Double Robustness of Local Projections
and Some Unpleasant VARithmetic*

José Luis Montiel Olea   Mikkel Plagborg-Møller

Cornell University   Princeton University

Eric Qian   Christian K. Wolf

Princeton University   MIT & NBER

May 15, 2024

Abstract: We consider impulse response inference in a locally misspecified stationary 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 follows from a "double robustness" property analogous to that of modern estimators for partially linear regressions. In contrast, VAR confidence intervals dramatically undercover even for misspecification so small that it is difficult to detect statistically and cannot be ruled out based on economic theory. This is because of a "no free lunch" result for VARs: the worst-case bias and coverage distortion are small if, and only if, the variance is close to that of LP. While VAR coverage can be restored by using a bias-aware critical value or a large lag length, the resulting confidence interval tends to be at least as wide as the LP interval.

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

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., see 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).

The main contribution of this paper is to provide a formal proof of the claim in Jordà (2005) that conventional LP confidence intervals for impulse responses are surprisingly robust to misspecification. At the same time, we show that VAR confidence intervals are unreliable in the face of even small amounts of misspecification.

To formalize our results, 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 of models covers many types of dynamic misspecification, such as under-specification of the lag length, failure to include relevant control variables, inappropriate aggregation, measurement error, and omitted nonlinearities. In particular, such an environment is consistent with essentially all linearized structural macroeconomic models (Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007). Intuitively, our set-up with local misspecification captures the idea that finite-order VAR models provide a good—but not perfect—approximation of reality.

In this setting, we prove that the conventional LP confidence interval has correct (pointwise) 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 (Newey, 1990; Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2018; Chernozhukov, Escanciano, Ichimura, Newey, and Robins, 2022).

In stark contrast to LP, small amounts of misspecification cause conventional VAR confidence intervals for impulse responses to suffer from severe under-coverage asymptotically. We derive analytically the worst-case bias and coverage of VARs over all possible misspecification processes, subject to a constraint only on the overall magnitude of the misspecification. 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. This worst-case result is practically relevant—VAR confidence intervals severely undercover even when the misspecification term: (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. While the coverage can be restored by using a larger bias-aware critical value (Armstrong and Kolesár, 2021) or by using a large number of lags for estimation, 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 first consider a univariate ARMA(1,1) model. While LP confidence intervals here have coverage close to the nominal level, confidence intervals based on short-lag autoregressions severely undercover when the MA coefficient is only moderately large. In fact, in this simple case the implied MA misspecification is very close to the theoretical worst case. We next run simulations based on the medium-scale structural macroeconomic model of Smets and Wouters (2007). The researcher estimates the impulse response function of inflation with respect to an observed cost-push shock, with lag length selected by the Akaike Information Criterion (AIC). VAR confidence intervals materially undercover—particularly at medium and long horizons—while LP again 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.[1]

OTHER LITERATURE.   Relative to the previously cited simulation studies of LPs and VARs, we derive *analytical* results on the worst-case asymptotic properties of these inference procedures that hold for a wide range of stationary, locally misspecified VAR models. The empirically calibrated simulations in Li, Plagborg-Møller, and Wolf (2024) suggest a stark bias-variance trade-off between LP (low bias, high variance) and 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, then even a moderate amount of VAR bias cannot be tolerated, as it causes the VAR confidence interval to be poorly centered. Effectively, a concern for confidence interval coverage induces a large weight on bias in the researcher's objective function, justifying the use of LP despite its higher variance.

The robustness of LP to misspecification discussed here—with stationary data and at fixed horizons—is conceptually and theoretically distinct from the robustness of LP 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 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, but do not provide analytical results on the performance of inference procedures. 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 $T^{-1/2}$. 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 formulas for the worst-case bias and coverage of VARs. As such, our results formalize concerns by some applied practitioners about the lack of VAR robustness (e.g., Chari, Kehoe, and McGrattan, 2008; Nakamura and Steinsson, 2018).

Whereas our paper deals with bias imparted by dynamic misspecification, our asymptotics

---

[1]We obtain similar results for simulations based on other observed shocks (including monetary shocks) and other shock identification schemes (including recursive identification).

abstract from the order-T^{-1} bias of LPs and VARs that is due to persistence in the data (Pope, 1990; Kilian, 1998; Herbst and Johannsen, 2023).

OUTLINE. Section 2 provides an overview of our main results in a univariate model. Section 3 introduces the general local-to-SVAR(p) model, defines LP and VAR estimators, and proves the robustness of LP and fragility of VAR confidence intervals. Section 4 derives analytically the worst-case bias and coverage of VARs, and shows that bias-aware VAR confidence intervals tend to be wider than the LP interval. Section 5 demonstrates the practical relevance of our results through simulations. Section 6 concludes. Proofs and supplementary results are relegated to the appendix and an online supplement.

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 Overview of results

In this section we illustrate our main results in a simple special case, namely a univariate autoregressive (AR) model subject to misspecification in the form of a small moving average (MA) process. In the subsequent sections we will show that all qualitative results—as well as many quantitative ones—from the simple model carry over to a general class of multivariate SVAR models with local MA misspecification.

MODEL AND ASSUMPTIONS. Suppose the data $\{y_t\}_{t=1}^{T}$ satisfies a univariate, stationary local-to-AR(1) model[2]

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

where $\alpha(L) = \sum_{\ell=1}^{\infty} \alpha_\ell L^\ell$ is a potentially infinite-order lag polynomial. This model captures the idea that the time series dynamics of the data are well approximated by a simple autoregressive model (here an AR(1)) driven by an unobserved white noise shock 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

---

[2] $\{y_t\}$ is a triangular array, but we suppress the dependence on T in the notation for convenience.

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 in Section 3.1 that MA misspecification of this form can capture many empirically relevant types of dynamic misspecification, such as under-specification of the lag length, missing control variables, inappropriate aggregation, and measurement error. 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 AR and LP estimators matter even asymptotically as the sample size diverges, allowing us to make tractable analytical comparisons between these two procedures.

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

i) $\varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2)$ *with* $\sigma^2 > 0$ *and* $E(\varepsilon_t^4) < \infty$.

ii) $\rho \in (-1, 1)$.

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

iv) $\zeta > 1/4$.

Absolute summability of $\alpha(L)$ is a weak regularity condition ensuring the $MA(\infty)$ 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 parameter of interest is the impulse response of y_{t+h} with respect to ε_t according to the model (2.1):

$$\theta_{h,T} \equiv E[y_{t+h} \mid \varepsilon_t = 1] - E[y_{t+h} \mid \varepsilon_t = 0] = \rho^h + T^{-\zeta} \sum_{\ell=1}^{h} \rho^{h-\ell} \alpha_\ell, \quad h = 0, 1, 2, \ldots.$$

The first term is the usual AR(1) 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 an ARMA form (with small but potentially non-zero MA terms), and we care about the full transmission mechanism of shocks in this model.

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

1. The *LP* estimator $\hat{\beta}_h$ is obtained from an OLS regression

$$y_{t+h} = \hat{\beta}_h y_t + \hat{\gamma}_h y_{t-1} + \hat{\xi}_{h,t}, \qquad (2.2)$$

   where we control for one lag of the data, and $\hat{\xi}_{h,t}$ is the OLS residual.

2. The *AR* estimator $\hat{\delta}_h$ is given by

$$\hat{\delta}_h = \hat{\rho}^h, \quad \text{where} \quad \hat{\rho} = \frac{\sum_{t=2}^T y_t y_{t-1}}{\sum_{t=2}^T y_{t-1}^2}.$$

The two estimators coincide at the impact horizon, $\hat{\beta}_0 = \hat{\delta}_0$. Conventional confidence intervals based on both these estimators would have correct asymptotic coverage in a well-specified AR(1) model. However, the presence of the 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, we shall now see that this dynamic misspecification has much more serious consequences for the AR procedure than for LP.

ROBUSTNESS OF LP. Our first main result is that the limiting distribution of the LP estimator is remarkably robust to the presence of local misspecification.

Proposition 2.1. *Under Assumption 2.1,*

$$\hat{\beta}_h - \theta_{h,T} = \frac{1}{\sigma^2} \frac{1}{T} \sum_{t=1}^T \xi_{h,t} \varepsilon_t + o_p(T^{-1/2}),$$

where $\xi_{h,t} \equiv \sum_{\ell=1}^h \rho^{h-\ell} \varepsilon_{t+\ell}$ is the h-step-ahead forecast error.

Since the above asymptotic representation is invariant to the MA misspecification, the limiting distribution of LP is the same as it would be in a correctly specified AR(1) model (provided $\zeta > 1/4$, as imposed in Assumption 2.1). Though this robustness of LP is with respect to local (i.e., asymptotically vanishing) misspecification, it is still quantitatively meaningful, as MA terms of order $T^{-\zeta}$ with $\zeta \in (1/4, 1/2)$ can be detected with probability 1 asymptotically by conventional AR model specification tests, such as the Hausman test considered in Section 3.3 below.

Why is LP robust to misspecification of such large magnitude? We will offer two mathematically equivalent pieces of intuition, with our discussion here deliberately heuristic. First, the classic omitted variable bias formula suggests that the bias of the LP impulse response estimator $\hat{\beta}_h$ in the regression (2.2) is proportional to the product of two factors: (i) the direct effect of omitted lags on y_{t+h}, and (ii) the covariance of the residualized regressor of interest $y_t - E[y_t \mid y_{t-1}] = \varepsilon_t + O_p(T^{-\zeta})$ with the omitted lags. The factor (i) is of order $T^{-\zeta}$ in the local-to-AR(1) model (2.1). The factor (ii) is also of order $T^{-\zeta}$, since ε_t is uncorrelated with any lagged data. Hence, the OVB is of order $T^{-2\zeta} = o(T^{-1/2})$ under the assumption $\zeta > 1/4$, and thus 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.

The preceding intuition is a special case of the *double robustness* property of partially linear regressions (see Example 1 in Chernozhukov, Escanciano, Ichimura, Newey, and Robins, 2022), which we will now argue applies also to LP, again settling for a heuristic argument. Consider any dynamic model (for example a VARMA(p,q)) that implies the following local projection representation:

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

Here $\theta_{0,h}$ is the true impulse response, and $\gamma_0(\cdot)$ is a function of lagged data. Define $\nu_0(y^{t-1}) \equiv E[y_t \mid y^{t-1}]$. By applying the Frisch-Waugh lemma to the regression (2.2), the LP estimator $\hat{\beta}_h$ is the sample analogue of the solution $\theta_{0,h}$ to the moment condition

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

If we evaluate the moment on the left-hand side at arbitrary functions $\gamma(\cdot)$ and $\nu(\cdot)$ rather than at the true ones $\gamma_0(\cdot)$ and $\nu_0(\cdot)$, a simple calculation shows that it equals $E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{\nu_0(y^{t-1}) - \nu(y^{t-1})\}]$.[3] Hence, the moment condition is satisfied at the true impulse response parameter $\theta_{0,h}$ as long as *either* $\gamma = \gamma_0$ or $\nu = \nu_0$, making the LP estimator *doubly robust*: it is consistent if we correctly specify either the controls $\gamma(y^{t-1})$ in the outcome equation or the controls $\nu(y^{t-1})$ in the implicit first-stage regression that isolates the shock $\varepsilon_t = y_t - \nu(y^{t-1})$. Whereas in the present univariate model the choices of $\gamma(\cdot)$ and $\nu(\cdot)$ are necessarily linked, in the general multivariate model in Section 3 they are decoupled,

---

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

and as a result it is possible to correctly specify the controls in the outcome equation while mis-identifying the shock, and *vice versa*. As a consequence of double robustness, and as argued more generally by Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) (and confirmed by our proof), it turns out that estimation error in γ_0 and ν_0 only affects the asymptotic distribution of the doubly robust estimator $\hat{\beta}_h$ through the *product* of the estimation errors $\|\hat{\gamma} - \gamma_0\| \times \|\hat{\nu} - \nu_0\|$. In our local-to-AR(1) 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 thus asymptotically negligible, consistent with our earlier intuition.

FRAGILITY OF AR. Our second main result is that this (double) robustness of LP to misspecification stands in stark contrast to the fragility of the AR estimator.

Proposition 2.2. *Under Assumption 2.1,*

$$\hat{\delta}_h - \theta_{h,T} = T^{-\zeta} \operatorname{aBias}(\hat{\delta}_h) + \frac{h\rho^{h-1}(1-\rho^2)}{\sigma^2} \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \tilde{y}_{t-1} + o_p(T^{-\zeta} + T^{-1/2}), \tag{2.3}$$

where $\tilde{y}_t = (1 - \rho L)^{-1}\varepsilon_t$ satisfies a correctly specified AR(1) model with $\alpha(L) = 0$, and the asymptotic bias is

$$\operatorname{aBias}(\hat{\delta}_h) \equiv \underbrace{h\rho^{h-1}}_{\frac{\partial(\rho^h)}{\partial\rho}} \underbrace{(1-\rho^2)\sum_{\ell=1}^{\infty}\rho^{\ell-1}\alpha_\ell}_{\operatorname{aBias}(\hat{\rho})=\frac{\operatorname{Cov}(\alpha(L)\varepsilon_t,\tilde{y}_{t-1})}{\operatorname{Var}(\tilde{y}_{t-1})}} - \underbrace{\sum_{\ell=1}^{h}\rho^{h-\ell}\alpha_\ell}_{\theta_{h,T}-\rho^h}. \tag{2.4}$$

The convergence rate of the AR estimator is $T^{-\min\{1/2,\zeta\}}$, potentially slower than the $T^{-1/2}$ rate achieved by LP. This is because the AR estimator suffers from an asymptotic bias of order $T^{-\zeta}$. This bias is due to two forces: first, the AR(1) coefficient $\hat{\rho}$ is biased due to the endogeneity caused by the MA terms, and second, the AR estimator extrapolates the horizon-h impulse response based on a parametric formula $\hat{\rho}^h$ that does not hold exactly in the true model (2.1). The second term on the right-hand side of the asymptotic representation (2.3) is the same as it would be in the correctly specified AR(1) model (with $\alpha(L) = 0$), and therefore of order $T^{-1/2}$. As a consequence, the AR 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 below in Section 3.3.

9

IMPLICATIONS FOR INFERENCE. Consider the conventional level-$(1-a)$ confidence intervals based on the LP and AR estimators, respectively:

$$\text{CI}(\hat{\beta}_h) \equiv \left[\hat{\beta}_h \pm z_{1-a/2}\sqrt{\text{aVar}(\hat{\beta}_h)/T}\right], \quad \text{CI}(\hat{\delta}_h) \equiv \left[\hat{\delta}_h \pm z_{1-a/2}\sqrt{\text{aVar}(\hat{\delta}_h)/T}\right], \quad (2.5)$$

where $z_{1-a/2} \equiv \Phi^{-1}(1-a/2)$ is the normal critical value, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and the asymptotic variances $\text{aVar}(\hat{\beta}_h)$ and $\text{aVar}(\hat{\delta}_h)$ correspond to the variances of the leading stochastic terms in Propositions 2.1 and 2.2, yielding the same expressions as in the correctly specified case.[4] While we use the asymptotic variances in the definition of the confidence intervals for mathematical simplicity, the results would be the same if we used the conventional (estimated) standard error formulas that apply under correct specification, since the latter are consistent under Assumption 2.1.

Corollary 2.1. *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* $\lim_{T\to\infty} P(\theta_{h,T} \in \text{CI}(\hat{\delta}_h)) = 0$ *for* $\zeta \in (1/4, 1/2)$, *and* $\lim_{T\to\infty} P(\theta_{h,T} \in \text{CI}(\hat{\delta}_h)) < 1 - a$ *for* $\zeta = 1/2$.

The corollary shows that the LP confidence interval robustly controls coverage. In contrast, except in the knife-edge case where the asymptotic bias equals 0, the AR confidence interval has coverage converging to 0 for misspecification of order $T^{-\zeta}$ with $\zeta \in (1/4, 1/2)$, while the asymptotic coverage is strictly below the nominal level $1 - a$ for $\zeta = 1/2$. This is because the AR interval has the same width as in the correctly specified model, but it is incorrectly centered due to the bias.

WORST-CASE BIAS AND COVERAGE. While we have shown that the AR confidence interval can severely under-cover asymptotically, it is natural to ask whether such poor performance is likely in practice. To answer this question, we set $\zeta = 1/2$, ensuring that the bias-variance trade-off is non-trivial; otherwise the bias either dominates or is negligible asymptotically. Next, we restrict the misspecification to be small, in the sense that the *noise-to-signal ratio* $\text{Var}(T^{-1/2}\alpha(L)\varepsilon_t)/\text{Var}(\varepsilon_t)$ in the local-to-AR(1) model (2.1) is bounded by M^2/T for a given M. Under this constraint, the *worst-case* VAR bias is strikingly simple.

---

[4]Under Assumption 2.1, the summands of the leading stochastic terms are serially uncorrelated, so the asymptotic variances depend on the simple variance rather than the long-run variance. Then $\text{aVar}(\hat{\beta}_h) \equiv \sigma^{-4}\text{Var}(\xi_{h,t}\varepsilon_t) = (1 - \rho^{2h})/(1 - \rho^2)$ and $\text{aVar}(\hat{\delta}_h) \equiv \sigma^{-4}[h\rho^{h-1}(1 - \rho^2)]^2\text{Var}(\varepsilon_t\tilde{y}_{t-1}) = (h\rho^{h-1})^2(1 - \rho^2)$.

Proposition 2.3. *Under Assumption 2.1, $\zeta = 1/2$, and $\mathrm{aVar}(\hat{\delta}_h) > 0$,*

$$\max_{\alpha(L):\ \mathrm{Var}(\alpha(L)\varepsilon_t)/\sigma^2 \leq M^2} \frac{|\mathrm{aBias}(\hat{\delta}_h)|}{\sqrt{\mathrm{aVar}(\hat{\delta}_h)}} = M\sqrt{\frac{\mathrm{aVar}(\hat{\beta}_h)}{\mathrm{aVar}(\hat{\delta}_h)} - 1}. \tag{2.6}$$

This formula implies that there is *no free lunch* for the AR estimator: whenever the relative standard error $\sqrt{\mathrm{aVar}(\hat{\delta}_h)/\mathrm{aVar}(\hat{\beta}_h)}$ of the AR and LP estimators is small, the AR estimator is vulnerable to misspecification in the sense of a large worst-case bias, unless we are certain *a priori* that the noise-to-signal ratio M^2/T is extremely small. It turns out that the exact same worst-case bias formula applies to multivariate locally misspecified VAR models with multiple lags, as shown below in Section 4.1. Hence, while it is possible to ameliorate the worst-case bias of AR estimators by increasing the lag length, this can *only* happen at the expense of increasing the variance.

As an implication of the above result, we show analytically in Section 4 that the AR confidence interval has severely distorted worst-case coverage when the relative standard error of AR relative to LP lies in an empirically relevant range. These severe distortions are achieved by MA misspecification processes that (i) are small, (ii) cannot be ruled out *ex ante* using economic theory, and (iii) are difficult to detect *ex post* using model specification tests. While it is possible to construct a valid *bias-aware* AR confidence interval with a larger critical value that takes into account the worst-case bias (following the general recipe of Armstrong and Kolesár, 2021), this interval tends to be wider than the equally valid conventional LP confidence interval unless the noise-to-signal ratio is restricted to be extremely small. We conclude that AR inference is only reliable when using a lag length that is so large that the procedure becomes equivalent with LP.

3 Robust local projections, fragile VARs

We now state our general model and assumptions, define the general LP and VAR impulse response estimators, derive the asymptotic representations of these estimators, and draw conclusions about their relative robustness to misspecification. While LP is shown to be robust to large amounts of misspecification, the VAR estimator is fragile in that it suffers from generically non-negligible asymptotic bias.

3.1 Model and assumptions

We consider a multivariate, stationary structural VARMA$(1, \infty)$ model that is local to an SVAR(1) model:

$$y_t = A y_{t-1} + H[I + T^{-\zeta}\alpha(L)]\varepsilon_t \quad \text{for all } t, \tag{3.1}$$

where $y_t = (y_{1,t}, \ldots, y_{n,t})'$ is n-dimensional, $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{m,t})'$ is m-dimensional, A is an $n \times n$ matrix, H is an $n \times m$ matrix, and $\alpha(L) = \sum_{\ell=1}^{\infty} \alpha_\ell L^\ell$ is an $m \times m$ lag polynomial. We allow the number of shocks m to potentially exceed the number of variables n, and *vice versa*. We will show below that equation (3.1) encompasses local-to-SVAR models with $p > 1$ lags by writing them in companion form.

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 $i^* \in \{1, \ldots, n\}$ and $j^* \in \{1, \ldots, m\}$. We define this parameter formally below.

Assumption 3.1. *For each T, $\{y_t\}_{t \in \mathbb{Z}}$ is the stationary solution to equation (3.1), given the following restrictions on parameters and shocks:*

i) *$\varepsilon_t \overset{i.i.d.}{\sim} (0_{m \times 1}, D)$, where $D \equiv \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$. For all $j = 1, \ldots, m$, $\sigma_j^2 > 0$ 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 $j^* \leq n$.*

iv) *$S \equiv \mathrm{Var}(\tilde{y}_t)$ is non-singular, where $\tilde{y}_t \equiv (I - AL)^{-1} H \varepsilon_t$ is the stationary solution to (3.1) when $\alpha(L) = 0$. Specifically, $\mathrm{vec}(S) = (I - A \otimes A)^{-1} \mathrm{vec}(\Sigma)$, 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), as this just requires ordering the instrument $y_{1,t}$ first and adding an additional measurement error term to the first equation (Stock and Watson, 2018; Plagborg-Møller and Wolf, 2021).

The impulse response of interest is now 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.

ADDITIONAL LAGS.   Our framework covers local-to-SVAR(p) models of the form

$$\check{y}_t = \sum_{\ell=1}^{p} \check{A}_\ell \check{y}_{t-\ell} + \check{H}[I + T^{-\zeta}\alpha(L)]\varepsilon_t, \tag{3.2}$$

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 3.1(iii). This fits into the original model (3.1) 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 extension will prove useful when we consider what happens as the lag length of the estimated VAR is increased.

TYPES OF MISSPECIFICATION.   The local-to-SVAR model (3.1) with MA misspecification covers several empirically relevant types of 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., Giacomini, 2013).[5] 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

---

[5]See also Section 5 for concrete examples.

failing to control for some of the variables in the true VAR system. Relatedly, MA terms may arise from 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 (e.g., 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 in magnitude, which is consistent with the spirit of our locally misspecified model (3.1).

In the most general terms, our framework can accommodate 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 $\{\varepsilon_{t+\ell}\}_{\ell \geq 0}$. The parameter of interest $\theta_{h,T}$ is 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 (in a potentially expanded vector of shocks that includes ε_t), ultimately yielding a model of the form (3.1). This argument shows that our framework allows for VAR misspecification in the form of omitted nonlinear terms or stationary time-varying parameters, as long as (i) such misspecification is small relative to the linear, time-invariant VAR(p) component; and (ii) we define the impulse response parameter as indicated above.

3.2 Estimators

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

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

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

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 $\check{y}_{t-1}, \ldots, \check{y}_{t-p}$.[6]

2. The *VAR* estimator is defined as the response of $y_{i^*,t+h}$ with respect to the j^*-th recursive innovation, where the magnitude of the innovation is normalized such that $y_{j^*,t}$ increases

---

[6]Controlling for lags in the LP is key to the double robustness property even if we directly observe the shock, if we want to allow for small contamination of the observed shock proxy.

by one unit on impact. That is, the estimator equals

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

where

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

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

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

3.3 Asymptotic coverage

As in the univariate model in Section 2, the asymptotic representation of the LP estimator is invariant to misspecification, provided that $\zeta > 1/4$.

Proposition 3.1. *Under Assumption 3.1,*

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

where

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

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

Proof. See Appendix B.1. □

The intuition for the robustness of LP is a straight-forward extension of the arguments given in the univariate case in Section 2 and so will not be repeated in detail here. LP is doubly robust in the sense that it is consistent if we correctly specify *either* the lagged controls in the outcome equation *or* the controls in the shock identification equation $\varepsilon_{j^*,t} = y_{j^*,t} - E[y_{j^*,t} \mid \underline{y}_{j^*,t}, y_{t-1}, y_{t-2}, \dots]$. Even if both parts are misspecified, the two specification errors only affect the limiting distribution of $\hat{\beta}_h$ through their product. Hence, small specification

15

errors of order $T^{-\zeta}$ impart an even smaller bias of order $T^{-2\zeta}$, which is asymptotically negligible relative to the standard deviation under our assumption $\zeta > 1/4$.

In contrast, the VAR estimator is fragile, again as in the univariate case.

Proposition 3.2. *Under Assumption 3.1,*

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

where

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

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

and $\{\tilde{y}_t\}$ and S are defined in Assumption 3.1.

Proof. See Appendix B.2. □

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. 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 $\hat{\nu}$.

To study the coverage of LP and VAR confidence intervals, we first derive the asymptotic variances of the estimators.

Corollary 3.1. *Under Assumption 3.1, the asymptotic covariance matrix for the order-$T^{-1/2}$ stochastic terms in the representations of the LP and VAR estimators in Propositions 3.1 and 3.2 is given by*

$$\mathrm{aVar}\left(\begin{pmatrix} \hat{\beta}_h \\ \hat{\delta}_h \end{pmatrix} \right) = \begin{pmatrix} \mathrm{aVar}(\hat{\beta}_h) & \mathrm{aVar}(\hat{\delta}_h) \\ \mathrm{aVar}(\hat{\delta}_h) & \mathrm{aVar}(\hat{\delta}_h) \end{pmatrix},$$

where

$$\mathrm{aVar}(\hat{\beta}_h) \equiv \frac{1}{\sigma_{j^*}^2} \left(e_{i^*,n}' A^h \overline{H}_{j^*} \overline{D}_{j^*} \overline{H}_{j^*}' (A')^h e_{i^*,n} + \sum_{\ell=1}^{h} e_{i^*,n}' A^{h-\ell} \Sigma (A')^{h-\ell} e_{i^*,n} \right),$$

$$\mathrm{aVar}(\hat{\delta}_h) \equiv \frac{1}{\sigma_{j^*}^2} e_{i^*,n}' A^h \overline{H}_{j^*} \overline{D}_{j^*} \overline{H}_{j^*}' (A')^h e_{i^*,n} + \mathrm{trace}(\Psi_h \Sigma \Psi_h' S^{-1}),$$

16

and $\overline{D}_{j^*} \equiv \mathrm{diag}(\sigma_{j^*+1}^2, \ldots, \sigma_m^2)$.

Proof. See Appendix C.1. □

For later reference, note that the corollary implies that $\hat{\beta}_h - \hat{\delta}_h$ is asymptotically independent of $\hat{\delta}_h$. This also follows from the general arguments of Hausman (1978), since Propositions 3.1 and 3.2 imply that the asymptotic covariance matrix of the LP and VAR estimators does not depend on the misspecification $\alpha(L)$, and the VAR estimator is the quasi-MLE (and thus efficient) in the correctly specified model. In particular, $\mathrm{aVar}(\hat{\delta}_h) \leq \mathrm{aVar}(\hat{\beta}_h)$.

The robustness of the LP confidence interval and fragility of the VAR confidence interval discussed in Section 2 carry over to the general model. We define the LP and VAR confidence intervals using the standard formula (2.5) in Section 2, but of course substituting the general expressions for the estimators and asymptotic variances defined in this section.

Corollary 3.2. *Under Assumption 3.1,* $\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = 1 - a$. *If, moreover,* $\mathrm{aVar}(\hat{\delta}_h) > 0$ *and* $\mathrm{aBias}(\hat{\delta}_h) \neq 0$, *then* $\underline{\lim}_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\delta}_h)) = \lim_{T\to\infty}\{1 - r\left(T^{1/2-\zeta}b_h; z_{1-a/2}\right)\}$, *where* $b_h \equiv \mathrm{aBias}(\hat{\delta}_h)/\sqrt{\mathrm{aVar}(\hat{\delta}_h)}$ *and* $r(b;c) \equiv P_{Z\sim N(0,1)}(|Z + b| > c) = \Phi(-c - b) + \Phi(-c + b)$.

Proof. See Appendix C.2. □

We again see that 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 $\zeta = 1/2$.

ROLE OF THE LAG LENGTH. A 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 horizon of interest, the more lags are required for bias reduction.

Corollary 3.3. *Suppose the model* (3.2) *written in companion form* (3.1) *satisfies Assumption 3.1. Let* $\tilde{\tilde{y}}_t$ *denote the stationary solution to equation* (3.2) *when* $\alpha(L) = 0$. *If* $\varepsilon_{j^*,t-\ell} \in \mathrm{span}(\tilde{\tilde{y}}_{t-1}, \ldots, \tilde{\tilde{y}}_{t-p})$ *for all* $\ell = 1, \ldots, h$, *then* $\mathrm{aBias}(\hat{\delta}_h) = 0$ *and* $\mathrm{aVar}(\hat{\delta}_h) = \mathrm{aVar}(\hat{\beta}_h)$. *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 C.3. □

See Plagborg-Møller and Wolf (2021) and Xu (2023) for related results in models without explicit MA misspecification.

HAUSMAN TEST.   To interpret the magnitude of the local misspecification, it is helpful to 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 $\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|/\sqrt{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)}$. A test of this kind was proposed by Stock and Watson (2018) in the context of testing for invertibility.

Proposition 3.3. *Impose Assumption 3.1 and assume* $\mathrm{aVar}(\hat{\beta}_h) > \mathrm{aVar}(\hat{\delta}_h) > 0$. *Then the asymptotic rejection probability of the Hausman test equals*

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

where b_h and $r(\cdot, \cdot)$ were defined in Corollary 3.2.

Proof. See Appendix B.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 $b_h = 0$. When $\zeta = 1/2$ and $b_h \neq 0$, the asymptotic rejection probability is strictly between the significance level a and 1. In the next section we shall use the Hausman test to quantify the difficulty of detecting especially pernicious types of model misspecification.

4   Some unpleasant VARithmetic

We now argue that our theoretical results on VAR fragility are likely to have serious implications for VAR inference in practice. Consistent with the idea that finite-order VARs can provide a good—but not perfect—approximation of reality, we impose an *a priori* constraint on the magnitude of the MA misspecification. We then show that, even if the amount of misspecification is small, the worst-case bias of the VAR estimator is small if, and only if,

its variance is similar to that of LP. When the relative standard errors of the VAR and LP estimators is in an empirically relevant range, the VAR confidence interval exhibits severe coverage distortions for misspecification that is small, hard to rule out *ex ante* on economic theory grounds, and also difficult to detect *ex post* with model validation tests. While the VAR under-coverage can be fixed by using a larger bias-aware critical value, we show that the resulting confidence interval is usually wider than the LP interval.

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

To quantify the amount of misspecification in the local-to-SVAR model (3.1) 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

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

Now 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.

Proposition 4.1. *Impose Assumption 3.1, $\zeta = 1/2$, 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. See Appendix B.4. □

19

Under our bound M^2/T on the noise-to-signal ratio, the worst-case (scaled) VAR bias is a simple function only of M and of the relative asymptotic precision $\mathrm{aVar}(\hat{\beta}_h)/\mathrm{aVar}(\hat{\delta}_h)$ 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 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 3.3), then the VAR estimator must necessarily be asymptotically equivalent with LP.

4.2   Worst-case mean squared error

To benchmark our later results on the worst-case properties of VAR inference procedures, we will first very briefly show 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 3.1, we define the asymptotic MSE of the VAR and LP estimators as follows:

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

Corollary 4.1. *Impose Assumption 3.1 and $\zeta = 1/2$. Then*

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

Proof. See Appendix C.4. □

In words, the worst-case MSE regret of VAR relative to LP is proportional to the variance reduction of VAR relative to LP, with proportionality constant $M^2 - 1$. 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.

We can also ask what is the minimax optimal way to average LP and VAR estimates.

Corollary 4.2. *Impose Assumption 3.1, $\zeta = 1/2$, and $\mathrm{aVar}(\hat{\beta}_h) > \mathrm{aVar}(\hat{\delta}_h)$. 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 $\mathrm{aMSE}(\hat{\theta}_h(\omega))$. Then*

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

Proof. See Appendix C.5. □

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

4.3 Worst-case coverage

We now turn to our main area of interest: the worst-case coverage of the conventional VAR confidence interval under small amounts of misspecification. We also characterize the least favorable type of misspecification and argue that it is difficult to rule out.

Corollary 3.2 and Proposition 4.1 immediately imply an expression for the worst-case asymptotic coverage of the conventional VAR confidence interval.

Corollary 4.3. *Impose Assumption 3.1, $\zeta = 1/2$, and $\mathrm{aVar}(\hat{\delta}_h) > 0$. Then*

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

Proof. This is an immediate consequence of Corollary 3.2 and Propositions 3.3 and 4.1. □

Figure 4.1 plots this worst-case coverage probability for different values of M, and given significance level $a = 10\%$. 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 standard error ratio commonly encountered in empirical work.

We conclude that, even for small amounts of VAR misspecification (as measured by the noise-to-signal ratio) and for relative estimator precisions of typical magnitudes, the scope for VAR undercoverage is material. This potential undercoverage 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.
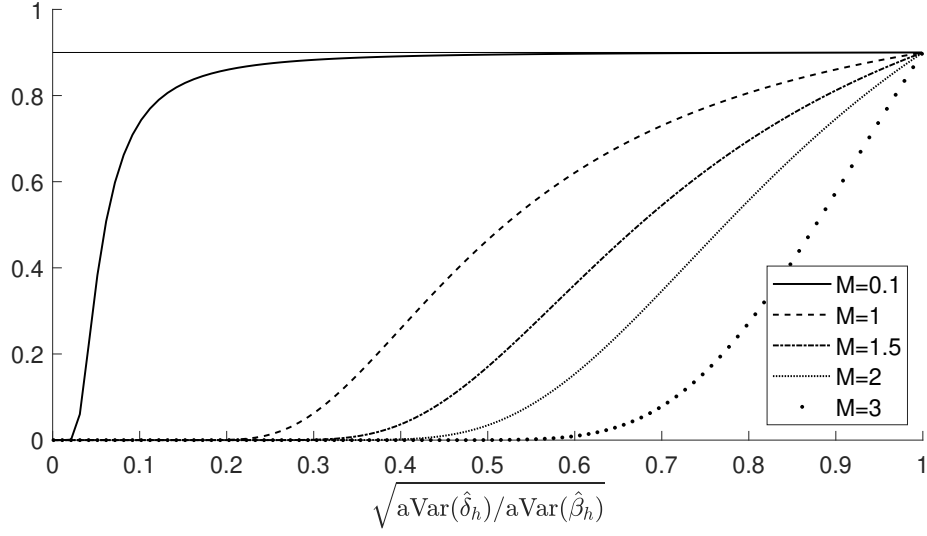
Figure 4.1: Worst-case asymptotic coverage probability of the conventional 90% VAR confidence interval. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. Different lines: different bounds M on $\|\alpha(L)\|$. The solid horizontal line marks the nominal coverage probability $1 - a = 90\%$.

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 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 aBias($\hat{\delta}_h$) is linear in $\alpha(L)$, the least favorable choice given the constraint $\|\alpha(L)\| \leq M$ follows easily from the Cauchy-Schwarz inequality (see the proof of Proposition 4.1):

$$\alpha^\dagger_{\ell,h,M} \propto D^{1/2} H' \Psi'_h S^{-1} A^{\ell-1} H D^{1/2} - \mathbb{1}(\ell \leq h) \sigma^{-1}_{j^*} D^{1/2} H' (A')^{h-\ell} e_{i^*,n} e'_{j^*,m}, \quad \ell = 1, 2, \ldots,$$

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 worst-case $\alpha^\dagger_{\ell,h,M}$. First, the magnitude of the MA coefficients 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 particularly small. Second, numerical examples shown in Appendix A.1 suggest that the MA coefficients tend to be largest in magnitude at horizon

22

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. In fact, as we will show formally in Section 5.2, simple canonical models like an ARMA(1,1) map into misspecification polynomials $\alpha(L)$ that look extremely similar to the worst-case misspecification $\alpha^{\dagger}(L)$. We thus view MA dynamics of the worst-case form as empirically and theoretically realistic types of misspecification.

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: over 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 3.1, $\zeta = 1/2$, and* $\mathrm{aVar}(\hat{\beta}_h) > \mathrm{aVar}(\hat{\delta}_h) > 0$. *Consider the joint event $\mathcal{A}_T$ that $\theta_{h,T} \notin \mathrm{CI}(\hat{\delta}_h)$ and the Hausman test in Proposition 3.3 fails to reject misspecification. Then*

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

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

Proof. See Appendix C.6. □

Figure 4.2 plots this worst-case probability for significance level $a = 10\%$, which as we have seen depends only on the ratio $\mathrm{aVar}(\hat{\delta}_h)/\mathrm{aVar}(\hat{\beta}_h)$.[7] Under correct specification, the

---

[7]It is straight-forward to numerically compute the supremum on the right-hand side of the display in Corollary 4.4, since the objective function is single-peaked, being the product of an increasing function and a decreasing function.
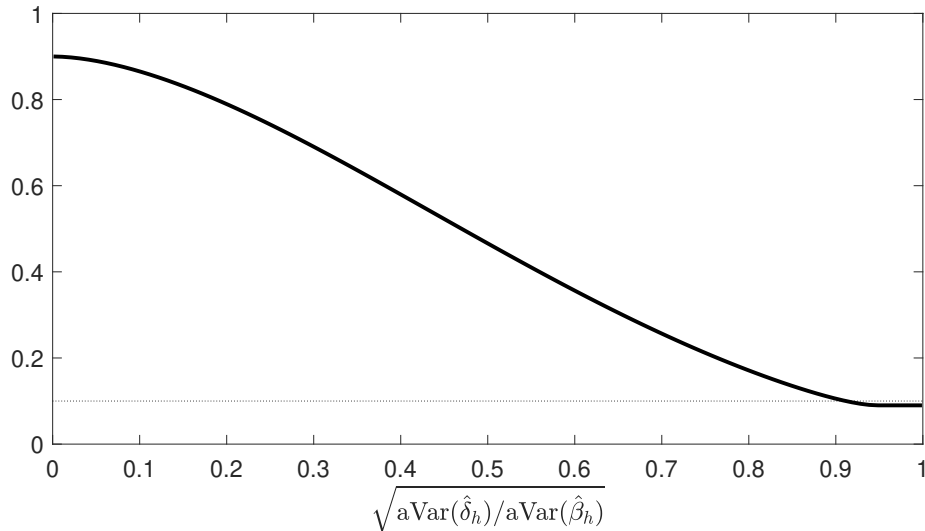
Figure 4.2: Worst-case asymptotic probability of the joint event that the conventional VAR confidence interval fails to cover the true impulse response and yet the Hausman test fails to reject misspecification. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. The dotted horizontal line marks the nominal significance level $a = 10\%$.

probability of the joint event is equal to $a(1-a)$ $(= 9\%$ when $a = 10\%)$. Yet we see that the joint probability can exceed 46% when the asymptotic standard deviation of the VAR estimator is less than half that of the LP estimator. As $\mathrm{aVar}(\hat{\delta}_h)/\mathrm{aVar}(\hat{\beta}_h) \to 0$, the worst-case joint probability approaches $1 - a$.[8] We thus again see that statistical tests fail to guard researchers adequately against the potential for severe VAR coverage distortions.

4.4 Bias-aware inference

To fix the under-coverage of the conventional VAR confidence interval, we can adjust the critical value upward to compensate for the bias, as suggested in a general setting by Armstrong and Kolesár (2021). Suppose again that we restrict the misspecification $\alpha(L)$ to satisfy $\|\alpha(L)\| \leq M$. Then we define the *bias-aware* VAR confidence interval

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

---

[8]The argument is as follows. On the one hand, the supremum of the function on the right-hand side of the display in Corollary 4.4 exceeds the function evaluated at $b = (\mathrm{aVar}(\hat{\beta}_h)/\mathrm{aVar}(\hat{\delta}_h) - 1)^{1/4}$, and it is easy to see that this function value converges to $1 - a$. On the other hand, the first factor in the function is bounded above by 1, and the second factor is bounded above by $1 - r(0; z_{1-a/2}) = 1 - a$.
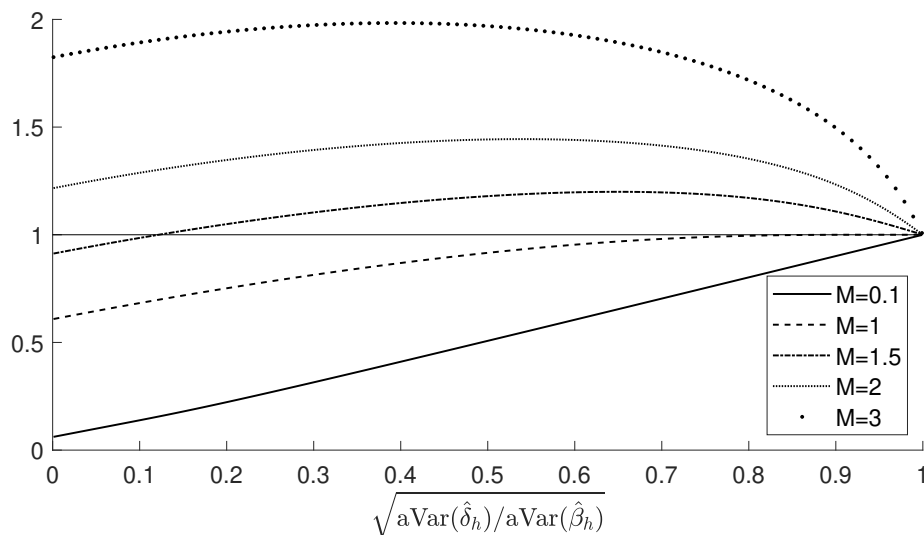
24

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.

where the bias-aware critical value $\mathrm{cv}_{1-a}(b)$ is given by the number such that $r(b; \mathrm{cv}_{1-a}(b)) = a$, and $r(\cdot, \cdot)$ is defined in Corollary 3.2. By construction, this bias-aware confidence interval has correct (but potentially conservative) asymptotic coverage.

Corollary 4.5. *Impose Assumption 3.1, $\zeta = 1/2$, and $\mathrm{aVar}(\hat{\delta}_h) > 0$. Then*

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

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

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. The ratio of the length of the bias-aware interval to that of the conventional LP interval is a function only of M and the asymptotic relative precision of LP and VAR. Figure 4.3 plots the relative interval length as a function of the relative asymptotic standard deviation, for significance level $a = 10\%$ 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 (and sometimes longer). For values of M above 2 (corresponding to a noise-to-signal ratio above $4/T$), bias-aware VAR is dominated by LP.

25

In Appendix A.2 we show that the conventional LP confidence interval is at most 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 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.

5 Simulations

In this section we show that our asymptotic results accurately reflect the finite-sample properties of LP and VAR procedures. The first set of simulations pertain to a simple ARMA(1,1) model, while the second set uses the popular structural macroeconomic model of Smets and Wouters (2007). Additional simulation results are reported in Supplemental Appendix D.

5.1 Deriving the VARMA representation

We first show how we transform any given simulation DGP into the VARMA form assumed in Sections 3 and 4. The DGPs we consider have a vector MA representation

$$y_t = C(L)u_t, \tag{5.1}$$

where $C(L) = I + \sum_{\ell=1}^{\infty} C_\ell L^\ell$, and u_t is white noise with $\Sigma_u \equiv E(u_t u_t')$. Given a choice of autoregressive lag length p, we seek to represent this process in VARMA(p, ∞) form with MA component that is as small as possible. To this end, consider a linear least-squares projection of y_t onto p of its lags. We write the projection coefficients as $A_\ell(p)$, $\ell = 1, 2, \ldots, p$, and define $A(L; p) \equiv I - \sum_{\ell=1}^{p} A_\ell(p)$. Letting $\tilde{C}(L; p) = \sum_{\ell=0}^{\infty} \tilde{C}_\ell(p)L^\ell \equiv A(L; p)C(L)$, equation (5.1) implies the VARMA(p, ∞) representation

$$y_t = \sum_{\ell=1}^{p} A_\ell(p)y_{t-\ell} + \sum_{\ell=0}^{\infty} \tilde{C}_\ell(p)u_{t-\ell}. \tag{5.2}$$

Given a choice of T as well as ζ, it is straightforward to translate the representation given in (5.2) into our baseline form (3.1).[9] We have thus arrived at a local-to-SVAR(p) model that perfectly matches the second-moment properties of the original process $\{y_t\}$ in (5.1). The MA residual in (5.2) then measures the magnitude of the departure of the DGP from the best-fitting VAR(p) model. As is the case for many structural dynamic macro models, the simulation DGPs we consider below cannot be represented exactly as finite-order VAR models. Our results will shed light on whether the resulting MA components are large enough to meaningfully distort VAR inference in practice.

5.2 Illustrative univariate model

We begin with an illustration based on a simple univariate ARMA(1,1) model. The purpose of this section is threefold: first, to get a sense of possible magnitudes for the degree of VAR misspecification M; second, to compare the actual with the worst-case lag polynomial $\alpha(L)$ in a simple and transparent environment; and third, to provide a visual illustration of our main theoretical results.

MODEL. We consider an ARMA(1,1) process:

$$y_t = \rho y_{t-1} + \varepsilon_t + \psi \varepsilon_{t-1}, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2). \tag{5.3}$$

Throughout this section we set $\rho = 0.9$, $\sigma^2 = 1$, and $\psi \in \{0, 0.25, 0.5, 0.75\}$. The lagged MA term thus accounts for up to 36 per cent of the overall variance of the error term in (5.3).

RESULTS. We first quantify the amount of misspecification. For a given lag length p as well as reference values $T = 240$ and $\zeta = 1/2$, we can proceed as in Section 5.1 to translate the model (5.3) into an ARMA(p, ∞) representation (5.2), with the AR parameters providing the best fit given p.[10] Table 5.1 shows the total amount of MA misspecification $M \equiv ||\alpha(L)||$ as a function of p and ψ. The table also reports the quantity $M^2/(1 + M^2)$, the minimax optimal *ex ante* model averaging weight on LP in Corollary 4.2. As expected, shorter lag lengths p or larger MA coefficients ψ mean greater misspecification, in the sense of larger M.

---

[9]Specifically, set $A_\ell = A_\ell(p)$, and recover H and D as the unique lower-triangular and diagonal matrices that satisfy $HDH' = \check{C}_0(p)\Sigma_u \check{C}_0(p)'$, with 1's along the main diagonal of H. Finally recover $\alpha_\ell = (HT^{-\zeta})^{-1}\tilde{C}_\ell(p)\tilde{C}_0(p)^{-1}H$ for $\ell = 1, 2, \ldots$.

[10]Note that these best-fitting AR parameters differ from the parameter ρ in (5.3) due to the serial correlation of the MA term.

p	M			$\frac{M^2}{1+M^2}$		
	$\psi = 0.25$	$\psi = 0.50$	$\psi = 0.75$	$\psi = 0.25$	$\psi = 0.50$	$\psi = 0.75$
1	3.622	7.396	11.234	0.929	0.982	0.992
2	0.882	3.337	6.821	0.437	0.918	0.979
3	0.220	1.631	4.682	0.046	0.727	0.956
4	0.055	0.811	3.361	0.003	0.397	0.919

Table 5.1: M and $\frac{M^2}{1+M^2}$ as a function of p and ψ in the ARMA(1,1) model (5.3), with the researcher estimating an AR(p).
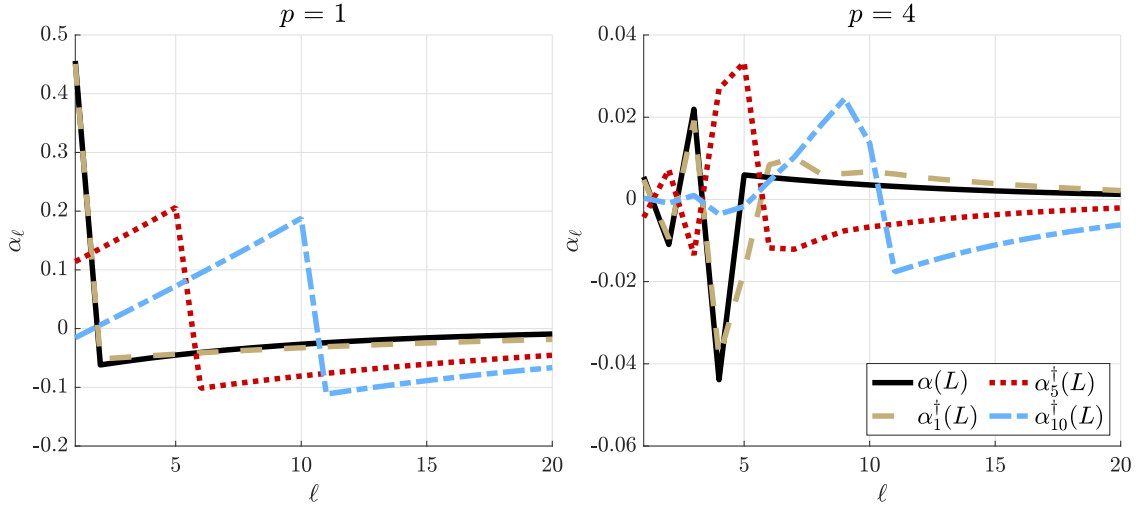
Figure 5.1: $\alpha(L)$ (black, solid) and least favorable $\alpha^\dagger(L)$ (colored, dashed and dotted) at various horizons for the simple ARMA(1,1) model (5.3), with the researcher estimating an AR(1) (left) and an AR(4) (right).

We also see that, even for moderate p and ψ, the degree of misspecification—and accordingly the optimal LP weight—can be large.

Next we show that, in this DGP, the implied misspecification polynomial $\alpha(L)$ can be very close to the theoretical least favorable polynomial $\alpha^\dagger(L; h)$ derived in Section 4.1, which represents the worst case for AR bias and coverage. Figure 5.1 shows $\alpha(L)$ (solid black) as well as $\alpha^\dagger(L; h)$ (colored, dashed and dotted) for horizons $h \in \{1, 5, 10\}$, and throughout setting $\psi = 0.5$. For both $p = 1$ as well as $p = 4$, the actual MA polynomial $\alpha(L)$ implied by the ARMA(1,1) model (5.3) is very close—though not quite identical—to the worst-case $\alpha^\dagger(L; 1)$ at horizon $h = 1$. Hence, the least favorable lag polynomial is not some practically immaterial theoretical curiosity in the particular DGP considered here.

Finally, and consistent with our theory, we find that coverage can be poor for VAR confidence intervals, while LP intervals are robust to the presence of the MA term. Figure 5.2

reports coverage rates and median confidence interval lengths for the cases $\psi = 0$ (no misspecification) and $\psi = 0.25$ (moderate misspecification, with the lagged MA term accounting for around 6% of the variance of the error term). We throughout consider $p = 1$, simulate 5,000 samples of size $T = 240$, and then for each construct delta method as well as bootstrap LP and AR confidence intervals (assuming homoskedasticity), in blue and red, respectively.[11] The top panel reveals that, when the AR(1) model is in fact correctly specified (i.e., for $\psi = 0$), then both LP and AR confidence intervals attain the nominal coverage probability of 90 per cent (left panel); furthermore, and also as expected, the AR confidence intervals are meaningfully shorter (right panel). In the misspecified case in the bottom panel, the AR confidence intervals instead substantially undercover, and particularly so at short horizons. This is consistent with our general theoretical results as well as our observation above that the actual misspecification lag polynomial $\alpha(L)$ is close to the horizon-1 worst-case one, $\alpha^{\dagger}(L; 1)$. We also see that LP, on the other hand, exhibits at worst mild undercoverage.

5.3   Misspecification in Smets and Wouters (2007)

For our quantitative simulation exercise, we take as the DGP 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 applied macroeconometric practice: the model is rich enough to match 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, like monetary or cost-push shocks.

FRAMEWORK.   We consider the model of Smets and Wouters, solved at its (quarterly) posterior mode parameterization. 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. Our specification follows recent work that studies the role of labor markets in the post-2021 inflation spike (e.g., Bernanke and Blanchard, 2023). Results for alternative set-ups with other structural shocks and different shock identification schemes are reported in Supplemental Appendix D.

RESULTS.   We begin by quantifying the amount of misspecification, having represented the model in VARMA form (see Section 5.1) for $T = 240$ and $\zeta = 1/2$. Table 5.2 shows the total

---

[11]We construct percentile-t bootstrap confidence intervals, with 5,000 bootstrap draws.

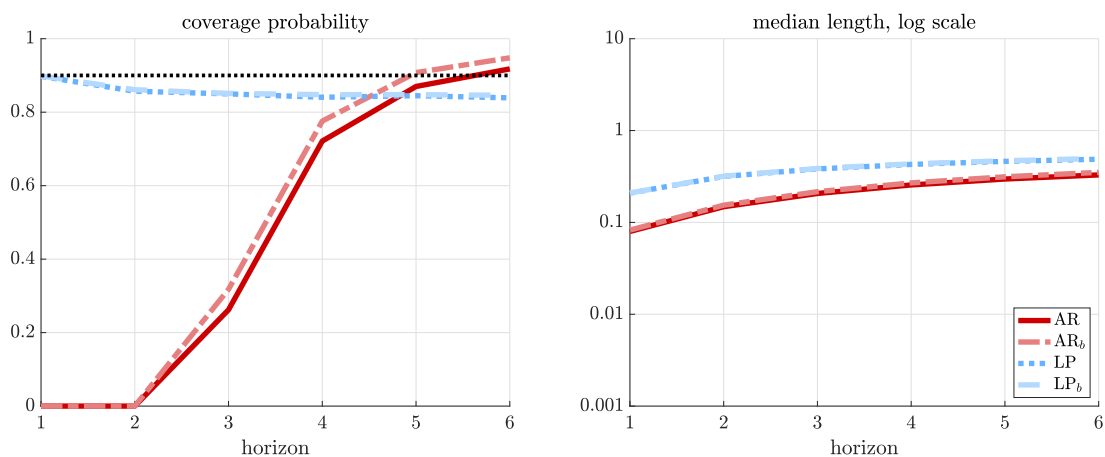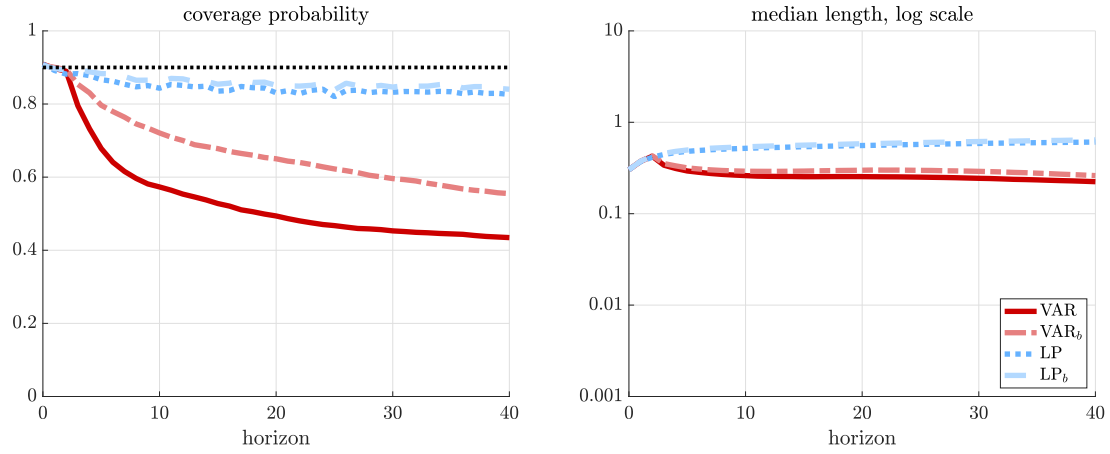## AR(1) − CORRECT SPECIFICATION
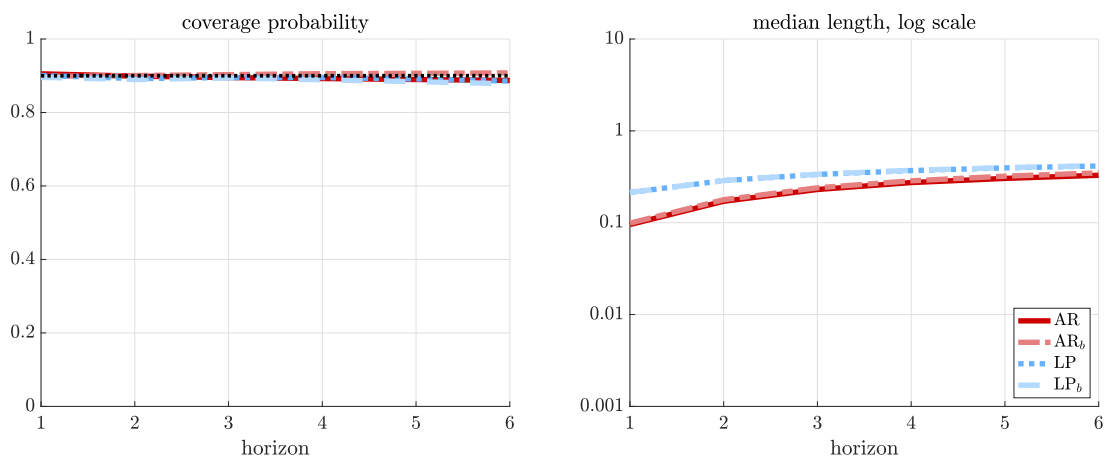
ARMA(1,1), $\psi = 0.25$

Figure 5.2: Coverage probabilities (left) and median confidence interval length (right) for AR (red) and LP (blue) for delta method and bootstrap (solid, dashed, and dotted). The DGP is (5.3) with $\psi = 0$ (top) and $\psi = 0.25$ (bottom), with estimation lag length fixed at $p = 1$.
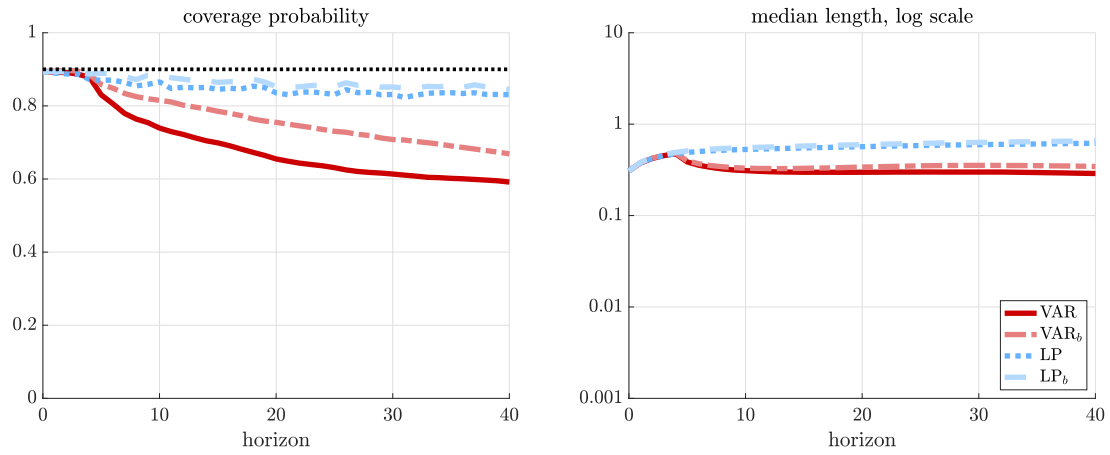
degree of misspecification M as well as the minimax MSE-optimal weight $M^2/(1 + M^2)$ on LP in Corollary 4.2 as a function of the VAR lag length p. As anticipated, the larger p, the smaller M. Importantly, however, M only declines extremely slowly with the lag length p. In particular, for lag lengths typical in applied practice (for quarterly data), misspecification is material, 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 materially affect VAR inference.

Figure 5.3 shows that VAR confidence intervals indeed severely undercover, while LP

coverage probability

median length, log scale

LAG LENGTH $p = 4$

coverage probability

median length, log scale

LAG LENGTH $p = 8$

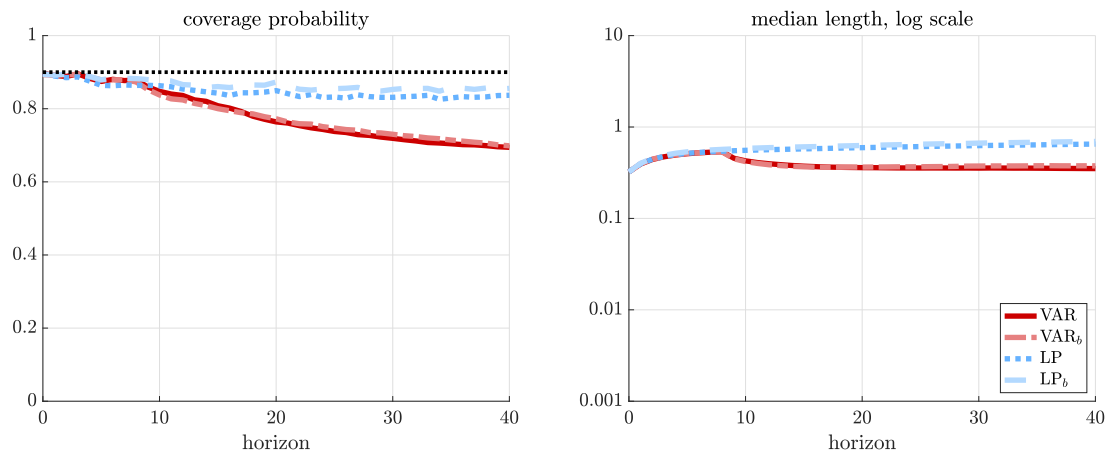coverage probability

median length, log scale

Figure 5.3: See Figure 5.2. The DGP is the model of Smets and Wouters, and the researcher estimates the response of inflation to an observed cost-push shock. Lag length p is selected using the AIC for the top panel, and set to $p = 4$ and $p = 8$ for the middle and bottom panels, respectively.

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 5.2: M and $\frac{M^2}{1+M^2}$ as a function of p in the structural model of Smets and Wouters (2007), with the researcher observing the cost-push shock, inflation, wages, and hours worked, and estimating a VAR(p).

intervals remain 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). The top panel displays our headline results, with lag length p selected using the AIC. 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. 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 material VAR bias and undercoverage. The middle and bottom panels then further illustrate our "no free lunch" results. 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 3.3), but at the same time confidence intervals become essentially as wide as for LP. At longer horizons we obtain the same picture as before: material undercoverage for VAR, and coverage close to the nominal level for LP.[12]

Supplemental Appendix D shows that these conclusions extend to alternative shocks and alternative shock identification schemes. We there furthermore investigate what happens if the actual lag polynomial $\alpha(L)$ is replaced by the theoretical least favorable one. In that case, the magnitudes of VAR undercoverage are broadly comparable with those obtained under the actual $\alpha(L)$ implied by the Smets and Wouters (2007) model, revealing that the least favorable MA polynomial $\alpha^{\dagger}(L, \cdot)$ is not particularly pathological, echoing our earlier finding in the univariate ARMA(1,1) simulation DGP.

---

[12]The VAR coverage distortion is not a small-sample phenomenon: Supplemental Appendix D shows that similar coverage distortions arise with a larger sample size.

SUMMARY. The simulation results in this section suggest that our theoretical results are likely to have practical bite. In empirically realistic settings, LP confidence intervals have close-to-nominal coverage, while VAR misspecification is plausibly large enough (and similar enough to the worst-case misspecification) to cause severe VAR coverage distortions. A researcher may guard against such distortions by increasing the lag length beyond the values suggested by standard information criteria; however, in doing so, she will end up with confidence intervals of length comparable to those of LPs.

6 Conclusion

We have shown that, while conventional LP inference is robust to surprisingly large amounts of misspecification, VAR inference is much more fragile. Conventional VAR confidence intervals suffer from severe coverage distortions even for misspecification that is small, difficult to rule out *ex ante* using economic theory, and difficult to detect *ex post* with VAR model specification tests. These analytical results demonstrate that—when the goal is to construct valid confidence intervals for impulse responses, as opposed to minimizing MSE—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.

A practical take-away for LP analysis is that researchers should control for lags of the data that are strong predictors of the outcome and 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, and conventional information criteria (such as AIC) suffice to select an appropriate lag length for LP in our framework. 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.

Is there a way forward for VAR inference? The conventional VAR confidence interval is valid if the estimation lag length is sufficiently large, but our results show that this can only happen by making the interval equivalent with the LP interval. Using a smaller lag length combined with a bias-aware critical value typically leads 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.[13] 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 realistic. Rather than restricting the parameter space, an alternative would be to weaken the coverage requirement, e.g., only requiring a certain coverage probability *on average* over a set of horizons (Armstrong, Kolesár, and Plagborg-Møller, 2022), or by changing the target for inference from the true impulse response function to a smooth projection of this function (Genovese and Wasserman, 2008). Finally, a subjectivist Bayesian VAR modeler need only worry about our negative results if their prior on potential misspecification attaches significant weight to MA processes that imply large VAR biases.

---

[13]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.

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 4.3. We focus here on the univariate local-to-AR(1) model from Section 2, 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 the lag ℓ. The figure shows that the shape of the coefficients either takes the form of a hump or 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.
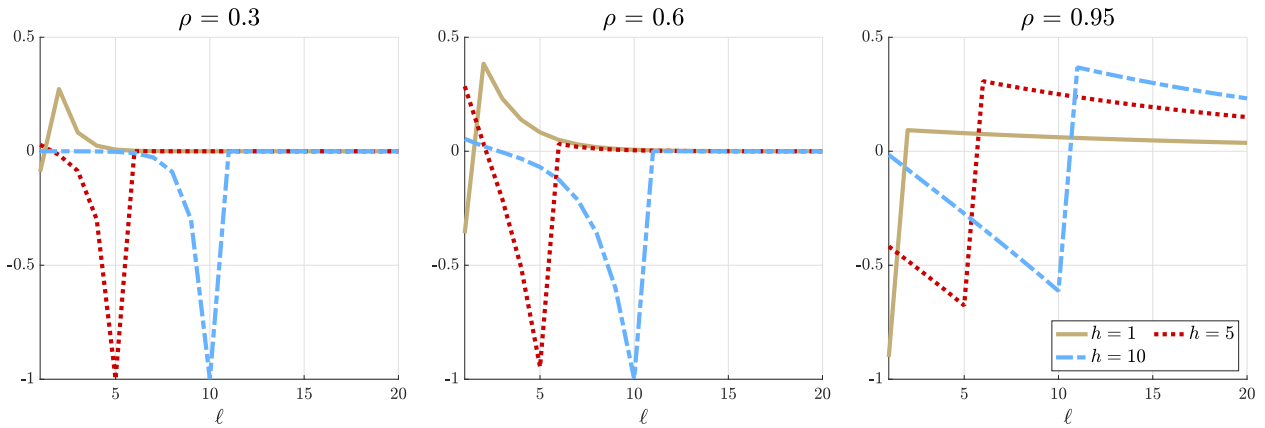

Figure A.1: Least favorable $\alpha^\dagger(L; h)$ for horizons $h \in \{1, 5, 10\}$ for local-to-AR(1) models with different persistence parameters ρ (left, middle, and right panel).

A.2   More efficient bias-aware confidence interval

Generalizing the bias-aware VAR confidence interval in Section 4.4, 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:

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

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

Corollary A.1. *Impose Assumption 3.1, $\zeta = 1/2$, 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 3.1 and the same calculations as in the proof of Corollary 4.2. $\qquad\square$

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 averaging weight $\omega = \omega^*$ is given by

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

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 $M \geq 2$, the length-optimal ω^* is numerically close to the MSE-optimal weight $M^2/(1+M^2)$ derived in Corollary 4.2.
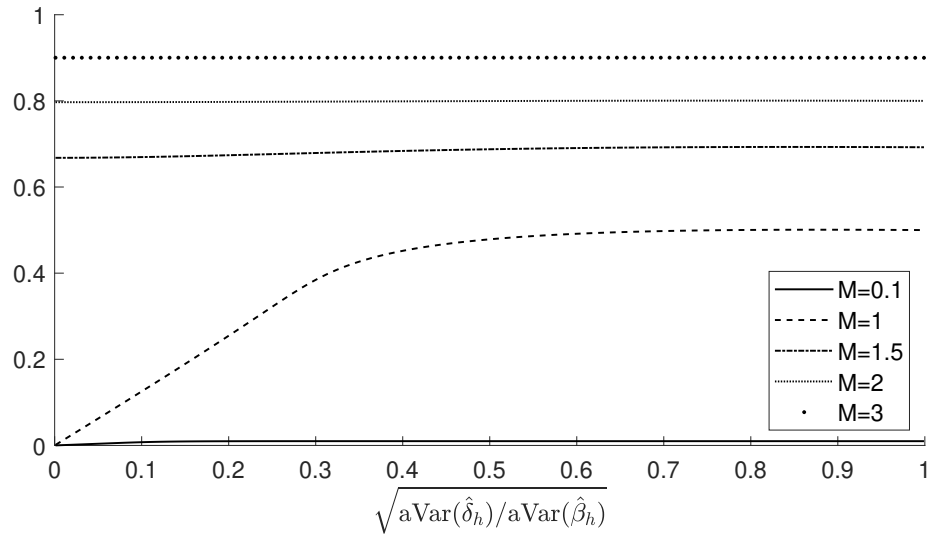
Figure A.2: Length-optimal weight on LP in bias-aware confidence interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. Different lines: different bounds M on $\|\alpha(L)\|$.
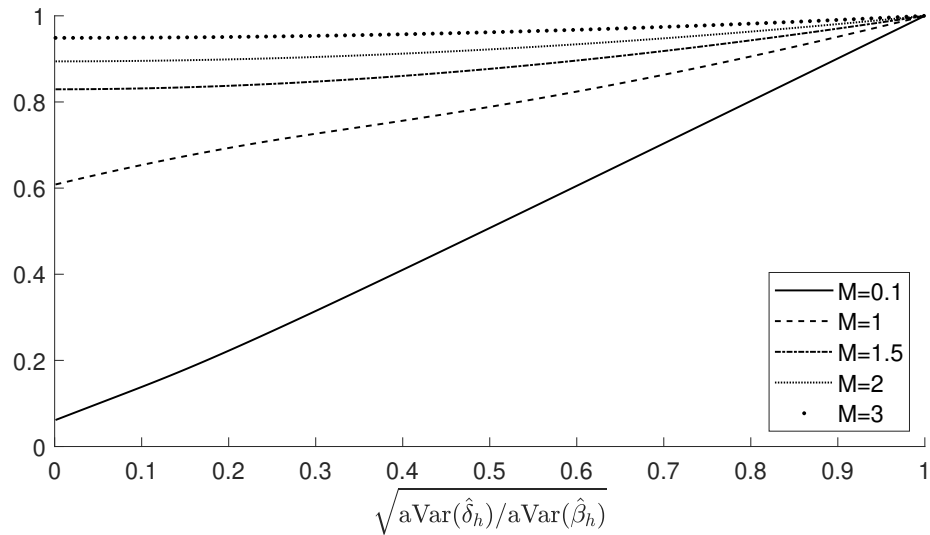

Figure A.3: Relative length of optimal bias-aware confidence interval vs. conventional LP interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of LP vs. VAR. Different lines: different bounds M on $\|\alpha(L)\|$.

Appendix B   Proofs of propositions

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,

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

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

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

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

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

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

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

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

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

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

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

where S was defined in Assumption 3.1. Lemma E.3 shows that

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

39

Define

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

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

B.3   Proof of Proposition 3.3

By Propositions 3.1 and 3.2 and Corollary 3.1,

$$
\hat{\beta}_h - \hat{\delta}_h = -T^{-\zeta}\,\mathrm{aBias}(\hat{\delta}_h) + R_T + o_p(T^{-\zeta} + T^{-1/2}),
$$

where

$$
\sqrt{T}R_T \xrightarrow{d} N(0, \mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)).
$$

Consider the following three cases:

1. $\zeta \in (1/4, 1/2)$: In this case

$$
\begin{aligned}
\lim_{T\to\infty} P\left(\frac{\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|}{\sqrt{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)}} > z_{1-a/2} \right) &= \lim_{T\to\infty} P\left(\frac{T^\zeta|\hat{\beta}_h - \hat{\delta}_h|}{\sqrt{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)}} > \frac{z_{1-a/2}}{T^{1/2-\zeta}} \right) \\
&= P\left(\frac{|\mathrm{aBias}(\hat{\delta}_h)|}{\sqrt{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)}} > 0 \right) \\
&= 1.
\end{aligned}
$$

    where the last inequality uses the fact that $\mathrm{aBias}(\hat{\delta}_h) \neq 0$.

2. $\zeta > 1/2$: In this case

$$
\sqrt{T}(\hat{\beta}_h - \hat{\delta}_h) \xrightarrow{d} N(0, \mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)),
$$

    and thus

$$
\lim_{T\to\infty} P\left(\frac{\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|}{\sqrt{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)}} > z_{1-a/2} \right) = a.
$$

3. $\zeta = 1/2$: In this case

$$
\sqrt{T}(\hat{\beta}_h - \hat{\delta}_h) \xrightarrow{d} N(\mathrm{aBias}(\hat{\delta}_h), \mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)).
$$

Thus,

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

where $b_h \equiv \mathrm{aBias}(\hat{\delta}_h)/\sqrt{\mathrm{aVar}(\hat{\delta}_h)}$ and $r(b;c) \equiv P_{Z \sim N(0,1)}(|Z + b| > c) = \Phi(-c - b) + \Phi(-c + b)$, as in Corollary 3.2.

In summary, the three cases above show that

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

B.4   Proof of Proposition 4.1

Rewrite the numerator of b_h as

$$\mathrm{trace}\left\{S^{-1}\Psi_h H \sum_{\ell=1}^{\infty} \alpha_\ell D H'(A')^{\ell-1}\right\} - e'_{i^*,n} \sum_{\ell=1}^{h} A^{h-\ell} H \alpha_\ell e_{j^*,m} = \sum_{\ell=1}^{\infty} \mathrm{trace}\left\{\Xi_{h,\ell} D^{-1/2} \alpha_\ell D^{1/2}\right\},$$

with

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

By Cauchy-Schwarz, we can bound the squared value of the numerator of b_h as follows:

$$\left(\sum_{\ell=1}^{\infty} \mathrm{trace}\left\{\Xi_{h,\ell} D^{-1/2} \alpha_\ell D^{1/2}\right\}\right)^2 \le \left(\sum_{\ell=1}^{\infty} \mathrm{trace}(\Xi'_{h,\ell} \Xi_{h,\ell})\right)$$
$$\times \underbrace{\left(\sum_{\ell=1}^{\infty} \mathrm{trace}\{(D^{-1/2} \alpha_\ell D^{1/2})' D^{-1/2} \alpha_\ell D^{1/2}\}\right)}_{=\|\alpha(L)\|^2}.$$

Moreover, the bound is achieved by choosing $\alpha_\ell \propto \Xi'_{h,\ell}$ for $\ell = 1, 2, \ldots$. Hence, the statement of the proposition follows if we can show that

$$\sum_{\ell=1}^{\infty} \mathrm{trace}(\Xi'_{h,\ell} \Xi_{h,\ell}) = \mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h).$$

Indeed, by multiplying out terms and using $\Sigma = HDH'$, we obtain

$$\sum_{\ell=1}^{\infty} \text{trace}(\Xi'_{h,\ell}\Xi_{h,\ell}) = \sum_{\ell=1}^{\infty} \text{trace}(A^{\ell-1}\Sigma(A')^{\ell-1}S^{-1}\Psi_h\Sigma\Psi'_h S^{-1})$$

$$- 2\sum_{\ell=1}^{h} \text{trace}(A^{\ell-1}H_{\bullet,j^*}e'_{i^*,n}A^{h-\ell}\Sigma\Psi'_h S^{-1})$$

$$+ \sum_{\ell=1}^{h}\sigma_{j^*}^{-2}e'_{i^*,n}A^{h-\ell}\Sigma(A')^{h-\ell}e_{i^*,n}.$$

The first term on the right-hand side above equals

$$\text{trace}\left\{\underbrace{\left(\sum_{\ell=1}^{\infty}A^{\ell-1}\Sigma(A')^{\ell-1}\right)}_{=\text{Var}(\tilde{y}_t)=S}S^{-1}\Psi_h\Sigma\Psi'_h S^{-1}\right\} = \text{trace}\{\Psi_h\Sigma\Psi'_h S^{-1}\}.$$

The second term on the right-hand side two displays ago equals

$$-2\,\text{trace}\left\{\underbrace{\left(\sum_{\ell=1}^{h}A^{\ell-1}H_{\bullet,j^*}e'_{i^*,n}A^{h-\ell}\right)}_{=\sum_{\ell=1}^{h}A^{h-\ell}H_{\bullet,j^*}e'_{i^*,n}A^{\ell-1}=\Psi_h}\Sigma\Psi'_h S^{-1}\right\} = -2\,\text{trace}\{\Psi_h\Sigma\Psi'_h S^{-1}\}.$$

We have shown

$$\sum_{\ell=1}^{\infty}\text{trace}(\Xi'_{h,\ell}\Xi_{h,\ell}) = \frac{1}{\sigma_{j^*}^2}\sum_{\ell=1}^{h}e'_{i^*,n}A^{h-\ell}\Sigma(A')^{h-\ell}e_{i^*,n} - \text{trace}\{\Psi_h\Sigma\Psi'_h S^{-1}\},$$

which in turn equals $\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)$ (cf. Corollary 3.1), as required. $\square$

Appendix C   Proofs of corollaries

C.1   Proof of Corollary 3.1

By Proposition 3.1 and the that fact $\xi_{h,i^*,t}\varepsilon_{j^*,t}$ is serially uncorrelated, the asymptotic variance of LP equals $\sigma_{j^*}^{-4}E[\xi_{h,i^*,t}^2\varepsilon_{j^*,t}^2] = \sigma_{j^*}^{-2}E[\xi_{h,i^*,t}^2]$, where $\xi_{h,i^*,t}$ is the i^*-th element of the $n \times 1$ vector

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

with $\overline{H}_{j*} \equiv (H_{\bullet, j*+1}, \ldots, H_{\bullet, m})$ and $\overline{\varepsilon}_{j*,t} \equiv (\varepsilon_{j*+1,t}, \ldots, \varepsilon_{m,t})'$. A quick calculation verifies the expression stated in the corollary.

Using the limiting representation for the VAR estimator in Proposition 3.2, notice first that $T^{-1} \sum_{t=1}^{T} \xi_{0,t} \varepsilon_{j*,t}$ and $T^{-1} \sum_{t=1}^{T} \varepsilon_t \tilde{y}'_{t-1}$ are asymptotically independent, since for any $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m\}$, and $\ell \geq 0$,

$$E[\xi_{0,i,t} \varepsilon_{j*,t} \tilde{y}_{t-\ell-1} \varepsilon_{j,t-\ell}] = E[\xi_{0,i,t}] E[\varepsilon_{j*,t} \varepsilon_{j,t-\ell}] E[\tilde{y}_{t-\ell-1}] = 0.$$

The asymptotic variance of the term

$$\frac{1}{\sigma_{j*}^2} e'_{i*,n} A^h T^{-1} \sum_{t=1}^{T} \xi_{0,t} \varepsilon_{j*,t},$$

which has serially uncorrelated summands, is given by

$$\frac{1}{\sigma_{j*}^4} e'_{i*,n} A^h E[\xi_{0,t} \xi'_{0,t}] E[\varepsilon_{j*,t}^2] A^h e_{i*,n} = \frac{1}{\sigma_{j*}^2} e'_{i*,n} A^h \overline{H}_{j*} \overline{D}_{j*} \overline{H}'_{j*} (A')^h e_{i*,n}.$$

The asymptotic variance of the term

$$\text{trace} \left\{ S^{-1} \Psi_h H T^{-1/2} \sum_{t=1}^{T} \varepsilon_t \tilde{y}'_{t-1} \right\} = T^{-1/2} \sum_{t=1}^{T} \tilde{y}'_{t-1} S^{-1} \Psi_h H \varepsilon_t,$$

which also has serially uncorrelated summands, equals

$$\begin{aligned}
\text{Var}\left(\tilde{y}'_{t-1} S^{-1} \Psi_h H \varepsilon_t \right) &= E[\tilde{y}'_{t-1} S^{-1} \Psi_h H \varepsilon_t \varepsilon'_t H' \Psi'_h S^{-1} \tilde{y}_{t-1}] \\
&= \text{trace}\left\{ E[S^{-1} \Psi_h H \varepsilon_t \varepsilon'_t H' \Psi'_h S^{-1} \tilde{y}_{t-1} \tilde{y}'_{t-1}] \right\} \\
&= \text{trace}\left\{ S^{-1} \Psi_h H E[\varepsilon_t \varepsilon'_t] H' \Psi'_h S^{-1} E[\tilde{y}_{t-1} \tilde{y}'_{t-1}] \right\} \\
&= \text{trace}\left\{ S^{-1} \Psi_h H D H' \Psi'_h S^{-1} S \right\} \\
&= \text{trace}\left\{ \Psi_h \Sigma \Psi'_h S^{-1} \right\}.
\end{aligned}$$

The expression for the asymptotic variance of VAR follows.

Finally, for the formula of the asymptotic covariance, note that by Propositions 3.1 and 3.2,

$$\begin{aligned}
\hat{\beta}_h - \theta_{h,T} &= R_{1,T} + R_{3,T} + o_p(T^{-1/2}), \\
\hat{\delta}_h - \theta_{h,T} &= T^{-\zeta} \, \text{aBias}(\hat{\delta}_h) + R_{2,T} + R_{3,T} + o_p(T^{-\zeta} + T^{-1/2}),
\end{aligned}$$

where

$$R_{1,T} \equiv \frac{1}{\sigma_{j*}^2} T^{-1} \sum_{t=1}^{T} \left(\sum_{\ell=1}^{h} e_{i*,n}' A^{h-\ell} H \varepsilon_{t+\ell} \right) \varepsilon_{j*,t},$$

$$R_{2,T} \equiv \text{trace} \left\{ S^{-1} \Psi_h H T^{-1} \sum_{t=1}^{T} \varepsilon_t \tilde{y}_{t-1}' \right\},$$

$$R_{3,T} \equiv \frac{1}{\sigma_{j*}^2} e_{i*,n}' A^h T^{-1} \sum_{t=1}^{T} \left(\overline{H}_{j*} \overline{\varepsilon}_{j*,t} \right) \varepsilon_{j*,t}.$$

The asymptotic covariances for the order $T^{-1/2}$ terms in the representation above is

$$\text{aCov}(R_{1,T} + R_{3,T}, R_{2,T} + R_{3,T}).$$

In the calculation of $\text{aVar}(\hat{\delta}_h)$, we have already shown that $\text{aCov}(R_{3,T}, R_{2,T}) = 0$. Analogous calculations also show that $\text{aCov}(R_{3,T}, R_{1,T}) = 0$. Thus,

$$\text{aCov}(R_{1,T} + R_{3,T}, R_{2,T} + R_{3,T}) = \text{aCov}(R_{1,T}, R_{2,T}) + \text{aVar}(R_{3,T}).$$

We next show that $\text{aCov}(R_{1,T}, R_{2,T}) = \text{aVar}(R_{2,T})$. Note that

$$\text{aCov}(R_{1,T}, R_{2,T}) = \frac{1}{\sigma_{j*}^2} \sum_{s=-\infty}^{\infty} \text{Cov} \left(\varepsilon_{j*,t} \sum_{\ell=1}^{h} e_{i*,n}' A^{h-\ell} H \varepsilon_{t+\ell}, \text{trace} \left\{ S^{-1} \Psi_h H \varepsilon_{t+s} \tilde{y}_{t+s-1}' \right\} \right)$$

$$= \frac{1}{\sigma_{j*}^2} \sum_{\ell=1}^{h} E \left[e_{i*,n}' A^{h-\ell} H \varepsilon_{t+\ell} \varepsilon_{t+\ell}' H' \Psi_h' S^{-1} (\varepsilon_{j*,t} \tilde{y}_{t+\ell-1}) \right]$$

$$= \frac{1}{\sigma_{j*}^2} \sum_{\ell=1}^{h} e_{i*,n}' A^{h-\ell} H E \left[\varepsilon_{t+\ell} \varepsilon_{t+\ell} \right]' H' \Psi_h' S^{-1} E \left[\varepsilon_{j*,t} \tilde{y}_{t+\ell-1} \right]$$

$$= \frac{1}{\sigma_{j*}^2} \text{trace} \left\{ \left(\sum_{\ell=1}^{h} E \left[\varepsilon_{j*,t} \tilde{y}_{t+\ell-1} \right] e_{i*,n}' A^{h-\ell} \right) H D H' \Psi_h' S^{-1} \right\}$$

$$= \frac{1}{\sigma_{j*}^2} \text{trace} \left\{ \left(\sum_{\ell=1}^{h} A^{\ell-1} H_{\bullet,j*} \sigma_{j*}^2 e_{i*,n}' A^{h-\ell} \right) \Sigma \Psi_h' S^{-1} \right\}$$

$$= \text{trace} \left\{ \left(\sum_{\ell=1}^{h} A^{\ell-1} H_{\bullet,j*} e_{i*,n}' A^{h-\ell} \right) \Sigma \Psi_h' S^{-1} \right\}.$$

Finally, reversing the order of the terms in the sum above gives

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

44

Thus, we conclude that

$$\mathrm{aCov}(R_{1,T}, R_{2,T}) = \mathrm{trace}\left\{\Psi_h \Sigma \Psi_h' S^{-1}\right\} = \mathrm{aVar}(R_{2,T}).$$

where the final equality follows from the derivation of $\mathrm{aVar}(\hat{\delta}_h)$. This implies that

$$\mathrm{aCov}(R_{1,T} + R_{3,T}, R_{2,T} + R_{3,T}) = \mathrm{aVar}(R_{2,T}) + \mathrm{aVar}(R_{3,T}) = \mathrm{aVar}(\hat{\delta}_h). \quad \square$$

C.2   Proof of Corollary 3.2

The result is an immediate consequence of Propositions 3.1 and 3.2. For the sake of exposition, we present details below. Recall the definitions of the VAR and LP confidence intervals in Corollary 3.2, and let the asymptotic variances $\mathrm{aVar}(\hat{\beta}_h)$ and $\mathrm{aVar}(\hat{\delta}_h)$ correspond to the formulae in Corollary 3.1.

First, we show that $\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = 1 - a$. Note that

$$P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = P\left(-z_{1-a/2} \le \frac{\sqrt{T}(\hat{\beta}_h - \theta_{h,T})}{\sqrt{\mathrm{aVar}(\hat{\beta}_h)}} \le z_{1-a/2}\right).$$

By Proposition 3.1 and Corollary 3.1,

$$\sqrt{T}(\hat{\beta}_h - \theta_{h,T}) \xrightarrow{d} N(0, \mathrm{aVar}(\hat{\beta}_h)).$$

We thus conclude that $\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = 1 - a$.

Second, we characterize $\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\delta}_h))$. Note that

$$P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = P\left(-\frac{z_{1-a/2}}{\sqrt{T}} \le \frac{(\hat{\delta}_h - \theta_{h,T})}{\sqrt{\mathrm{aVar}(\hat{\delta}_h)}} \le \frac{z_{1-a/2}}{\sqrt{T}}\right).$$

By Proposition 3.2 and Corollary 3.1,

$$\hat{\delta}_h - \theta_{h,T} = T^{-\zeta} \mathrm{aBias}(\hat{\delta}_h) + \tilde{R}_T + o_p(T^{-\zeta} + T^{-1/2}),$$

where

$$\sqrt{T}\tilde{R}_T \xrightarrow{d} N(0, \mathrm{aVar}(\hat{\delta}_h)).$$

Consider the following three cases.

1. $\zeta \in (1/4, 1/2)$: In this case

$$\lim_{T \to \infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = \lim_{T \to \infty} P\left(-\frac{z_{1-a/2}}{T^{1/2-\zeta}} \leq \frac{T^{\zeta}(\hat{\delta}_h - \theta_{h,T})}{\sqrt{\mathrm{aVar}(\hat{\delta}_h)}} \leq \frac{z_{1-a/2}}{T^{1/2-\zeta}}\right)$$

$$= P\left(0 \leq \frac{\mathrm{aBias}(\hat{\delta}_h)}{\sqrt{\mathrm{aVar}(\hat{\delta}_h)}} \leq 0\right) = 0,$$

where the last equality uses the fact that $\mathrm{aBias}(\hat{\delta}_h) \neq 0$.

2. $\zeta > 1/2$: In this case

$$\sqrt{T}(\hat{\delta}_h - \theta_{h,T}) \xrightarrow{d} N(0, \mathrm{aVar}(\hat{\delta}_h)),$$

and thus

$$\lim_{T \to \infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = 1 - a.$$

3. $\zeta = 1/2$: In this case

$$\sqrt{T}(\hat{\delta}_h - \theta_{h,T}) \xrightarrow{d} N(\mathrm{aBias}(\hat{\delta}_h), \mathrm{aVar}(\hat{\delta}_h)).$$

Thus,

$$\lim_{T \to \infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = 1 - r(b_h, z_{1-a/2}),$$

where $b_h \equiv \mathrm{aBias}(\hat{\delta}_h)/\sqrt{\mathrm{aVar}(\hat{\delta}_h)}$ and $r(b; c) \equiv P_{Z \sim N(0,1)}(|Z + b| > c) = \Phi(-c - b) + \Phi(-c + b)$.

In summary, $\lim_{T \to \infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\delta}_h)) = \lim_{T \to \infty}\{1 - r(T^{1/2-\zeta}b_h; z_{1-a/2})\}$. $\qquad \square$

C.3 Proof of Corollary 3.3

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

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

$$= \sum_{\ell=1}^{h} (A')^{h-\ell} e_{i^*,n} E^*(\varepsilon_{j^*,t-\ell} \mid \tilde{y}_{t-1})$$

$$= \sum_{\ell=1}^{h} (A')^{h-\ell} e_{i^*,n} \varepsilon_{j^*,t-\ell},$$

46

where the last equality uses the assumption $\varepsilon_{j^*,t-\ell} \in \text{span}(\tilde{y}_{t-1}, \ldots, \tilde{y}_{t-p})$ for $\ell = 1, \ldots, h$. 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 3.1 that $\text{aVar}(\hat{\beta}_h) = \text{aVar}(\hat{\delta}_h)$. Then Proposition 4.1 implies that $\text{aBias}(\hat{\delta}_h) = 0$. □

C.4 Proof of Corollary 4.1

By Proposition 4.1,

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

Thus,

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

C.5 Proof of Corollary 4.2

Write $\hat{\theta}_h(\omega) = \hat{\delta}_h + \omega(\hat{\beta}_h - \hat{\delta}_h)$. By Proposition 3.3, 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. $\qquad\square$

C.6 Proof of Corollary 4.4

Proposition 4.1 implies that the absolute relative VAR bias $|b_h|$ can be made to take any value in $[0, \infty)$ as $\alpha(L)$ varies over the set of all absolutely summable lag polynomials. The corollary then follows from Corollary 3.2 and Proposition 3.3. $\qquad\square$

References

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.

FERNÁNDEZ-VILLAVERDE, J., J. F. RUBIO-RAMÍREZ, T. J. SARGENT, AND M. W. WATSON (2007): "ABCs (and Ds) of understanding VARs," *American Economic Review*, 97(3), 1021–1026.

GENOVESE, C., AND L. WASSERMAN (2008): "Adaptive confidence bands," *The Annals of Statistics*, 36(2), 875 – 905.

GIACOMINI, R. (2013): "The Relationship Between DSGE and VAR Models," in *VAR Models in Macroeconomics — New Developments and Applications: Essays in Honor of Christopher A. Sims*, ed. by T. B. Fomby, L. Kilian, and A. Murphy, vol. 32 of *Advances in Econometrics*, pp. 1–25. Emerald Group Publishing Limited.

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., AND B. K. JOHANNSEN (2023): "Bias in Local Projections," Manuscript, Board of Governors of the Federal Reserve.

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

JORDÀ, O. (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.

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.