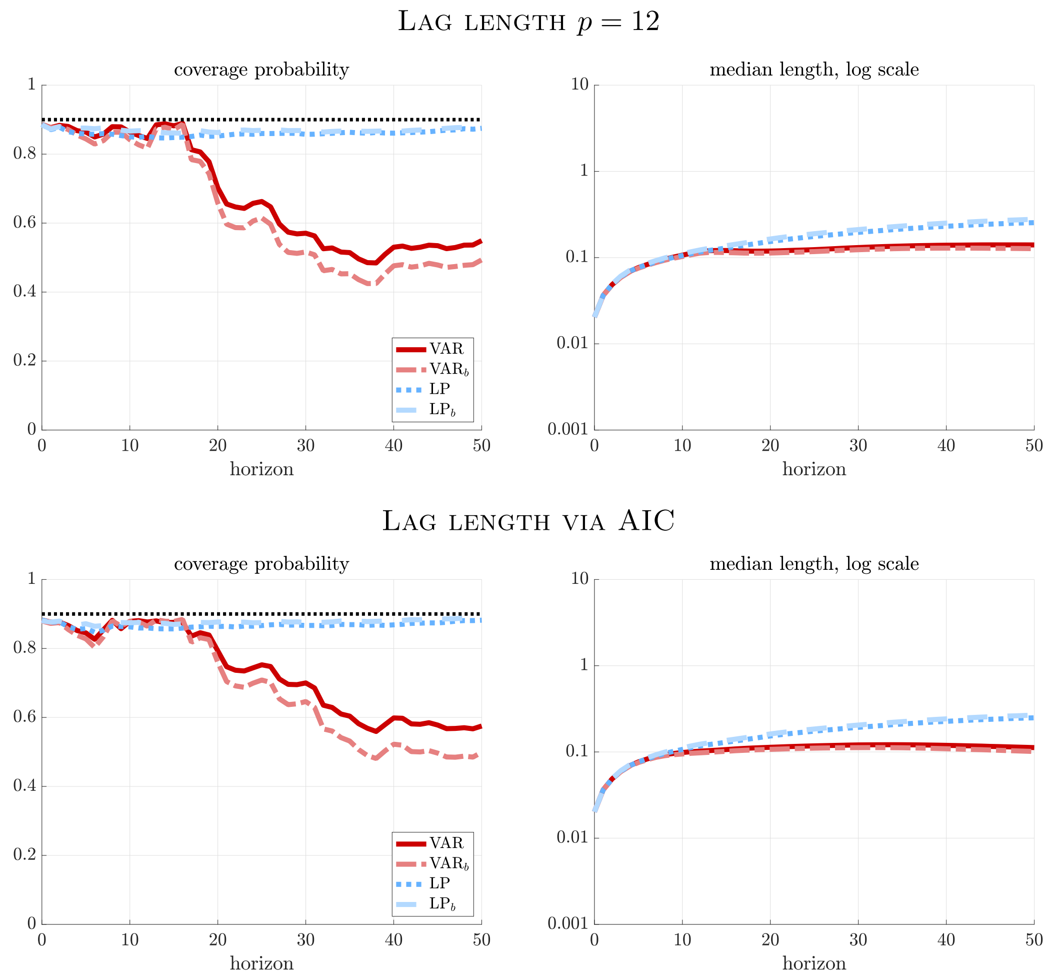

Figure 5.1: Coverage probability (left) and median length (right) for VAR (red) and LP (blue) nominal 90% confidence intervals computed via the delta method or bootstrap (the latter are indicated with subscript "b" in the legends). Lag length: fixed at $p = 12$ in the top panel, and selected using AIC in the bottom panel.

6 Conclusion

Our theoretical results suggest the following practical take-aways:

1. When the goal is to construct confidence intervals for impulse responses that have accurate coverage in a wide range of empirically relevant DGPs—as opposed to minimizing MSE—then the smaller bias of LPs documented in simulations by Li, Plagborg-Møller, and Wolf (2024) is more valuable than the smaller variance enjoyed by VAR estimators.

2. Researchers who use LP should control for those lags of the data that are strong predictors

of the outcome or impulse variables. This is important not only when the shock is recursively identified, but even if the researcher directly observes a near-perfect proxy for the shock of interest. However, unlike for VAR inference, it is *not* necessary to get the lag length or set of control variables exactly right to achieve correct coverage. To select the number of lags to control for in the LP, we recommend running an auxiliary VAR in all variables used in the analysis and selecting the lag length to minimize the AIC; the auxiliary VAR is only used as a device to select the lag length and is otherwise discarded.

3. With the moderate lag lengths typical in current applied practice, VAR confidence intervals will only have accurate coverage at short horizons, and only because they are approximately equivalent with LP intervals at these horizons. If, at some horizons of interest, an estimated VAR yields confidence intervals that are substantially narrower than the corresponding LP intervals, we recommend increasing the VAR lag length until that is no longer the case, to guarantee robust confidence interval coverage. Conventional tests of correct VAR specification do not suffice to guard against coverage distortions.

Is there a way forward for VAR inference, beyond just including a large number of lags? We showed how to construct a VAR confidence interval with a bias-aware critical value that robustly controls coverage, but found that it will typically lead to wider confidence intervals than LP. Another option would be to estimate VARMA models rather than pure VARs, though this would be computationally expensive, and the bias-variance trade-off relative to LPs is unclear. In principle, VAR procedures may work better under additional restrictions on the misspecification, such as shape restrictions on the impulse response functions.[11] However, it appears that detailed application-specific restrictions would be required to generate a negligible worst-case bias, since we have shown that the least favorable misspecification in our baseline analysis cannot generally be ruled out based on economic theory. Rather than restricting the parameter space, future research could instead investigate weakening the coverage requirement, e.g., only requiring a certain coverage probability *on average* over a set of horizons (Armstrong, Kolesár, and Plagborg-Møller, 2022), or by changing the target for inference from the true impulse response function to a smooth projection of this function (Genovese and Wasserman, 2008). Finally, a subjectivist Bayesian VAR modeler need only worry about our negative results if their prior on potential misspecification attaches significant weight to MA processes that imply large VAR biases.

[11] Given any convex parameter space for the misspecification MA polynomial $\alpha(L)$, the worst-case bias of the VAR estimator (see Proposition 3.2) can be computed using convex programming.

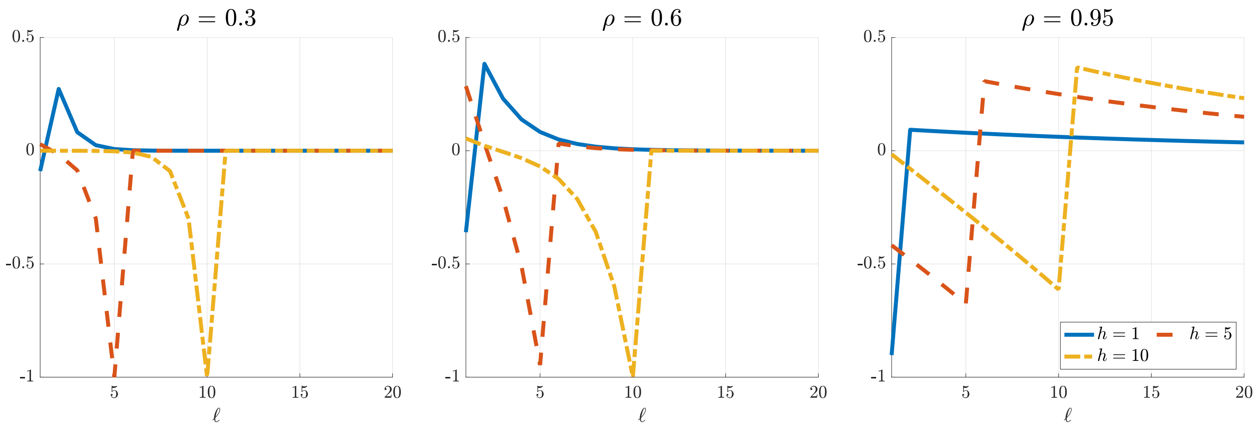

Figure A.1: Least favorable $\alpha^\dagger(L;h)$ for horizons $h \in \{1, 5, 10\}$ for local-to-AR(1) models with different persistence parameters ρ (left, middle, and right panel).

Appendix A Further theoretical results

A.1 Least favorable misspecification

Figure A.1 plots some numerical examples of the least favorable MA polynomial $\alpha^\dagger(L;h,M) = \sum_{\ell=1}^{\infty} \alpha^\dagger_{\ell,h,M} L^\ell$ discussed in Section 4.2. We focus here on a univariate local-to-AR(1) model $y_t = \rho y_{t-1} + [1 + T^{-1/2}\alpha(L)]\varepsilon_t$, though unreported numerical experiments suggest that the qualitative features mentioned below also apply to multivariate models. Recall that the least favorable MA coefficients depend on the horizon h of interest, while M only influences the overall scale of the coefficients, and not their shape as a function of ℓ. The figure shows that the shape of the coefficients either takes the form of a hump or of a single zig-zag pattern, with the largest absolute value of the coefficients generally occurring at $\ell = h$. Notice that we can flip the signs of all coefficients without changing the absolute value of the bias.

A.2 More efficient bias-aware confidence interval

Generalizing the bias-aware VAR confidence interval in Section 4.3, consider a bias-aware confidence interval that is centered at the model averaging estimator $\hat{\theta}_h(\omega) = \omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h$ from Corollary 4.2:

$$\mathrm{CI}_B(\hat{\theta}_h(\omega); M) \equiv \left[\hat{\theta}_h(\omega) \pm \mathrm{cv}_{1-a}\left(\frac{(1-\omega)M\tau}{\sqrt{1+\omega^2\tau^2}}\right)\sqrt{(1+\omega^2\tau^2)\,\mathrm{aVar}(\hat{\delta}_h)/T}\right],$$

where $\tau \equiv \sqrt{\mathrm{aVar}(\hat{\beta}_h)/\,\mathrm{aVar}(\hat{\delta}_h) - 1}$. This interval equals the conventional LP interval when $\omega = 1$ and the bias-aware VAR interval when $\omega = 0$.

Corollary A.1. *Impose Assumption 2.1, $\zeta = 1/2$, and $\mathrm{aVar}(\hat{\delta}_h) > 0$. Then, for any $\omega \in [0,1]$,*

$$\inf_{\alpha(L)\colon \|\alpha(L)\| \le M} \lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}_B(\hat{\theta}_h(\omega); M)) = 1 - a.$$

Proof. The result follows from Propositions 3.1, 3.2 and 4.1, Corollary A.2, and the same calculations as in the proof of Corollary 4.2. □

Even if we choose the weight ω to minimize confidence interval length, the resulting bias-aware interval tends to be nearly as long as the LP interval. The length-optimal weight $\omega = \omega^*$ is given by

$$\omega^* \equiv \operatorname*{argmin}_{\omega\in[0,1]} \; \mathrm{cv}_{1-a}\left(\frac{(1-\omega)M\tau}{\sqrt{1+\omega^2\tau^2}}\right)\sqrt{1+\omega^2\tau^2}.$$

Figure A.2 shows this optimal weight as a function of M and the relative asymptotic standard deviation of the VAR and LP estimators, while Figure A.3 shows the length of the resulting optimal bias-aware confidence interval relative to the length of the conventional LP interval. We see that, for $M \ge 2$, there is little gain from reporting the optimal bias-aware interval rather than the LP interval, regardless of the relative precision of VAR and LP. An additional observation is that, for $M \ge 1.5$, the length-optimal ω^* is numerically close to the MSE-optimal weight $M^2/(1+M^2)$ derived in Corollary 4.2.

A.3 Covariance structure of LP and VAR estimators

The following result provides the asymptotic variance-covariance matrix of the LP and VAR estimators in the general multi-dimensional set-up of Section 4.4. Define $\Psi_{i^*,j^*,h}$ as in Proposition 3.2, but making the dependence on (i^*, j^*) explicit in the notation.

Corollary A.2. *Impose Assumption 2.1, with part (iii) holding for all shock indices $j_1^*, \dots, j_k^*$. Then for any $a, b \in \{1, \dots, k\}$,*

$$\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_b^*,h_b}) = \mathbb{1}(j_a^* = j_b^*)\sigma_{j_a^*}^{-2}\left(\psi_{a,b} + \sum_{\ell=1}^{\min\{h_a,h_b\}} e'_{i_a^*,n} A^{h_a-\ell}\Sigma(A')^{h_b-\ell} e_{i_b^*,n}\right),$$
$$\mathrm{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) = \mathbb{1}(j_a^* = j_b^*)\sigma_{j_a^*}^{-2}\psi_{a,b} + \mathrm{trace}\left(\Psi_{i_a^*,j_a^*,h_a}\Sigma\Psi'_{i_b^*,j_b^*,h_b}S^{-1}\right),$$

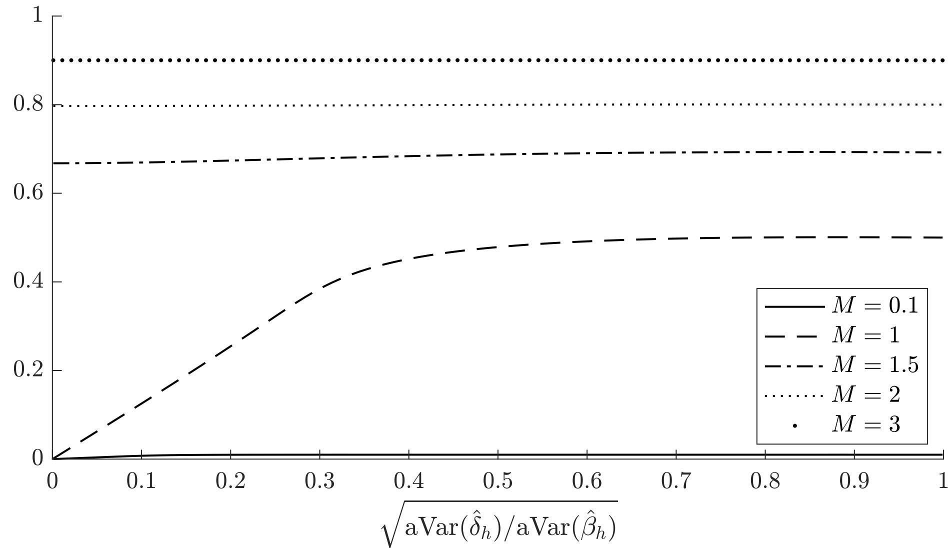

Figure A.2: Length-optimal weight on LP in bias-aware confidence interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. Different lines: different bounds M on $\|\alpha(L)\|$.

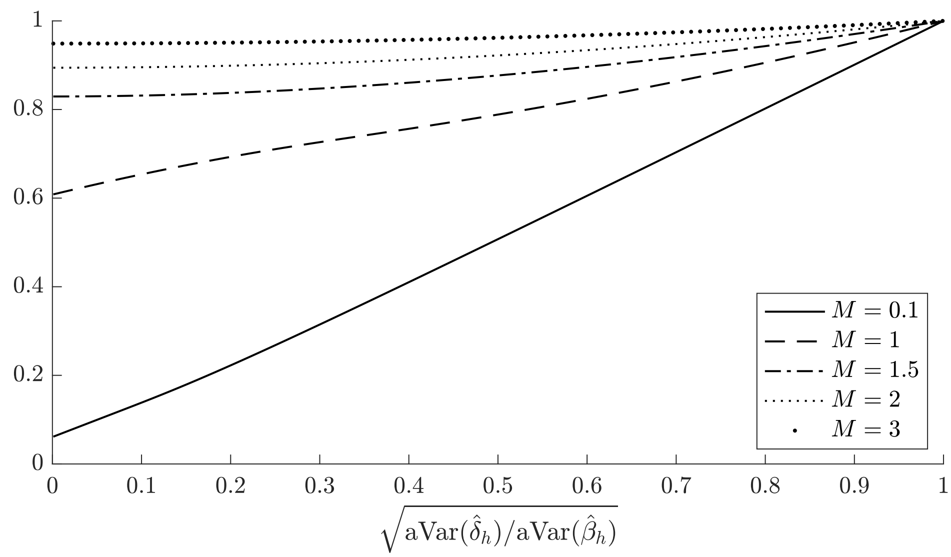

Figure A.3: Relative length of optimal bias-aware confidence interval vs. conventional LP interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. Different lines: different bounds M on $\|\alpha(L)\|$.

$$\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) = \mathrm{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}),$$

where

$$\psi_{a,b} \equiv e'_{i_a^*,n} A^{h_a} \overline{H}_{j_a^*} \overline{D}_{j_a^*} \overline{H}'_{j_a^*} (A')^{h_b} e_{i_b^*,n},$$

*and the "*aCov*" notation refers to elements of the asymptotic variance-covariance matrix in Equation* (C.2) *in Supplemental Appendix C.4. In particular,* $\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a} - \hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) = 0$.

Proof. See Appendix B.7. □

Appendix B Proofs

Lemmas whose name begins with "E" can be found in Supplemental Appendix E.

B.1 Proof of Proposition 3.1

Lemma E.1 shows that we can represent

$$y_{i^*,t+h} = \theta_{h,T}\varepsilon_{j^*,t} + \underline{B}'_{h,y}\underline{y}_{j^*,t} + B'_{h,y}y_{t-1} + \xi_{i^*,h,t} + T^{-\zeta}\Theta_h(L)\varepsilon_t, \tag{B.1}$$

where the expressions for the coefficient matrices and the $1 \times n$ two-sided lag polynomial $\Theta_h(L) = \sum_{\ell=-\infty}^{\infty} \Theta_{h,\ell} L^\ell$ are given in Lemma E.1.

Let $\hat{x}_{h,t}$ be the residual in a regression of $y_{j^*,t}$ on $\underline{y}_{j^*,t}$ and y_{t-1}, using data points $1, 2, \ldots, T-h$. By definition, $\hat{x}_{h,t}$ is in-sample orthogonal to $\underline{y}_{j^*,t}$ and y_{t-1}. Hence,

$$\begin{aligned}
\hat{\beta}_h &= \frac{\sum_{t=1}^{T-h} y_{i^*,t+h}\hat{x}_{h,t}}{\sum_{t=1}^{T-h} \hat{x}_{h,t}^2} \\
&= \theta_{h,T} + \frac{\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\hat{x}_{h,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t}}{\sum_{t=1}^{T-h} \hat{x}_{h,t}^2} \quad \text{by orthogonality} \\
&= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\hat{x}_{h,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t}}{\sigma_{j^*}^2 + o_p(1)} \quad \text{by Lemma E.4(v)} \\
&= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\varepsilon_{j^*,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2})}{\sigma_{j^*}^2 + o_p(1)} \\
&\qquad \text{by Lemma E.4(iv)}
\end{aligned}$$

$$= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(\xi_{i^*,h,t} + T^{-\zeta}\Theta_h(L)\varepsilon_t)\hat{x}_{h,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2})}{\sigma_{j^*}^2 + o_p(1)} \quad \text{by (B.1)}$$

$$= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(\xi_{i^*,h,t} + T^{-\zeta}\Theta_h(L)\varepsilon_t)\varepsilon_{j^*,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2})}{\sigma_{j^*}^2 + o_p(1)}$$

by Lemma E.4(iii) and (vi)

$$= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}\xi_{i^*,h,t}\varepsilon_{j^*,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2})}{\sigma_{j^*}^2 + o_p(1)} \quad \text{by Lemma E.1},$$

and the result follows. □

B.2 Proof of Proposition 3.2

Note first that

$$\begin{aligned}\hat{\delta}_h - e'_{i^*,n}A^h H_{\bullet,j^*} &= e'_{i^*,n}\hat{A}^h\hat{\nu} - e'_{i^*,n}A^h H_{\bullet,j^*} \\ &= e'_{i^*,n}\hat{A}^h H_{\bullet,j^*} - e'_{i^*,n}A^h H_{\bullet,j^*} + e'_{i^*,n}\hat{A}^h(\hat{\nu} - H_{\bullet,j^*}).\end{aligned}$$

Lemma E.2 shows that $\hat{A} - A = O_p(T^{-\zeta} + T^{-1/2})$. By Magnus and Neudecker (2007, Table 7, p. 208),

$$\left(\frac{\partial(e'_{i^*,n}A^h H_{\bullet,j^*})}{\partial\,\mathrm{vec}(A)}\right)' = (H'_{\bullet,j^*}\otimes e'_{i^*,n})\left(\sum_{\ell=1}^{h}(A')^{h-\ell}\otimes A^{\ell-1}\right) = \sum_{\ell=1}^{h} H'_{\bullet,j^*}(A')^{h-\ell}\otimes e'_{i^*,n}A^{\ell-1},$$

so

$$\begin{aligned}\hat{\delta}_h - e'_{i^*,n}A^h H_{\bullet,j^*} &= \left(\sum_{\ell=1}^{h} H'_{\bullet,j^*}(A')^{h-\ell}\otimes e'_{i^*,n}A^{\ell-1}\right)\mathrm{vec}(\hat{A}-A) + e'_{i^*,n}A^h(\hat{\nu} - H_{\bullet,j^*}) + o_p(T^{-\zeta} + T^{-1/2}) \\ &= \sum_{\ell=1}^{h} e'_{i^*,n}A^{\ell-1}(\hat{A}-A)A^{h-\ell}H_{\bullet,j^*} + e'_{i^*,n}A^h(\hat{\nu} - H_{\bullet,j^*}) + o_p(T^{-\zeta} + T^{-1/2}) \\ &= \mathrm{trace}\left\{\Psi_h(\hat{A}-A)\right\} + e'_{i^*,n}A^h(\hat{\nu} - H_{\bullet,j^*}) + o_p(T^{-\zeta} + T^{-1/2}),\end{aligned}$$

where $\Psi_h \equiv \sum_{\ell=1}^{h} A^{h-\ell}H_{\bullet,j^*}e'_{i^*,n}A^{\ell-1}$. Lemma E.2 further implies that

$$\begin{aligned}\mathrm{trace}\left\{\Psi_h(\hat{A}-A)\right\} = T^{-\zeta}\,&\mathrm{trace}\left\{S^{-1}\Psi_h H\sum_{\ell=1}^{\infty}\alpha_\ell D H'(A')^{\ell-1}\right\} \\ &+ \mathrm{trace}\left\{S^{-1}\Psi_h H T^{-1}\sum_{t=1}^{T}\varepsilon_t\tilde{y}'_{t-1}\right\} + o_p(T^{-\zeta}),\end{aligned}$$

where S was defined in Assumption 2.1. Lemma E.3 shows that

$$\hat{\nu} - H_{\bullet,j^*} = \frac{1}{\sigma_{j^*}^2} T^{-1} \sum_{t=1}^{T} \xi_{0,t} \varepsilon_{j^*,t} + o_p(T^{-\zeta} + T^{-1/2}).$$

Using the definition of $\theta_{h,T}$ and re-arranging terms gives the desired result. □

B.3 Proof of Corollary 3.2

Use the notation $E^*(z \mid w) = \mathrm{Cov}(z, w) \,\mathrm{Var}(w)^{-1} w$ for mean-square projection. Then

$$\begin{aligned}
\sigma_{j^*}^2 \Psi_h' S^{-1} \tilde{y}_{t-1} &= \left(\sum_{\ell=1}^{h} (A')^{h-\ell} e_{i^*,n} \underbrace{\sigma_{j^*}^2 H_{\bullet,j^*}' (A')^{\ell-1}}_{=\mathrm{Cov}(\varepsilon_{j^*,t-\ell}, \tilde{y}_{t-1})} \right) S^{-1} \tilde{y}_{t-1} \\
&= \sum_{\ell=1}^{h} (A')^{h-\ell} e_{i^*,n} E^*(\varepsilon_{j^*,t-\ell} \mid \tilde{y}_{t-1}) \\
&= \sum_{\ell=1}^{h} (A')^{h-\ell} e_{i^*,n} \varepsilon_{j^*,t-\ell},
\end{aligned}$$

where the last equality uses $\varepsilon_{j^*,t-\ell} \in \mathrm{span}(\tilde{\tilde{y}}_{t-1}, \dots, \tilde{\tilde{y}}_{t-p})$ for $\ell = 1, \dots, h$. Thus,

$$\begin{aligned}
\mathrm{Var}(\varepsilon_t' H' \Psi_h' S^{-1} \tilde{y}_{t-1}) &= \mathrm{Var}\left(\frac{1}{\sigma_{j^*}^2} \sum_{\ell=1}^{h} \varepsilon_{j^*,t-\ell} \varepsilon_t' H' (A')^{h-\ell} e_{i^*,n} \right) \\
&= \frac{1}{\sigma_{j^*}^4} \sum_{\ell=1}^{h} \mathrm{Var}\left(\varepsilon_{j^*,t-\ell} \varepsilon_t' H' (A')^{h-\ell} e_{i^*,n} \right) \\
&= \frac{1}{\sigma_{j^*}^4} \sum_{\ell=1}^{h} E(\varepsilon_{j^*,t-\ell}^2) \,\mathrm{Var}(\varepsilon_t' H' (A')^{h-\ell} e_{i^*,n}) \\
&= \frac{1}{\sigma_{j^*}^2} \mathrm{Var}\left(e_{i^*,n}' \sum_{\ell=1}^{h} A^{h-\ell} H \varepsilon_{t+\ell} \right).
\end{aligned}$$

It now follows as in the proof of Corollary A.2 that $\mathrm{aVar}(\hat{\beta}_h) = \mathrm{aVar}(\hat{\delta}_h)$. Then Proposition 4.1 implies that $\mathrm{aBias}(\hat{\delta}_h) = 0$. □

B.4 Proof of Corollary 4.1

By Proposition 4.1, $\sup_{\alpha(L) \colon \|\alpha(L)\| \le M} \mathrm{aBias}(\hat{\delta}_h; \alpha(L))^2 = M^2 \{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)\}$. The result follows. □

B.5 Proof of Corollary 4.2

Write $\hat{\theta}_h(\omega) = \hat{\delta}_h + \omega(\hat{\beta}_h - \hat{\delta}_h)$. By Corollary A.2, the two terms are asymptotically independent of each other, and the second term has asymptotic variance $\omega^2\{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}$. Hence,

$$\text{aMSE}(\hat{\theta}_h(\omega)) = \{(1-\omega)\,\text{aBias}(\hat{\delta}_h)\}^2 + \text{aVar}(\hat{\delta}_h) + \omega^2\{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}.$$

By Proposition 4.1, the supremum of the above expression over $\alpha(L)$ satisfying $\|\alpha(L)\| \leq M$ equals

$$(1-\omega)^2 M^2\{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\} + \text{aVar}(\hat{\delta}_h) + \omega^2\{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}.$$

To find the ω that minimizes the above expression, we can equivalently minimize the function $(1-\omega)^2 M^2 + \omega^2$. The result follows. □

B.6 Proof of Corollary 4.4

Proposition 4.1 implies that the absolute relative VAR bias $|b_h|$ can be made to take any value in $[0,\infty)$ as $\alpha(L)$ varies over the set of all absolutely summable lag polynomials. The corollary then follows from Corollaries 3.1 and A.2 and Proposition 3.3. □

B.7 Proof of Corollary A.2

We first use Proposition 3.1 to compute $\text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_b^*,h_b})$. Define $\xi_{j^*,h,t} = (\xi_{1,j^*,h,t}, \dots, \xi_{n,j^*,h,t})'$ as in Proposition 3.1, but making the dependence on both i^* and j^* explicit in the notation. Observe that

$$E[\xi_{i_a^*,j_a^*,h_a,t}\varepsilon_{j_a^*,t}\xi_{i_b^*,j_b^*,h_b,s}\varepsilon_{j_b^*,s}] = 0 \quad \text{for all } s \neq t.$$

Hence,

$$\text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_b^*,h_b}) = \frac{1}{\sigma_{j_a^*}^2\sigma_{j_b^*}^2} E[\xi_{i_a^*,j_a^*,h_a,t}\varepsilon_{j_a^*,t}\xi_{i_b^*,j_b^*,h_b,t}\varepsilon_{j_b^*,t}].$$

If $j_a^* < j_b^*$, then $\varepsilon_{j_a^*,t}$ is independent of all the other terms in the above expectation, so the expectation equals zero; similarly if $j_a^* > j_b^*$. Now consider the case $j_a^* = j_b^*$:

$$\text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_a^*,h_b}) = \frac{1}{\sigma_{j_a^*}^4} E[\xi_{i_a^*,j_a^*,h_a,t}\xi_{i_b^*,j_a^*,h_b,t}\varepsilon_{j_a^*,t}^2]$$

$$
\begin{aligned}
&= \frac{1}{\sigma_{j_a^*}^4} E[\xi_{i_a^*,j_a^*,h_a,t}\xi_{i_b^*,j_a^*,h_b,t}]E[\varepsilon_{j_a^*,t}^2] \\
&= \frac{1}{\sigma_{j_a^*}^2} E[\xi_{i_a^*,j_a^*,h_a,t}\xi_{i_b^*,j_a^*,h_b,t}] \\
&= \frac{1}{\sigma_{j_a^*}^2} \Bigg(E[e'_{i_a^*,n} A^{h_a} \overline{H}_{j_a^*} \bar{\varepsilon}_{j_a^*,t} \bar{\varepsilon}'_{j_a^*,t} \overline{H}'_{j_a^*} (A')^{h_b} e_{i_b^*,n}] \\
&\qquad\qquad + E\left[e'_{i_a^*,n} \sum_{\ell_1=1}^{h_a} \sum_{\ell_2=1}^{h_b} A^{h_a-\ell_1} H \varepsilon_{t+\ell_1} \varepsilon'_{t+\ell_2} H' (A')^{h_b-\ell_2} e_{i_b^*,n} \right] \Bigg) \\
&= \frac{1}{\sigma_{j_a^*}^2} \left(\psi_{a,b} + \sum_{\ell=1}^{\min\{h_a,h_b\}} e'_{i_a^*,n} A^{h_a-\ell} \Sigma (A')^{h_b-\ell} e_{i_b^*,n} \right),
\end{aligned}
$$

as claimed.

We now derive $\text{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b})$ using Proposition 3.2. Observe that the vector process $(\varepsilon'_t \otimes \tilde{y}'_{t-1}, \xi'_{j_a^*,0,t}\varepsilon_{j_a^*,t}, \xi'_{j_b^*,0,t}\varepsilon_{j_b^*,t})'$ is a martingale difference sequence with respect to the filtration generated by $\{\varepsilon_t\}$. Moreover, $E[(\varepsilon_t \otimes \tilde{y}_{t-1})\xi'_{j^*,0,t}\varepsilon_{j^*,t}] = 0$ for any j^*. Hence,

$$
\begin{aligned}
\text{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) &= E\left[\text{trace}\left(S^{-1}\Psi_{i_a^*,j_a^*,h_a} H \varepsilon_t \tilde{y}'_{t-1} \right) \text{trace}\left(S^{-1}\Psi_{i_b^*,j_b^*,h_b} H \varepsilon_t \tilde{y}'_{t-1} \right) \right] \\
&\quad + \frac{1}{\sigma_{j_a^*}^2 \sigma_{j_b^*}^2} E\left[e'_{i_a^*,n} A^{h_a} \xi_{j_a^*,0,t} \xi'_{j_b^*,0,t} (A')^{h_b} e_{i_b^*,n} \varepsilon_{j_a^*,t} \varepsilon_{j_b^*,t} \right].
\end{aligned}
$$

The second term on the right-hand side above equals $\mathbb{1}(j_a^* = j_b^*)\sigma_{j_a^*}^{-2}\psi_{a,b}$, by similar arguments as in the earlier LP calculation. The first term on the right-hand side above equals

$$
\begin{aligned}
&E\left[\tilde{y}'_{t-1} S^{-1} \Psi_{i_a^*,j_a^*,h_a} H \varepsilon_t \varepsilon'_t H' \Psi'_{i_b^*,j_b^*,h_b} S^{-1} \tilde{y}_{t-1} \right] \\
&= \text{trace}\left(E\left[\tilde{y}_{t-1}\tilde{y}'_{t-1} S^{-1} \Psi_{i_a^*,j_a^*,h_a} H \varepsilon_t \varepsilon'_t H' \Psi'_{i_b^*,j_b^*,h_b} S^{-1} \right] \right) \\
&= \text{trace}\left(E\left[\tilde{y}_{t-1}\tilde{y}'_{t-1} \right] S^{-1} \Psi_{i_a^*,j_a^*,h_a} H E\left[\varepsilon_t \varepsilon'_t \right] H' \Psi'_{i_b^*,j_b^*,h_b} S^{-1} \right) \\
&= \text{trace}\left(S S^{-1} \Psi_{i_a^*,j_a^*,h_a} H D H' \Psi'_{i_b^*,j_b^*,h_b} S^{-1} \right) \\
&= \text{trace}\left(\Psi_{i_a^*,j_a^*,h_a} \Sigma \Psi'_{i_b^*,j_b^*,h_b} S^{-1} \right),
\end{aligned}
$$

as claimed.

Finally, we compute $\text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b})$ using Propositions 3.1 and 3.2. Using arguments similar to above, we obtain

$$
\text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b})
$$

$$= \frac{1}{\sigma^2_{j_a^*}} \sum_{s=-\infty}^{\infty} E\left[e'_{i_a^*,n} \sum_{\ell=1}^{h_a} A^{h_a-\ell} H \varepsilon_{t+\ell} \varepsilon_{j_a^*,t} \operatorname{trace}\left(S^{-1} \Psi_{i_b^*,j_b^*,h_b} H \varepsilon_{t+s} \tilde{y}'_{t+s-1} \right) \right] + \mathbb{1}(j_a^* = j_b^*) \sigma^{-2}_{j_a^*} \psi_{a,b}.$$

The first term on the left-hand side above equals

$$\begin{aligned}
&\frac{1}{\sigma^2_{j_a^*}} \sum_{\ell=1}^{h_a} \sum_{s=-\infty}^{\infty} E\left[e'_{i_a^*,n} A^{h_a-\ell} H \varepsilon_{t+\ell} \varepsilon'_{t+s} H' \Psi'_{i_b^*,j_b^*,h_b} S^{-1} \tilde{y}_{t+s-1} \varepsilon_{j_a^*,t} \right] \\
&= \frac{1}{\sigma^2_{j_a^*}} \sum_{\ell=1}^{h_a} E\left[e'_{i_a^*,n} A^{h_a-\ell} H \varepsilon_{t+\ell} \varepsilon'_{t+\ell} H' \Psi'_{i_b^*,j_b^*,h_b} S^{-1} \tilde{y}_{t+\ell-1} \varepsilon_{j_a^*,t} \right] \\
&= \frac{1}{\sigma^2_{j_a^*}} \sum_{\ell=1}^{h_a} e'_{i_a^*,n} A^{h_a-\ell} H E[\varepsilon_{t+\ell} \varepsilon'_{t+\ell}] H' \Psi'_{i_b^*,j_b^*,h_b} S^{-1} E[\tilde{y}_{t+\ell-1} \varepsilon_{j_a^*,t}] \\
&= \frac{1}{\sigma^2_{j_a^*}} \sum_{\ell=1}^{h_a} e'_{i_a^*,n} A^{h_a-\ell} H D H' \Psi'_{i_b^*,j_b^*,h_b} S^{-1} A^{\ell-1} H_{\bullet,j_a^*} \sigma^2_{j_a^*} \\
&= \operatorname{trace}\left(\underbrace{\sum_{\ell=1}^{h_a} A^{\ell-1} H_{\bullet,j_a^*} e'_{i_a^*,n} A^{h_a-\ell}}_{=\sum_{\ell=1}^{h_a} A^{h_a-\ell} H_{\bullet,j_a^*} e'_{i_a^*,n} A^{\ell-1} = \Psi_{i_a^*,j_a^*,h_a}} \Sigma \Psi'_{i_b^*,j_b^*,h_b} S^{-1} \right).
\end{aligned}$$

It follows that $\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) = \mathrm{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b})$, as claimed. □

References

Ai, C., and X. Chen (2007): "Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables," *Journal of Econometrics*, 141(1), 5–43, Special issue on "Semiparametric methods in econometrics".

Alessi, L., M. Barigozzi, and M. Capasso (2011): "Non-Fundamentalness in Structural Econometric Models: A Review," *International Statistical Review*, 79(1), 16–47.

Armstrong, T. B., and M. Kolesár (2021): "Sensitivity analysis using approximate moment condition models," *Quantitative Economics*, 12(1), 77–108.

Armstrong, T. B., M. Kolesár, and M. Plagborg-Møller (2022): "Robust Empirical Bayes Confidence Intervals," *Econometrica*, 90(6), 2567–2602.

Bang, H., and J. M. Robins (2005): "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61(4), 962–972.

Belloni, A., and V. Chernozhukov (2013): "Least squares after model selection in high-dimensional sparse models," *Bernoulli*, 19(2), 521–547.

Braun, P. A., and S. Mittnik (1993): "Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions," *Journal of Econometrics*, 59(3), 319–341.

Brockwell, P. J., and R. A. Davis (1991): *Time Series: Theory and Methods*, Springer Series in Statistics. Springer, 2nd edn.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2008): "Are structural VARs with long-run restrictions useful in developing business cycle theory?," *Journal of Monetary Economics*, 55(8), 1337–1352.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21(1), C1–C68.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2022): "Locally Robust Semiparametric Estimation," *Econometrica*, 90(4), 1501–1535.

Genovese, C., and L. Wasserman (2008): “Adaptive confidence bands,” *The Annals of Statistics*, 36(2), 875 – 905.

González-Casasús, O., and F. Schorfheide (2025): “Misspecification-Robust Shrinkage and Selection for VAR Forecasts and IRFs,” Discussion Paper 33474, National Bureau of Economic Research.

Granger, C. W. J., and M. J. Morris (1976): “Time Series Modelling and Interpretation,” *Journal of the Royal Statistical Society. Series A (General)*, 139(2), 246–257.

Hausman, J. (1978): “Specification Tests in Econometrics,” *Econometrica*, 46(6), 1251–1271.

Herbst, E. P., and B. K. Johannsen (2024): “Bias in local projections,” *Journal of Econometrics*, 240(1), 105655.

Inoue, A., and L. Kilian (2002): “Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR(∞) Models,” *International Economic Review*, 43(2), 309–331.

——— (2020): “The uniform validity of impulse response inference in autoregressions,” *Journal of Econometrics*, 215(2), 450–472.

Jordà, Ò. (2005): “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, 95(1), 161–182.

——— (2023): “Local Projections for Applied Economics,” *Annual Review of Economics*, 15(1), 607–631.

Känzig, D. R. (2021): “The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements,” *American Economic Review*, 111(4), 1092–1125.

Kilian, L. (1998): “Small-sample Confidence Intervals for Impulse Response Functions,” *Review of Economics and Statistics*, 80(2), 218–230.

Kilian, L., and Y. J. Kim (2011): “How Reliable Are Local Projection Estimators of Impulse Responses?,” *Review of Economics and Statistics*, 93(4), 1460–1466.

Kilian, L., and H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Leeb, H., and B. M. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21(01), 21–59.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2024): "Local projections vs. VARs: Lessons from thousands of DGPs," *Journal of Econometrics*, 244(2), 105722, Themed Issue: Macroeconometrics.

Lütkepohl, H. (1984): "Linear aggregation of vector autoregressive moving average processes," *Economics Letters*, 14(4), 345–350.

Magnus, J., and H. Neudecker (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley Series in Probability and Statistics. John Wiley & Sons, 3rd edn.

Montiel Olea, J. L., and M. Plagborg-Møller (2021): "Local Projection Inference Is Simpler and More Robust Than You Think," *Econometrica*, 89(4), 1789–1823.

Montiel Olea, J. L., M. Plagborg-Møller, E. Qian, and C. K. Wolf (2025): "Local Projections or VARs? A Primer for Macroeconomists," *NBER Macroeconomics Annual*, Forthcoming.

Müller, U. K., and J. H. Stock (2011): "Forecasts in a Slightly Misspecified Finite Order VAR," Manuscript, Princeton University.

Nakamura, E., and J. Steinsson (2018): "Identification in Macroeconomics," *Journal of Economic Perspectives*, 32(3), 59–86.

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

Pope, A. L. (1990): "Biases of Estimators in Multivariate Non-Gaussian Autoregressions," *Journal of Time Series Analysis*, 11(3), 249–258.

Ramey, V. A. (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2, chap. 2, pp. 71–162. Elsevier.

Robins, J., and A. Rotnitzky (2001): "Comments on "Inference for Semiparametric Models: Some Questions and an Answer", by P.J. Bickel and J. Kwon," *Statistica Sinica*, 11(4), 920–936.

Robins, J. M., S. D. Mark, and W. K. Newey (1992): "Estimating Exposure Effects by Modelling the Expectation of Exposure Conditional on Confounders," *Biometrics*, 48(2), 479–495.

Robins, J. M., and A. Rotnitzky (1995): "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90(429), 122–129.

Robins, J. M., A. Rotnitzky, and M. van der Laan (2000): "On Profile Likelihood: Comment," *Journal of the American Statistical Association*, 95(450), 477–482.

Rothenberg, T. J. (1984): "Approximating the distributions of econometric estimators and test statistics," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2, chap. 15, pp. 881–935. Elsevier.

Schorfheide, F. (2005): "VAR forecasting under misspecification," *Journal of Econometrics*, 128(1), 99–136.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48(1), 1–48.

Stock, J. H., and M. W. Watson (2016): "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2, chap. 8, pp. 415–525. Elsevier.

Stock, J. H., and M. W. Watson (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal*, 128(610), 917–948.

Xu, K.-L. (2023): "Local Projection Based Inference under General Conditions," Manuscript, Indiana University Bloomington.