NBER WORKING PAPER SERIES

DOUBLE ROBUSTNESS OF LOCAL PROJECTIONS AND SOME UNPLEASANT VARITHMETIC

José Luis Montiel Olea
Mikkel Plagborg-Møller
Eric Qian
Christian K. Wolf

Working Paper 32495
http://www.nber.org/papers/w32495

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2024, revised August 2024

We received helpful comments from 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 Columbia, Princeton, the Federal Reserve Banks of Cleveland and Philadelphia, the 2024 NBER Summer Institute Monetary Meeting, and the 2024 Conference in Honor of Christopher A. Sims. Plagborg-Møller acknowledges that this material is based upon work supported by the NSF under Grant #2238049, and Wolf does the same for Grant #2314736. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or the National Bureau of Economic Research.

Double Robustness of Local Projections and Some Unpleasant VARithmetic
José Luis Montiel Olea, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf
NBER Working Paper No. 32495
May 2024, revised August 2024
JEL No. C22,C32

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. In contrast, the conventional VAR confidence interval with short-to-moderate lag length can severely undercover, even for misspecification that is small, economically plausible, and difficult to detect statistically. There is no free lunch: the VAR confidence interval has robust coverage only if the lag length is so large that the interval is as wide as the LP interval.

José Luis Montiel Olea
Cornell University
Department of Economics
440 Uris Hall
Ithaca, NY 14853
montiel.olea@gmail.com

Mikkel Plagborg-Møller
Princeton University
Department of Economics
Julis Romo Rabinowitz Building
Princeton, NJ 08544
mikkelpm@princeton.edu

Eric Qian
Department of Economics
Julis Romo Rabinowitz Building
Princeton, NJ 08544
ericqian@princeton.edu

Christian K. Wolf
MIT Department of Economics
The Morris and Sophie Chang Building
50 Memorial Drive
Cambridge, MA 02139
and NBER
ckwolf@mit.edu

A data appendix is available at http://www.nber.org/data-appendix/w32495

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. For VAR confidence intervals, we on the other hand show that there is no free lunch: they are robust if, *and only if*, they are as wide as LP intervals asymptotically, as is the case when they feature 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 (e.g., see Newey, 1990; Ai and Chen, 2007; Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2018; Chernozhukov, Escanciano, Ichimura, Newey, and Robins, 2022).

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. From here a "no free lunch" result for VARs emerges: 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 VAR confidence intervals 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 through simulations that our asymptotic results are useful to understand the finite-sample trade-off between LP and VAR confidence intervals. We consider a researcher that observes data generated from the Smets and Wouters (2007) model, and uses LPs and VARs to estimate the dynamic causal effects of cost-push or monetary shocks. If the lag length is selected by the Akaike Information Criterion (AIC), then VAR confidence intervals materially undercover—particularly at medium and long horizons—while LP throughout attains close to nominal coverage. Consistent with our theoretical results, increasing the estimation lag length ameliorates the VAR coverage, but 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 (2024) 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 contribute by: (i) focusing on structural impulse responses rather than forecasting; (ii) allowing for more general rates of local misspecification, 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 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 arising 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 and the LP and VAR estimators. Section 3 proves the robustness of LP and fragility of VAR confidence intervals. Section 4 derives analytically the worst-case bias and coverage of VARs, presents our "no free lunch" result, and shows that bias-aware VAR confidence intervals tend to be wider than the LP interval. Section 5 demonstrates the practical relevance of our results through simulations. Section 6 concludes. Replication codes are available online.

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 $y_t = (y_{1,t}, \dots, y_{n,t})'$ is n-dimensional, the shock vector $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{m,t})'$ 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 $m \times m$ 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 $j^* \in \{1,\dots,m\}$. We define this parameter formally 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) $\zeta > 1/4$.

The assumption of shock homoskedasticity is made for analytical convenience, though we expect that our qualitative conclusions about the robustness of LP and the asymptotic bias of VAR will go through under various forms of conditional heteroskedasticity. 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., $j^* = 1$). It is a minor extension to allow for identification via external instruments, also known as 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). The significance of the assumption that misspecification vanishes faster than $T^{-1/4}$ will become clear below.

The impulse response of interest is defined as

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

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 DSGE 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 (but misspecification in other parts of the model) and misspecification in the 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 is not entirely accurate.

We conjecture, but do not prove, that our framework can also be extended to accommodate omitted nonlinearities and time-varying parameters, as long as those features are small compared to the linear, time-invariant model components.[1]

2.2 Estimators

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

[1]Specifically, we conjecture that our proofs can be extended to allow for arbitrary additive local misspecification of the form

$$y_t = Ay_{t-1} + H\varepsilon_t + T^{-\zeta}v_t,$$

where v_t is an unobserved, stationary, non-deterministic process that is independent of $\{\varepsilon_{t+\ell}\}_{\ell \geq 0}$. Importantly, the parameter of interest $\theta_{h,T}$ must be defined as the coefficient in a population projection of $y_{i^*,t+h}$ onto $\varepsilon_{j^*,t}$. The Wold decomposition theorem implies that v_t has an MA representation, ultimately yielding a model of the form (2.1). However, since the orthogonalized Wold innovations will not satisfy all the independence conditions in Assumption 2.1, our current proofs do not apply directly.

1. The *LP* estimator is the coefficient $\hat{\beta}_h$ in a regression of $y_{i^*,t+h}$ on $y_{j^*,t}$, controlling for $\underline{y}_{j^*,t} \equiv (y_{1,t}, \ldots, y_{j^*-1,t})'$ (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 $\check{y}_{t-1}, \ldots, \check{y}_{t-p}$.

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

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

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.6 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 invariant to misspecification.

Proposition 3.1. *Under Assumption 2.1,*

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

where

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

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

Proof. See Appendix B.1. □

In words, the asymptotic behavior of LP does not depend on the misspecification parameters $\alpha(L)$ and ζ, provided $\zeta > 1/4$ as imposed in Assumption 2.1. 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 $\zeta \in (1/4, 1/2)$ 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 $T^{-2\zeta} = o(T^{-1/2})$ when $\zeta > 1/4$, so 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, 2024).

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 local projection 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 $E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{\nu_0(y^{t-1}) - \nu(y^{t-1})\}]$.[2] 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 $T^{-2\zeta} = o(T^{-1/2})$ and

[2] 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 $\nu_0(\cdot)$.

thus asymptotically negligible, consistent with our earlier intuition.

The fragility of VARs. In contrast to LP, the VAR estimator is fragile, as it is subject to non-negligible asymptotic bias under misspecification.

Proposition 3.2. *Under Assumption 2.1,*

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

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. □

The convergence rate $T^{-\min\{1/2,\zeta\}}$ of the VAR estimator is potentially slower than the $T^{-1/2}$ 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.[3] 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

[3] 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 $\hat{\nu}$.

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}.[4]

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:

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

Here $z_{1-a/2}$ is the $1-a/2$ quantile of the standard normal distribution, and $\text{aVar}(\hat{\beta}_h)$ and $\text{aVar}(\hat{\delta}_h)$ 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 $\text{aVar}(\hat{\beta}_h) \geq \text{aVar}(\hat{\delta}_h)$. 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).[5]

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

[4]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.

[5]Under Assumption 2.1, homoskedastic OLS standard errors suffice. In practice we recommend heteroskedasticity-robust standard errors or the bootstrap, as in Montiel Olea and Plagborg-Møller (2021).

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, thus causing coverage distortions.

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 $\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|/\sqrt{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)}$. 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 and assume* $\text{aVar}(\hat{\beta}_h) > \text{aVar}(\hat{\delta}_h) > 0$. *Then the asymptotic rejection probability of the Hausman test equals*

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

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 is an immediate consequence of Propositions 3.1 and 3.2 as well as Corollary A.2 in Appendix A.3. □

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 $\text{aBias}(\hat{\delta}_h) = 0$. When $\zeta = 1/2$ and $\text{aBias}(\hat{\delta}_h) \neq 0$, 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 The role of lag length

One simple way to remove the asymptotic bias of the VAR estimator is to control for sufficiently many lags, since in this case the estimator is asymptotically equivalent with the LP estimator. The larger the impulse horizon of interest, the more lags are required for bias reduction. 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. Let* $\tilde{\tilde{y}}_t$ *denote the stationary solution to equation* (2.2) *when* $\alpha(L) = 0$*. If* $\varepsilon_{j^*,t-\ell} \in \text{span}(\tilde{\tilde{y}}_{t-1}, \ldots, \tilde{\tilde{y}}_{t-p})$ *for all* $\ell = 1, \ldots, h$*, 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.4. □

We show in Section 4 that controlling for so many lags that LP and VAR are asymptotically equivalent is in fact the *only* way to guarantee that the asymptotic bias of the VAR estimator is zero.

4 Some unpleasant VARithmetic

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 there is no free lunch: the conventional VAR confidence interval is robust to misspecification if, and only if, the LP and VAR intervals coincide asymptotically. VARs with short-to-moderate lag lengths instead suffer from severe coverage distortions for small amounts of misspecification that is hard to rule out 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 generalize all our results 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*

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

where we define the norm

$$\|\alpha(L)\| \equiv \sqrt{\sum_{\ell=1}^{\infty}\text{trace}\{D\alpha_\ell' D^{-1}\alpha_\ell\}}.$$

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 $\text{aVar}(\hat{\delta}_h) > 0$. Then*

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

where we recall the definition $b_h = \text{aBias}(\hat{\delta}_h)/\sqrt{\text{aVar}(\hat{\delta}_h)}$.

Proof. The claim is a special case of Proposition 4.2 below. □

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, in terms of bias, there is "no free lunch" for VAR estimation: 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*

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

Proof. See Appendix B.5. □

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*

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

Proof. See Appendix B.6. □

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*

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

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).[6] 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 $M \geq 1$. It is only at the very right side of the figure—when the VAR includes enough lags 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 easily from the

[6] 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 Supplemental Appendix C for details.

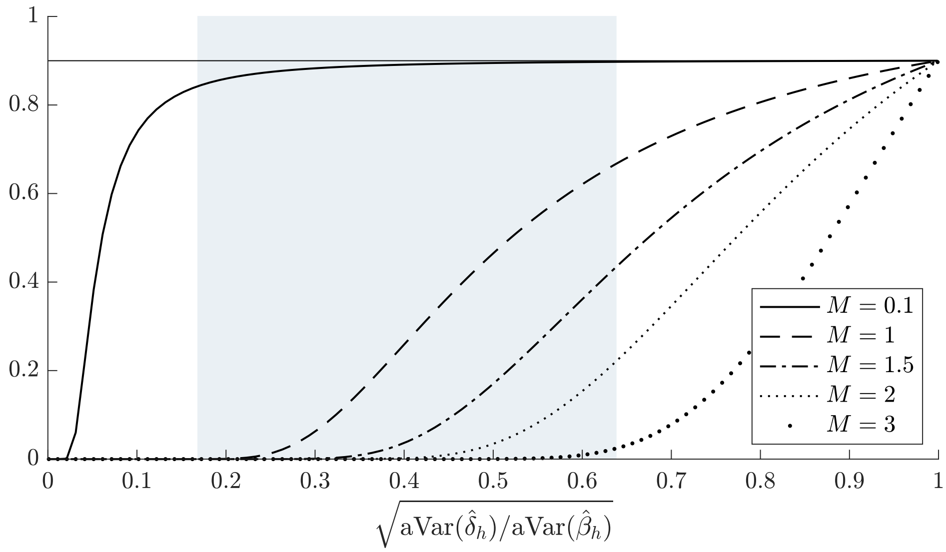

Figure 4.1: Worst-case asymptotic coverage probability of the conventional 90% VAR confidence interval. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. Different lines: different bounds M on $\|\alpha(L)\|$. Shaded area: empirical 10th–90th percentile range of relative standard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

Cauchy-Schwarz inequality (see the proof of Proposition 4.2 below):

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

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 types of misspecification.[7]

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 $\alpha(L)$.

Proof. See Appendix B.7. □

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 $\operatorname{aVar}(\hat{\delta}_h)/\operatorname{aVar}(\hat{\beta}_h)$. 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 $\operatorname{aVar}(\hat{\delta}_h)/\operatorname{aVar}(\hat{\beta}_h) \to 0$, 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.

[7] 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.

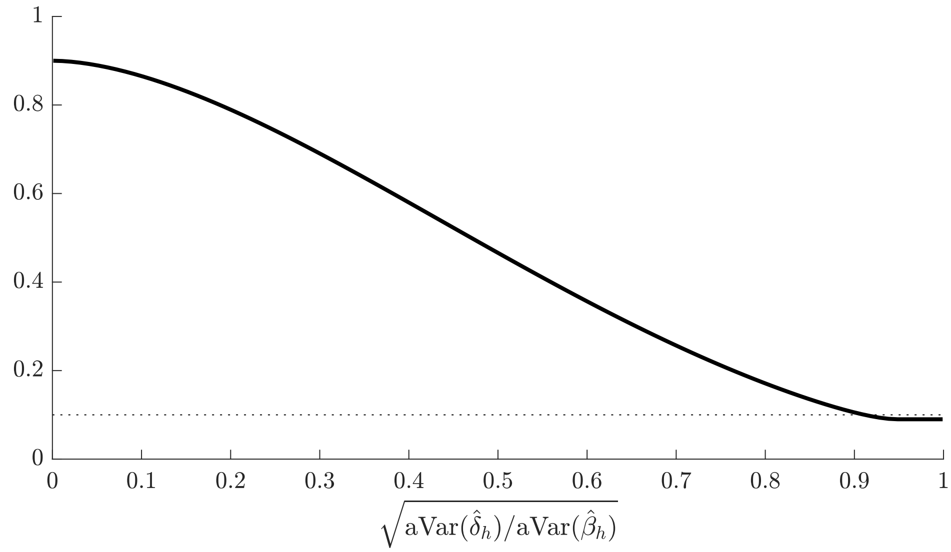

Figure 4.2: Worst-case asymptotic probability of the joint event that the conventional VAR confidence interval fails to cover the true impulse response and yet the Hausman test fails to reject misspecification. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. The dotted horizontal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

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

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

where the bias-aware critical value $\text{cv}_{1-a}(b)$ is given by the number such that $r(b; \text{cv}_{1-a}(b)) = a$, 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 $\text{aVar}(\hat{\delta}_h) > 0$. Then*

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

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

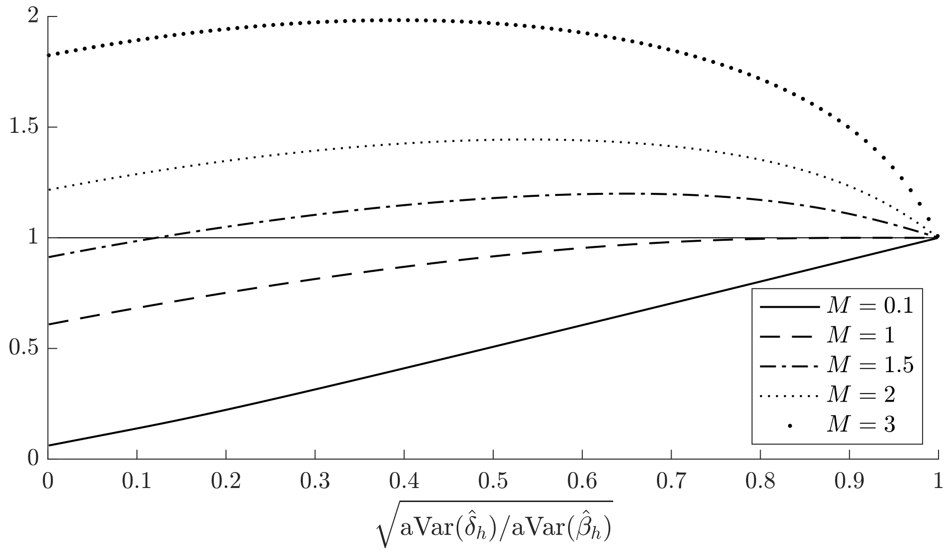

Figure 4.3: Relative length of bias-aware VAR confidence interval vs. conventional LP interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. Different lines: different bounds M on $\|\alpha(L)\|$. The solid horizontal line marks the value 1.

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, we now show that this is not the case by generalizing the worst-case bias formula to the multi-dimensional case and deriving the worst-case coverage of the Wald confidence ellipsoid.

Set-up. We now consider inference on any combination of impulse responses for various horizons h, response variables i^*, and shocks j^*. When referring to impulse responses and estimators of these, we need to make the response variable and shock explicit in the notation. Thus, we write $\theta_{i^*,j^*,h,T}$, $\hat{\beta}_{i^*,j^*,h}$, and $\hat{\delta}_{i^*,j^*,h}$, with the definitions being the same as in Section 2. Let k denote the total number of impulse responses of interest. We refer to the list of impulse responses by the collection of triples $\{(i^*_a, j^*_a, h_a)\}_{a=1}^k$ indexing the response variable, shock variable, and horizon, respectively. Define the k-dimensional vectors of true impulse responses and LP and VAR estimators:

$$\boldsymbol{\theta}_T \equiv \begin{pmatrix} \theta_{i^*_1,j^*_1,h_1,T} \\ \vdots \\ \theta_{i^*_k,j^*_k,h_k,T} \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} \equiv \begin{pmatrix} \hat{\beta}_{i^*_1,j^*_1,h_1} \\ \vdots \\ \hat{\beta}_{i^*_k,j^*_k,h_k} \end{pmatrix}, \quad \hat{\boldsymbol{\delta}} \equiv \begin{pmatrix} \hat{\delta}_{i^*_1,j^*_1,h_1} \\ \vdots \\ \hat{\delta}_{i^*_k,j^*_k,h_k} \end{pmatrix}.$$

It follows from Propositions 3.1 and 3.2 that

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\theta}_T \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\theta}_T \end{pmatrix} \overset{d}{\to} N\left(\begin{pmatrix} 0_{k\times 1} \\ \text{aBias}(\hat{\boldsymbol{\delta}}) \end{pmatrix}, \begin{pmatrix} \text{aVar}(\hat{\boldsymbol{\beta}}) & \text{aCov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}) \\ \text{aCov}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}) & \text{aVar}(\hat{\boldsymbol{\delta}}) \end{pmatrix}\right), \tag{4.2}$$

for a k-dimensional vector $\text{aBias}(\hat{\boldsymbol{\delta}})$ (defined in the proof of Proposition 4.2 below) and $k \times k$ matrices $\text{aVar}(\hat{\boldsymbol{\beta}})$, $\text{aVar}(\hat{\boldsymbol{\delta}})$, and $\text{aCov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}})$ given in Corollary A.2 in Appendix A.3. This corollary also implies that the difference $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\delta}}$ is asymptotically independent of $\hat{\boldsymbol{\delta}}$, which is not surprising given the general arguments of Hausman (1978) and the facts that (i) the asymptotic variances of the estimators are the same as in the model with $\alpha(L) = 0$ and (ii) the VAR estimator is the quasi-MLE in such a model. It follows that $\text{aVar}(\hat{\boldsymbol{\beta}}) \geq \text{aVar}(\hat{\boldsymbol{\delta}})$ in the positive semidefinite sense.

Worst-case bias. The following result generalizes the univariate worst-case bias formula in Proposition 4.1.

Proposition 4.2. *Impose Assumption 2.1, with part (iii) holding for all shock indices $j_1^*, \dots, j_k^*$, and let $\zeta = 1/2$. Let R be a constant matrix with k rows. Then*

$$\max_{\alpha(L)\colon \|\alpha(L)\| \le M} \|R \, \text{aBias}(\hat{\boldsymbol{\delta}})\|^2 = M^2 \lambda_{\max}\Big(R[\text{aVar}(\hat{\boldsymbol{\beta}}) - \text{aVar}(\hat{\boldsymbol{\delta}})]R'\Big),$$

where $\lambda_{\max}(B)$ denotes the largest eigenvalue of the matrix B.

Proof. See Appendix B.3. □

The proposition shows that the worst-case squared norm of the bias of the VAR estimator $R\hat{\boldsymbol{\delta}}$ of $R\boldsymbol{\theta}_T$ is a function of two simple quantities: the bound M on misspecification, and the largest eigenvalue of the difference $\text{aVar}(R\hat{\boldsymbol{\beta}}) - \text{aVar}(R\hat{\boldsymbol{\delta}})$ between the variance-covariance matrices for the LP and VAR estimators. The latter eigenvalue equals $\max_{\|\varsigma\|=1}\{\text{aVar}(\varsigma' R\hat{\boldsymbol{\beta}}) - \text{aVar}(\varsigma' R\hat{\boldsymbol{\delta}})\}$, i.e., the *largest* efficiency gain for VAR over LP across all linear combinations (with norm 1) of the estimated parameters. We thus see that there is still no free lunch: the worst-case bias is non-negligible if the VAR offers efficiency gains for *any* linear combination of the parameters of interest, echoing our univariate results. When R is a row vector, then the proposition implies that our conclusions from Section 4.2 extend to inference on any *linear combination* of impulse responses (across horizons, variables, and/or shocks). In particular, the VAR confidence interval has fragile coverage even if the target parameter is the integral (i.e., sum) of impulse responses across multiple horizons, as in the fiscal multiplier applications reviewed in Ramey (2016).

Worst-case coverage of confidence ellipsoid. We next derive the coverage of the conventional Wald confidence ellipsoid based on the VAR estimator. The level-$(1-a)$ confidence ellipsoid is given by

$$\text{CE}(\hat{\boldsymbol{\delta}}) \equiv \Big\{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^k \colon T(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\theta}})' \, \text{aVar}(\hat{\boldsymbol{\delta}})^{-1}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\theta}}) \le \chi^2_{1-a,k}\Big\},$$

where $\chi^2_{1-a,k}$ is the $1-a$ quantile of the χ^2 distribution with k degrees of freedom.

Corollary 4.6. *Impose Assumption 2.1, with part (iii) holding for all shock indices $j_1^*, \dots, j_k^*$, and let $\zeta = 1/2$. Assume also that $\text{aVar}(\hat{\boldsymbol{\delta}})$ is non-singular. Then*

$$\min_{\alpha(L)\colon \|\alpha(L)\| \le M} \lim_{T\to\infty} P(\boldsymbol{\theta}_T \in \text{CE}(\hat{\boldsymbol{\delta}})) = F_k\left(\chi^2_{1-a,k}; M^2\left[\lambda_{\max}(\text{aVar}(\hat{\boldsymbol{\beta}})\, \text{aVar}(\hat{\boldsymbol{\delta}})^{-1}) - 1\right]\right),$$

where $F_k(x;c)$ is the cumulative distribution function, evaluated at point $x \geq 0$, of a non-central χ^2 distribution with k degrees of freedom and non-centrality parameter $c \geq 0$.

Proof. See Appendix B.8. □

The worst-case coverage probability of the VAR confidence ellipsoid depends on three scalars: the bound M on misspecification, the dimension k of the ellipsoid, and the "multivariate relative standard error"

$$\sqrt{\lambda_{\min}(\text{aVar}(\hat{\boldsymbol{\delta}})\,\text{aVar}(\hat{\boldsymbol{\beta}})^{-1})} = [\lambda_{\max}(\text{aVar}(\hat{\boldsymbol{\beta}})\,\text{aVar}(\hat{\boldsymbol{\delta}})^{-1})]^{-1/2} = \min_{\varsigma\in\mathbb{R}^k}\sqrt{\text{aVar}(\varsigma'\hat{\boldsymbol{\delta}})/\,\text{aVar}(\varsigma'\hat{\boldsymbol{\beta}})}.$$

Again, the worst-case coverage distortion is an increasing function of the *largest* efficiency gain for VAR over LP across *all* linear combinations of the impulse responses. Since VAR impulse response estimates are often highly correlated across horizons, this suggests that the VAR undercoverage can in fact be particularly severe in the multivariate case.

Figures 4.4 and 4.5 show that the worst-case coverage probability of the confidence ellipsoid can be very poor even for small amounts of misspecification. The panels in the figures correspond to choices of the dimension $k \in \{2, 5, 10\}$ or the bound $M \in \{1, 1.5, 2\}$ on misspecification, respectively. Evidently, the coverage distortions can be severe for all of the k considered in these plots even when $M = 1$ (corresponding to a noise-to-signal ratio of $1/T$), especially if the multivariate relative standard error $\sqrt{\lambda_{\min}(\text{aVar}(\hat{\boldsymbol{\delta}})\,\text{aVar}(\hat{\boldsymbol{\beta}})^{-1})} \leq 1/4$. It further appears that the worst-case coverage distortion is monotonically decreasing in k, holding everything else constant; however, loosely speaking, the larger is k, the larger is the "chance" that there is some linear combination of the parameters for which VAR is much more efficient than LP, instead yielding a larger coverage distortion. The multivariate case thus overall echoes and reinforces our conclusions from the earlier univariate analysis.

5 Simulations

We now show that our asymptotic results accurately reflect the finite-sample properties of LP and VAR procedures. To do so, we consider a standard structural macroeconomic model as our DGP; in Supplemental Appendix D.1 we present further illustrative results from a simple univariate model.

Framework. Our DGP is the well-known structural macroeconomic model of Smets and Wouters (2007). This environment is ideal for our purposes, as it allows us to closely mimic

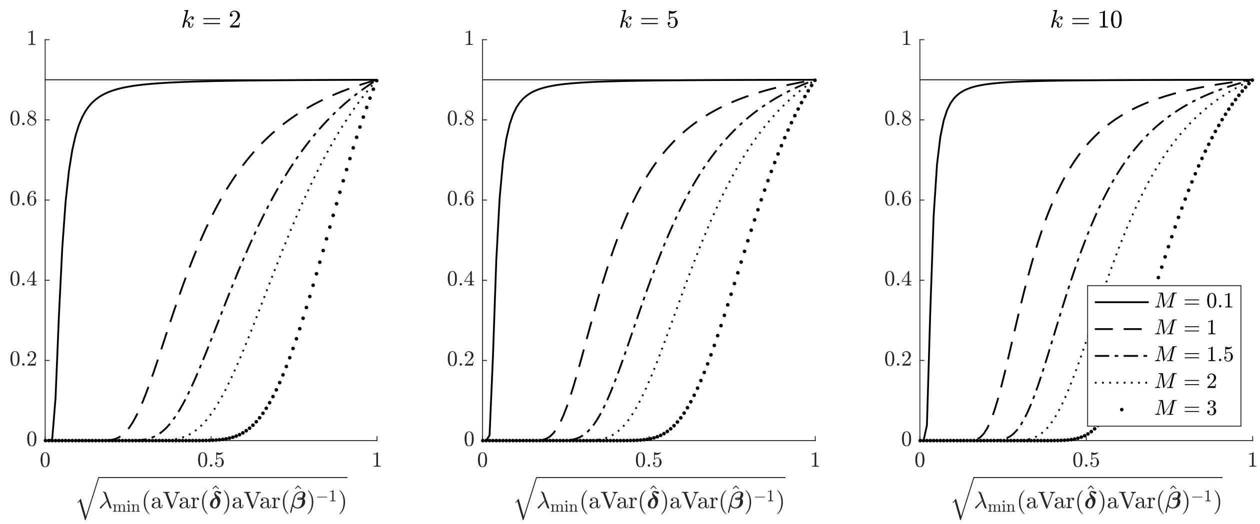

Figure 4.4: Worst-case asymptotic coverage probability of the conventional 90% VAR confidence ellipsoid. Different panels: choices of the number k of parameters in the ellipsoid. Horizontal axes: square root of smallest eigenvalue of the ratio of the VAR and LP variance-covariance matrices. Different lines: different bounds M on $\|\alpha(L)\|$. The solid horizontal lines mark the nominal coverage probability $1 - a = 90\%$.

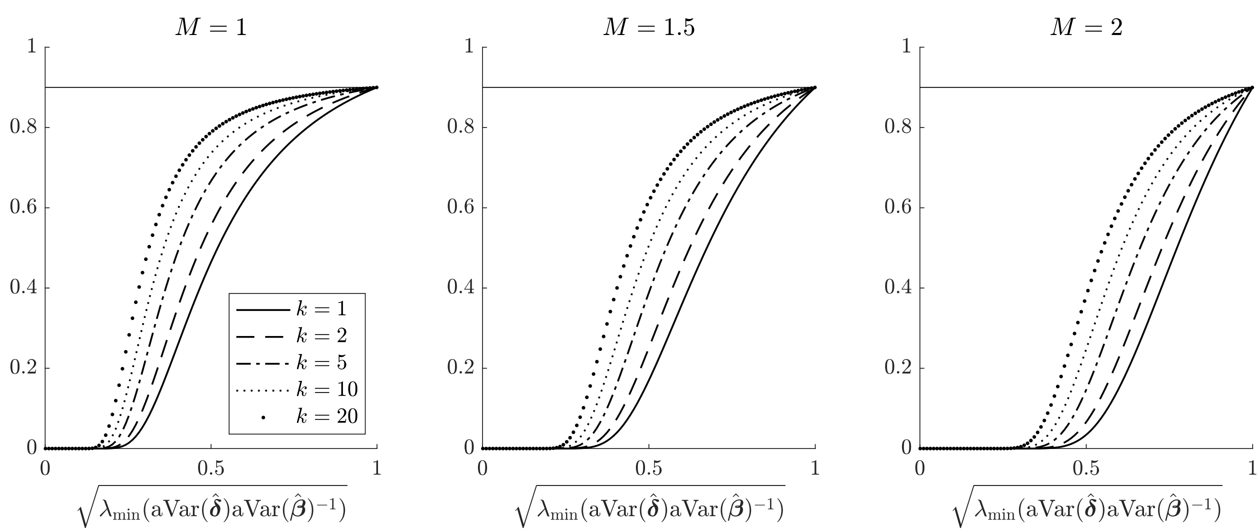

Figure 4.5: Worst-case asymptotic coverage probability of the conventional 90% VAR confidence ellipsoid. Different panels: different bounds M on $\|\alpha(L)\|$. Horizontal axes: square root of smallest eigenvalue of the ratio of the VAR and LP variance-covariance matrices. Different lines: choices of the number k of parameters in the ellipsoid. The solid horizontal lines mark the nominal coverage probability $1 - a = 90\%$.

p	M	$\frac{M^2}{1+M^2}$
1	34.053	0.999
2	5.412	0.967
4	3.225	0.912
8	1.893	0.782
12	1.324	0.637
20	0.915	0.456
40	0.614	0.274

Table 5.1: M and $\frac{M^2}{1+M^2}$ as functions of p in the model of Smets and Wouters, with the researcher observing the cost-push shock, inflation, wages, and hours worked, and estimating a VAR(p).

applied macroeconometric practice: the model is rich enough to match quite well the second-moment properties of standard macroeconomic time series data, yet at the same time features the interpretable, structural shocks typically studied in applied work. We solve the model at the posterior mode estimated by Smets and Wouters. For our main exercise we will assume that the econometrician observes the wage cost-push shock, inflation, wages, and total hours worked. The impulse response function of interest is that of inflation with respect to the cost-push shock. In addition to being topical (e.g., see the recent work of Bernanke and Blanchard, 2023), this exercise is well-suited to cleanly illustrate the potential bite of our theoretical conclusions: wage cost-push shocks in the model of Smets and Wouters follow an ARMA(1,1) process, and so short-lag VARs may be subject to material biases. Results for a monetary shock specification are reported in Supplemental Appendix D.2.

Results. We begin by quantifying the amount of VAR misspecification in this DGP. Given a choice of lag length p, we represent the Smets and Wouters model in VARMA(p, ∞) form (2.1) with $T = 240$ and $\zeta = 1/2$, where the VAR coefficients are selected to minimize the population sum of squared residuals.[8] Table 5.1 shows, as a function of p, the total degree of misspecification $M = \|\alpha(L)\|$ as well as $M^2/(1 + M^2)$, the minimax optimal *ex ante* model averaging weight on LP in Corollary 4.2. As anticipated, the larger p, the smaller M. Importantly, however, M only declines extremely slowly with the lag length p. In particular, for lag lengths typical in applied practice (for quarterly data), misspecification is material,

[8] Let $y_t = C(L)u_t$ be the Wold decomposition of the observed data. Denote the linear least-squares projection coefficients of y_t onto p of its lags as $A_\ell(p)$, $\ell = 1, 2, \dots, p$, and define $A(L; p) \equiv I - \sum_{\ell=1}^{p} A_\ell(p)$. Letting $\tilde{C}(L; p) = \sum_{\ell=0}^{\infty} \tilde{C}_\ell(p) L^\ell \equiv A(L; p)C(L)$, we obtain a representation $y_t = \sum_{\ell=1}^{p} A_\ell(p) y_{t-\ell} + \sum_{\ell=0}^{\infty} \tilde{C}_\ell(p) u_{t-\ell}$. It is straightforward to orthogonalize the innovations in this representation and write it in the form (2.1), thus yielding the implied MA polynomial $\alpha(L)$.

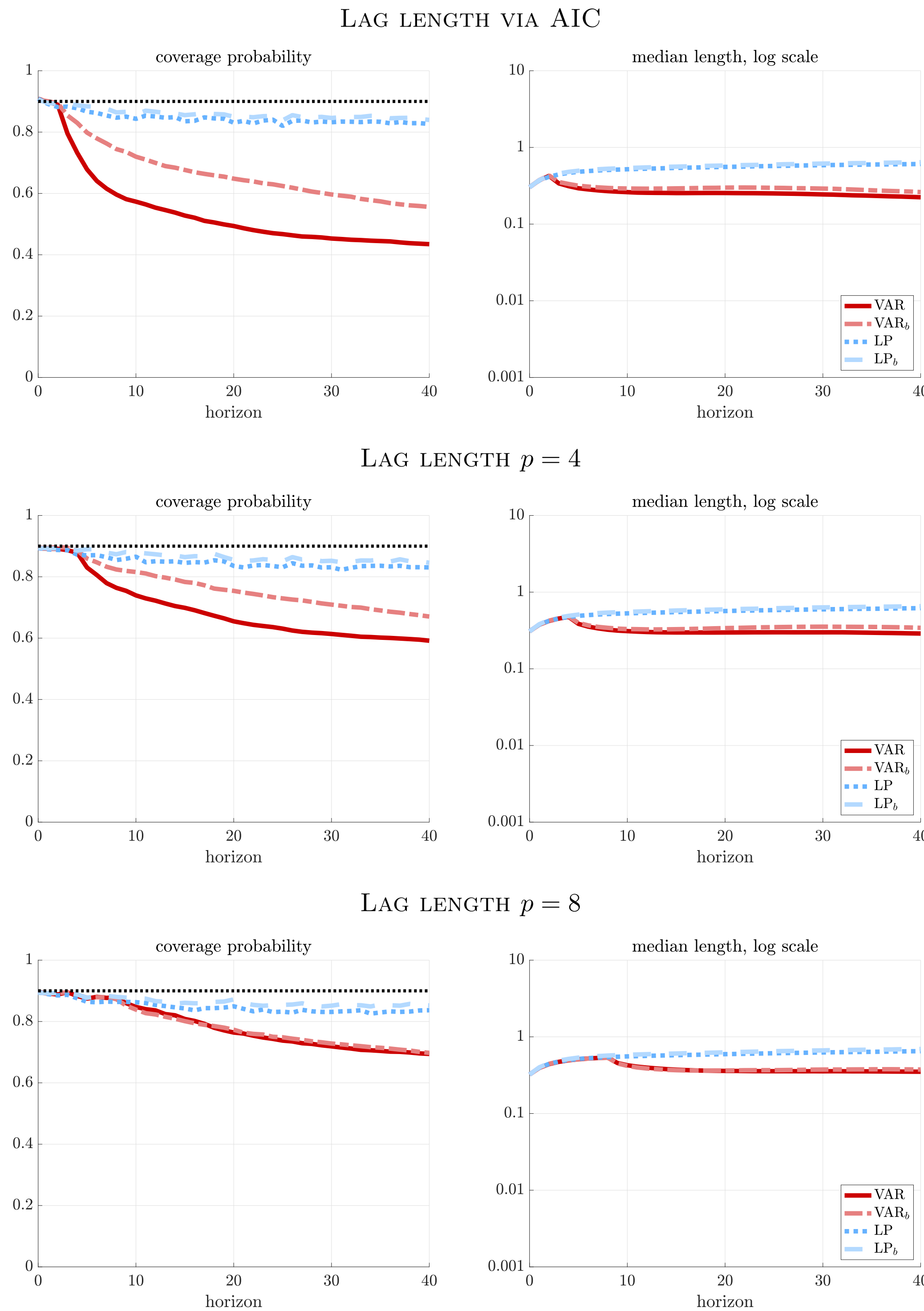

Figure 5.1: Coverage probabilities (left) and median confidence interval length (right) for VAR (red) and LP (blue) confidence intervals computed via the delta method or bootstrap (the latter are indicated with subscript "b" in the figure legends). DGP: Smets and Wouters (2007), cost-push shock. Lag length p is selected using the AIC for the top panel, and fixed at $p = 4$ and $p = 8$ for the middle and bottom panels, respectively.

with $M \approx 3.23$ for a standard lag length of $p = 4$, and $M \approx 1.89$ even for a long lag length of $p = 8$. This suggests that—depending on the shape of $\alpha(L)$—there is potential for misspecification to substantially affect VAR inference.

Figure 5.1 shows that VAR confidence intervals indeed severely undercover, while LP intervals are robust. We simulate 5,000 samples of size $T = 240$, and for each construct delta method as well as bootstrap LP and VAR confidence intervals (assuming homoskedasticity).[9] The top panel considers the empirically common case where the VAR lag length p is selected using the AIC; we then use the same p for LP. At all but very short horizons, VAR confidence intervals materially undercover, while LP throughout attains close to the nominal coverage level, consistent with our theoretical results. The bootstrap somewhat improves the VAR coverage, but large distortions remain.[10] We emphasize that these results for VAR inference are obtained even though the lag length p is selected using the AIC: the median selected lag length is $p = 2$, which here is evidently insufficient to guard against VAR bias and undercoverage.[11] The middle and bottom panels illustrate our "no free lunch" result. For those panels, we instead manually select longer lag lengths: $p = 4$ for the middle panel and $p = 8$ for the bottom panel. VAR coverage is now closer to the nominal level for all horizons $h \leq p$ (consistent with Corollary 3.2), but at the same time confidence intervals become essentially as wide as for LP. At longer horizons we obtain the same picture as before: substantial undercoverage for VAR, and coverage close to the nominal level for LP.

Supplemental Appendix D.2 shows that these conclusions extend to monetary shock specifications. We furthermore show that the VAR coverage distortions in the actual Smets and Wouters DGP are comparable to those in a slightly perturbed DGP where we replace the model-implied MA lag polynomial $\alpha(L)$ by the theoretical least favorable one $\alpha^{\dagger}(L)$; thus, the latter is not particularly pathological, consistent with our discussion in Section 4.2.

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

[9] We construct equal-tailed percentile-t bootstrap confidence intervals, using 2,000 bootstrap draws. See Inoue and Kilian (2002) for VAR and Montiel Olea and Plagborg-Møller (2021) for LP.

[10] Indeed, the coverage distortions are not primarily a small-sample phenomenon: Supplemental Appendix D.2 shows that similar distortions arise with $T = 2{,}000$.

[11] This finding is consistent with Kilian and Lütkepohl (2017, Chapter 2.6.5), who emphasize that standard lag order selection criteria such as the AIC often tend to select too few lags for accurate VAR coverage.

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 even if the researcher directly observes a near-perfect proxy for the shock of interest. However, it is *not* necessary to get the lag length exactly right to achieve correct coverage. To select the number of lags to control for in the LP, it suffices to run a VAR in all variables used in the analysis and select the lag length that minimizes conventional information criteria (such as AIC). Our results complement the finding of Montiel Olea and Plagborg-Møller (2021) that lag-augmented LP confidence intervals are also more robust than VAR intervals to persistence in the data and to the length of the impulse response horizon.

3. There is no free lunch for VARs: if 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.[12] 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 is economically plausible. 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

[12] 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.

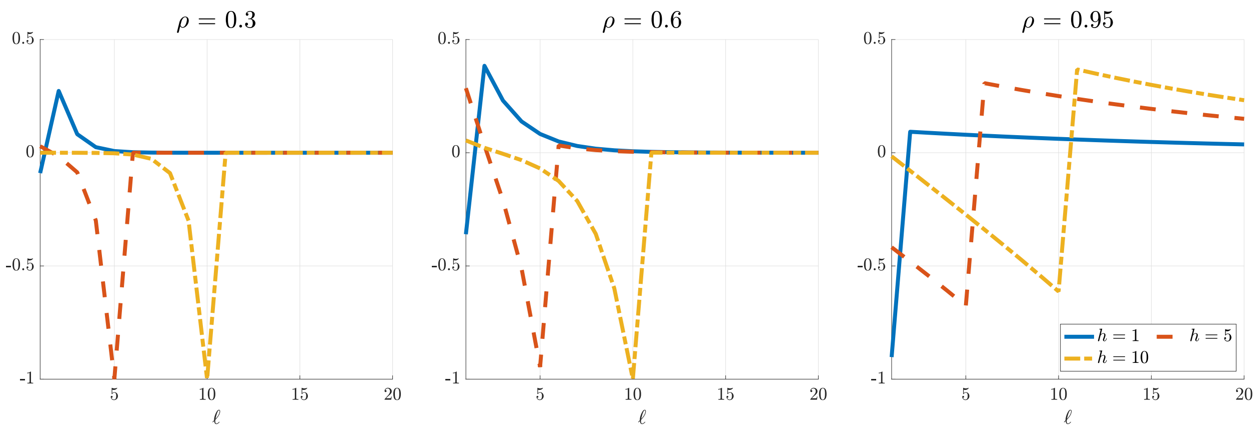

Figure A.1: Least favorable $\alpha^\dagger(L; h)$ for horizons $h \in \{1, 5, 10\}$ for local-to-AR(1) models with different persistence parameters ρ (left, middle, and right panel).

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.

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 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 $\omega \in [0, 1]$,*

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

Proof. The result follows from Propositions 3.1, 3.2 and 4.1 and 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

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

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 $M \geq 2$, 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 $M \geq 1.5$, 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.

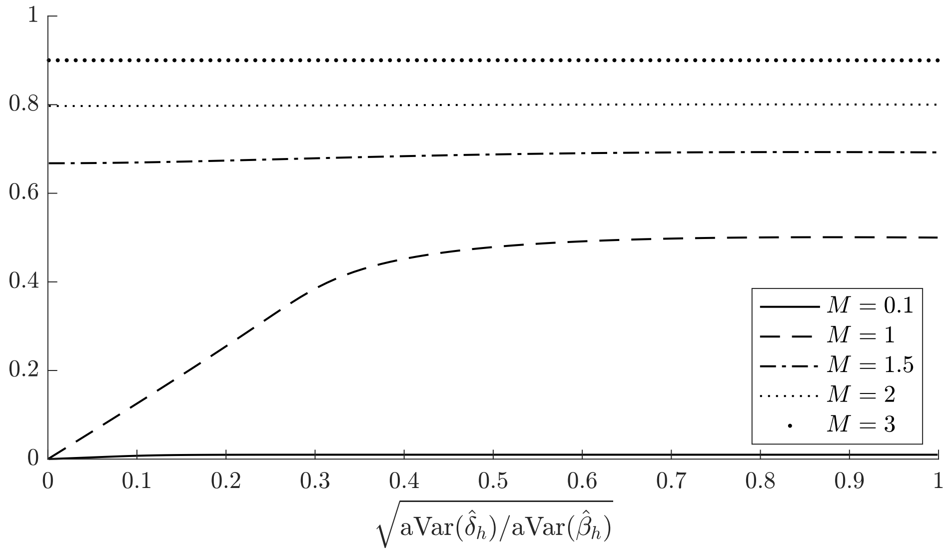

Figure A.2: Length-optimal weight on LP in bias-aware confidence interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. Different lines: different bounds M on $\|\alpha(L)\|$.

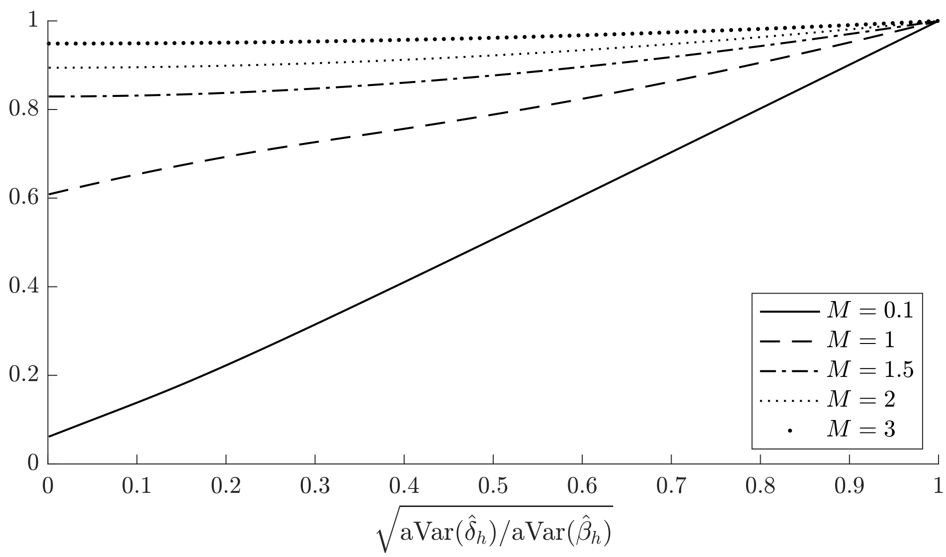

Figure A.3: Relative length of optimal bias-aware confidence interval vs. conventional LP interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. Different lines: different bounds M on $\|\alpha(L)\|$.

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

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

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 is understood to refer to elements of the asymptotic variance-covariance matrix in (4.2). In particular, $\text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a} - \hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) = 0$.

Proof. See Appendix B.9. □

Appendix B Proofs

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

B.1 Proof of Proposition 3.1

Lemma E.1 shows that we can represent

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

where the precise expressions for the coefficient matrices are given in Lemma E.1. The lemma also shows that the $1 \times n$ two-sided lag polynomial $\Theta_h(L) = \sum_{\ell=-\infty}^{\infty} \Theta_{h,\ell}L^\ell$ is absolutely summable and satisfies $\Theta_{h,0,j^*} = 0$. That is, $\Theta_h(L)\varepsilon_t$ is independent of $\varepsilon_{j^*,t}$ (but not of $\varepsilon_{j^*,s}$ for $s \neq t$).

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

$$\hat{\beta}_h = \frac{\sum_{t=1}^{T-h} y_{i^*,t+h}\hat{x}_{h,t}}{\sum_{t=1}^{T-h} \hat{x}_{h,t}^2}$$

$$= \theta_{h,T} + \frac{\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\hat{x}_{h,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t}}{\sum_{t=1}^{T-h}\hat{x}^2_{h,t}} \quad \text{by orthogonality}$$

$$= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\hat{x}_{h,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t}}{\sigma^2_{j^*} + o_p(1)} \quad \text{by Lemma E.4(v)}$$

$$= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\varepsilon_{j^*,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t} + o_p(T^{-1/2})}{\sigma^2_{j^*} + o_p(1)} \quad \text{by Lemma E.4(iv)}$$

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

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

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

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

Finally, Lemma E.1 shows that

$$T^{-1}\sum_{t=1}^{T-h}(\Theta_h(L)\varepsilon_t)\varepsilon_{j^*,t} = O_p(T^{-1/2}).$$

Thus, we conclude that

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

which completes the proof. □

B.2 Proof of Proposition 3.2

Note first that

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

Lemma E.2 shows that $\hat{A} - A = O_p(T^{-\zeta} + T^{-1/2})$. Since it is known that

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

see for example Magnus and Neudecker (2007, Table 7, p. 208), the delta method gives

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

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

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

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

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

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

B.3 Proof of Proposition 4.2

Define $\tilde{\alpha}_\ell = D^{-1/2}\alpha_\ell D^{1/2}$ for all $\ell \geq 1$. Notice that $\|\alpha(L)\|^2 = \sum_{\ell=1}^{\infty}\|\tilde{\alpha}_\ell\|^2$.

By Proposition 3.2, we have $\operatorname{aBias}(\hat{\delta}_{i^*,j^*,h}) = \sum_{\ell=1}^{\infty}\operatorname{trace}(\Xi_{i^*,j^*,h,\ell}\tilde{\alpha}_\ell)$, where

$$\Xi_{i^*,j^*,h,\ell} \equiv D^{1/2}H'(A')^{\ell-1}S^{-1}\Psi_{i^*,j^*,h}HD^{1/2} - \mathbb{1}(\ell \leq h)D^{-1/2}e_{j^*,m}e'_{i^*,n}A^{h-\ell}HD^{1/2}.$$

Since $\operatorname{trace}(\Xi_{i^*,j^*,h,\ell}\tilde{\alpha}_\ell) = \operatorname{vec}(\Xi_{i^*,j^*,h,\ell})'\operatorname{vec}(\tilde{\alpha}'_\ell)$, we can write

$$\operatorname{aBias}(\hat{\boldsymbol{\delta}}) = \sum_{\ell=1}^{\infty}\Upsilon_\ell \operatorname{vec}(\tilde{\alpha}'_\ell),$$

where

$$\Upsilon_\ell \equiv \left(\operatorname{vec}(\Xi_{i_1^*,j_1^*,h_1,\ell}), \ldots, \operatorname{vec}(\Xi_{i_k^*,j_k^*,h_k,\ell})\right)' \in \mathbb{R}^{k\times m^2}.$$

Hence,

$$\max_{\alpha(L)\colon \|\alpha(L)\|\le M} \|R\,\text{aBias}(\hat{\boldsymbol{\delta}})\|^2 = \max_{\{\tilde{\alpha}_\ell\}_{\ell=1}^\infty\colon \sum_{\ell=1}^\infty \|\tilde{\alpha}_\ell'\|^2\le M^2} \left\| \sum_{\ell=1}^\infty R\Upsilon_\ell \operatorname{vec}(\tilde{\alpha}_\ell') \right\|^2.$$

Lemma E.5 shows that the final expression above equals $M^2\lambda_{\max}(\sum_{\ell=1}^\infty R\Upsilon_\ell\Upsilon_\ell' R')$ (the lemma only explicitly considers the case $M=1$, but the general case then follows from the homogeneity of degree 1 of the norm). Finally, Lemma B.1 below shows that $\sum_{\ell=1}^\infty \Upsilon_\ell\Upsilon_\ell' = \text{aVar}(\hat{\boldsymbol{\beta}}) - \text{aVar}(\hat{\boldsymbol{\delta}})$. This completes the proof of the proposition. The proof of Lemma E.5 shows that the maximum above is achieved when $\operatorname{vec}(\tilde{\alpha}_\ell') \propto \Upsilon_\ell' v$ (with the constant of proportionality being independent of ℓ and chosen to satisfy the norm constraint), where v is the eigenvector corresponding to the largest eigenvalue of $R[\text{aVar}(\hat{\boldsymbol{\beta}}) - \text{aVar}(\hat{\boldsymbol{\delta}})]R'$. In the univariate case $k=1$, this reduces to expression (4.1) in Section 4.2. □

Lemma B.1. *Under the assumptions of Proposition 4.2, and using the notation in the proof of that proposition, we have*

$$\sum_{\ell=1}^\infty \Upsilon_\ell\Upsilon_\ell' = \text{aVar}(\hat{\boldsymbol{\beta}}) - \text{aVar}(\hat{\boldsymbol{\delta}}).$$

Proof. By definition of Υ_ℓ, it suffices to show that, for any indices $a,b\in\{1,\dots,k\}$,

$$\sum_{\ell=1}^\infty \operatorname{vec}(\Xi_{i_a^*,j_a^*,h_a,\ell})' \operatorname{vec}(\Xi_{i_b^*,j_b^*,h_b,\ell}) = \text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_b^*,h_b}) - \text{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}). \tag{B.2}$$

Multiplying out terms, we find that the left-hand side above equals

$$\begin{aligned}
\sum_{\ell=1}^\infty \operatorname{trace}(\Xi_{i_a^*,j_a^*,h_a,\ell}'\Xi_{i_b^*,j_b^*,h_b,\ell}) = &\sum_{\ell=1}^\infty \operatorname{trace}\left(A^{\ell-1}\Sigma(A')^{\ell-1}S^{-1}\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi_{i_a^*,j_a^*,h_a}'S^{-1}\right)\\
&-\sum_{\ell=1}^{h_b} \operatorname{trace}\left(A^{\ell-1}H_{\bullet,j_b^*}e_{i_b^*,n}'A^{h_b-\ell}\Sigma\Psi_{i_a^*,j_a^*,h_a}'S^{-1}\right)\\
&-\sum_{\ell=1}^{h_a} \operatorname{trace}\left(A^{\ell-1}H_{\bullet,j_a^*}e_{i_a^*,n}'A^{h_a-\ell}\Sigma\Psi_{i_b^*,j_b^*,h_b}'S^{-1}\right)\\
&+\mathbb{1}(j_a^*=j_b^*)\sigma_{j_a^*}^{-2}\sum_{\ell=1}^{\min\{h_a,h_b\}} e_{i_b^*,n}'A^{h_b-\ell}\Sigma(A')^{h_a-\ell}e_{i_a^*,n}.
\end{aligned}$$

We now evaluate each of the four terms on the right-hand side above. The *first* term equals

$$\operatorname{trace}\Bigg(\underbrace{\sum_{\ell=1}^{\infty} A^{\ell-1}\Sigma(A')^{\ell-1}}_{=S} S^{-1}\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi'_{i_a^*,j_a^*,h_a}S^{-1}\Bigg) = \operatorname{trace}\left(\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi'_{i_a^*,j_a^*,h_a}S^{-1}\right).$$

The *second* term (in the earlier display) equals

$$-\operatorname{trace}\Bigg(\underbrace{\sum_{\ell=1}^{h_b} A^{\ell-1}H_{\bullet,j_b^*}e'_{i_b^*,n}A^{h_b-\ell}}_{=\sum_{\ell=1}^{h_b}A^{h_b-\ell}H_{\bullet,j_b^*}e'_{i_b^*,n}A^{\ell-1}=\Psi_{i_b^*,j_b^*,h_b}} \Sigma\Psi'_{i_a^*,j_a^*,h_a}S^{-1}\Bigg) = -\operatorname{trace}\left(\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi'_{i_a^*,j_a^*,h_a}S^{-1}\right),$$

and the *third* term (in the earlier display) also equals this quantity by a symmetric calculation. In conclusion, we have shown

$$\begin{aligned}
&\sum_{\ell=1}^{\infty}\operatorname{trace}(\Xi'_{i_a^*,j_a^*,h_a,\ell}\Xi_{i_b^*,j_b^*,h_b,\ell}) \\
&= \mathbb{1}(j_a^*=j_b^*)\sigma_{j_a^*}^{-2}\sum_{\ell=1}^{\min\{h_a,h_b\}} e'_{i_b^*,n}A^{h_b-\ell}\Sigma(A')^{h_a-\ell}e_{i_a^*,n} - \operatorname{trace}\left(\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi'_{i_a^*,j_a^*,h_a}S^{-1}\right).
\end{aligned}$$

The desired result (B.2) now follows from Corollary A.2. □

B.4 Proof of Corollary 3.2

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

$$\begin{aligned}
\sigma_{j^*}^2\Psi'_h S^{-1}\tilde{y}_{t-1} &= \Bigg(\sum_{\ell=1}^{h}(A')^{h-\ell}e_{i^*,n}\underbrace{\sigma_{j^*}^2 H'_{\bullet,j^*}(A')^{\ell-1}}_{=\operatorname{Cov}(\varepsilon_{j^*,t-\ell},\tilde{y}_{t-1})}\Bigg)S^{-1}\tilde{y}_{t-1} \\
&= \sum_{\ell=1}^{h}(A')^{h-\ell}e_{i^*,n}E^*(\varepsilon_{j^*,t-\ell} \mid \tilde{y}_{t-1}) \\
&= \sum_{\ell=1}^{h}(A')^{h-\ell}e_{i^*,n}\varepsilon_{j^*,t-\ell},
\end{aligned}$$

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

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

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

B.5 Proof of Corollary 4.1

By Proposition 4.1,

$$\sup_{\alpha(L) \colon \|\alpha(L)\| \leq M} \text{aBias}(\hat{\delta}_h; \alpha(L))^2 = M^2 \{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}.$$

Thus,

$$\begin{aligned}
&\sup_{\alpha(L) \colon \|\alpha(L)\| \leq M} \text{aMSE}(\hat{\delta}_h; \alpha(L)) - \text{aMSE}(\hat{\beta}_h) \\
&= M^2 \{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\} + \text{aVar}(\hat{\delta}_h) - \text{aVar}(\hat{\beta}_h) \\
&= (M^2 - 1) \{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}. \quad \square
\end{aligned}$$

B.6 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.7 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.8 Proof of Corollary 4.6

The result follows straightforwardly from (4.2) if we can show that the maximal non-centrality parameter equals

$$\max_{\alpha(L)\colon \|\alpha(L)\| \leq M} \text{aBias}(\hat{\boldsymbol{\delta}})' \, \text{aVar}(\hat{\boldsymbol{\delta}})^{-1} \, \text{aBias}(\hat{\boldsymbol{\delta}}) = M^2 \left[\lambda_{\max}(\text{aVar}(\hat{\boldsymbol{\beta}}) \, \text{aVar}(\hat{\boldsymbol{\delta}})^{-1}) - 1\right].$$

But this follows from applying Proposition 4.2 with $R = \text{aVar}(\hat{\boldsymbol{\delta}})^{-1/2}$, since

$$\begin{aligned}
&\lambda_{\max}\Big(\text{aVar}(\hat{\boldsymbol{\delta}})^{-1/2}[\text{aVar}(\hat{\boldsymbol{\beta}}) - \text{aVar}(\hat{\boldsymbol{\delta}})] \, \text{aVar}(\hat{\boldsymbol{\delta}})^{-1/2\prime}\Big) \\
&= \lambda_{\max}\Big(\text{aVar}(\hat{\boldsymbol{\delta}})^{-1/2} \, \text{aVar}(\hat{\boldsymbol{\beta}}) \, \text{aVar}(\hat{\boldsymbol{\delta}})^{-1/2\prime} - I_k\Big) \\
&= \lambda_{\max}\Big(\text{aVar}(\hat{\boldsymbol{\delta}})^{-1/2} \, \text{aVar}(\hat{\boldsymbol{\beta}}) \, \text{aVar}(\hat{\boldsymbol{\delta}})^{-1/2\prime}\Big) - 1 \\
&= \lambda_{\max}\Big(\text{aVar}(\hat{\boldsymbol{\beta}}) \, \text{aVar}(\hat{\boldsymbol{\delta}})^{-1}\Big) - 1. \quad \square
\end{aligned}$$

B.9 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 $\xi_{j^*,h,t} = (\xi_{1,j^*,h,t}, \ldots, \xi_{n,j^*,h,t})'$ 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 $j_a^* = j_b^*$:

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

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

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

$$= \operatorname{trace}\left(\Psi_{i_a^*,j_a^*,h_a}\Sigma\Psi'_{i_b^*,j_b^*,h_b}S^{-1}\right),$$

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

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

The first term on the left-hand side above equals

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

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

Bernanke, B., and O. Blanchard (2023): "What caused the US pandemic-era inflation?," Peterson Institute for International Economics Working Paper 23-4.

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 (2024): "VAR Hyperparameter Determination Under Misspecification," Manuscript, University of Pennsylvania.

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.

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.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2024): “Local Projections vs. VARs: Lessons From Thousands of DGPs,” *Journal of Econometrics*, forthcoming.

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.

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.

Newey, W. K. (1990): "Semiparametric efficiency bounds," *Journal of Applied Econometrics*, 5(2), 99–135.

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.

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.

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

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.