Universitat Pompeu Fabra Barcelona
Department of Economics and Business

Economics Working Paper Series

Working Paper No. 1953

Quantifying dynamic misspecification biases in VARs

Oriol González-Casasús

September 2026

Quantifying Dynamic Misspecification Biases in VARs

Oriol González-Casasús*

Universitat Pompeu Fabra and Barcelona School of Economics

September 25, 2026

Abstract

Applied macroeconomists favor parsimonious VARs, which selection criteria and specification pre-tests routinely endorse. I show that weak, long-lived omitted dynamics escape those diagnostics and cannot be consistently estimated, yet the bias they induce in impulse responses and forecasts admits valid confidence intervals. Researchers can then keep the parsimonious VAR and state the direction and magnitude of the distortion it carries. In US data, the bias in the price response to a monetary contraction is significantly positive at several horizons, accounting for part of the price puzzle without expanding the cross-sectional information set.

JEL Classification: C12, C32, C52, E52.

Key words: Robust Inference, Vector Autoregressions, Dynamic Misspecification, Price Puzzle.

*Department of Economics and Business, Universitat Pompeu Fabra. Email: oriol.gonzalez@upf.edu. I thank Mikkel Plagborg-Møller, Frank Schorfheide and participants at the EIEF Conference on Macroeconometrics and Time Series for helpful comments. I acknowledge financial support from the Spanish Agencia Estatal de Investigación, through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2024-001476-S).

1 Introduction

Vector autoregressions (VARs) have been the workhorse of empirical macroeconomics since the pioneering work by Sims (1980), and are routinely used to predict macroeconomic aggregates or to quantify the transmission mechanisms of structural shocks, such as monetary policy interventions and technology innovations (e.g., Bernanke, 1983; Blanchard and Quah, 1993; Galí, 1999). In practice, these models are almost invariably truncated at a short lag structure, rarely spanning more than four quarters for quarterly data or twelve months for monthly data (e.g., Stock and Watson, 2001, 2002). Microfounded macrotheory points the other way, where standard environments, such as modern dynamic stochastic general equilibrium models, yield reduced-form representations featuring long or infinite autoregressive lag structures (e.g., Fernández-Villaverde, Rubio-Ramírez, Sargent and Watson, 2007; Ravenna, 2007), as do moving average dynamics or omitted variables. Limiting empirical specifications to short lag structures thus leaves an underfitted model, whose distortions propagate into forecasts and impulse response functions (IRFs).

Short lag structures are not an oversight, but a deliberate compromise rooted in the bias-variance trade-off. Estimating a much longer lag structure would remove the truncation bias, yet at a variance cost that is precisely what makes parsimony attractive in the first place (Stock and Watson, 2001; Giannone, Lenza and Primiceri, 2015). Low-order specifications accordingly outperform heavily parameterized alternatives in forecasting so regularly that avoiding unnecessarily large lag structures has become a textbook recommendation (see, e.g., Lütkepohl, 1985, 2005). The objective is seldom to uncover the true lag order of the data-generating process, but to construct an effective approximation for the task at hand, rendering lag length selection inherently conditional on the ultimate empirical objective (Ivanov and Kilian, 2005; Baumeister, 2025; González-Casasús and Schorfheide, 2026). Crucially, this preference for parsimony is entirely compatible with distant, higher-order lags that are small yet not zero. In systems characterized by deep habit persistence, adjustment costs, or learning dynamics under bounded rationality, the coefficients on remote lags may decay at a weak or local-to-zero rate (see, e.g., Evans and Honkapohja, 2001; Fernández-Villaverde et al., 2007; Ravenna, 2007; Milani, 2007). Such dynamics are too small to be consistently estimated, and too small for information criteria or specification pre-tests to rule out, yet their collective omission can induce a cumulative underfitting bias that compromises structural conclusions. Reinforcing this exact vulnerability, Montiel Olea, Plagborg-Møller, Qian and Wolf (2026) document that VARs with short-to-moderate lag lengths suffer severe coverage distortions even for such small amounts of misspecification.

This paper takes the omitted dynamics as given and makes the resulting bias the object of inference. The starting point is a simple observation: when higher-order dynamics are weak, the estimate of the omitted lag block obtained from a VAR is not consistent for the true drift, but is instead asymptotically normal and centered on it. A test of the null hypothesis of no misspecification compares that estimate to zero, and therefore retains the unknown drift, which is precisely why conventional pre-tests are not consistent against these local alternatives. Comparing it instead to the true drift cancels the unestimable term and leaves a pivotal limiting distribution. Inverting it delivers valid confidence sets for the misspecification bias in any forecast or impulse response, as well as misspecification-robust confidence bands for those objects themselves.

It is worth being explicit at the outset about which of the two is new, since the paper's main contribution rests on the first. The confidence interval for the bias has no counterpart among the tools available to

applied VAR practice. It quantifies how much a given IRF or forecast is distorted by lag truncation, in the units of the object itself, and it is available precisely where a pre-test is uninformative. The robust bands are a byproduct. Because the correction is orthogonal to the score of the restricted model, they inherit, to first order, the variance of an unrestricted VAR, so they deliver what estimating the larger system would have delivered anyway, up to finite-sample differences. What they add is that the correction is applied to the parsimonious specification the researcher reports, and that the same ingredients also produce the interval for the bias. Section 6.4 makes the equivalence precise and Section 7.3 documents it numerically. Both objects depend on a maintained window of omitted lag blocks, which is a user choice, much as the degree of lag augmentation is in a lag-augmented local projection.

To formally analyze this environment, I implement a drifting-parameter framework where the true unrestricted VAR(p) model drifts towards the restricted VAR(p_1), $p_1 < p$, used by the econometrician. The wedge between the two models captures the dynamic misspecification intrinsic to lag truncation, and is designed to yield non-trivial mean-squared error rankings among VARs of different lag orders. Parsimonious lag selection is thus accepted as a deliberate and practical compromise rather than an error to be corrected: the researcher retains the specification that would have been reported anyway and, alongside it, reports how much of its implied dynamics may be distorted. Section 7.1 summarizes the procedure step by step.

Efforts to explicitly quantify macroeconometric misspecification have historically relied on mapping structural models against unrestricted time-series benchmarks. Del Negro and Schorfheide (2004) and Del Negro, Schorfheide, Smets and Wouters (2007) leverage structural economic restrictions as VAR priors to scale aggregate dynamic misspecification, while Inoue, Kuo and Rossi (2020) embed non-structural wedges to pinpoint localized breakdowns in equilibrium conditions. More recently, the focus has shifted toward standalone lag truncation distortions without conditioning on a structural benchmark. Li, Plagborg-Møller and Wolf (2024) and Montiel Olea et al. (2026) document how truncating lag profiles to optimize finite-sample risk introduces substantial extrapolation biases into downstream IRFs, while De Graeve and Westermark (2025) challenge the conventional preference for parsimony altogether. Relative to this literature, which evaluates aggregate risk or ranks candidate specifications, the analysis here shifts the focus from model selection and averaging to direct structural inference, treating the misspecification bias itself as the primary object of interest. Dynamic misspecification is known to alter the ranking of predictors or estimators of IRFs and to yield suboptimal hyperparameter choices or model averaging configurations (Schorfheide, 2005; Lohmeyer, Palm, Reuvers and Urbain, 2019; González-Casasús and Schorfheide, 2026), yet such criteria remain agnostic about the direction and magnitude of the distortion embedded within the chosen specification.

Methodologically, the framework also connects the two dominant paradigms in empirical macroeconomics, providing an alternative to the practice of switching to multi-step regressions when concerned about model misspecification (Weiss, 1991; Jordà, 2005). The relation between the two is by now well understood: Plagborg-Møller and Wolf (2021) show that local projections and VARs estimate the same impulse responses in population, so that the choice between them is one of finite-sample bias against variance rather than of estimand (see, also, Ludwig, 2024), and Montiel Olea, Plagborg-Møller, Qian and Wolf (2025) survey the practical terms of that trade-off. It is precisely because the estimand is common that the bias studied here can be expressed as a property of the VAR and compared with what a projection would deliver. While lag-augmented local projections deliver valid inference under dynamic misspecification, they pay a steep variance penalty at longer horizons (Montiel Olea et al., 2026). The misspecification-robust VAR approach

proposed here internalizes the analytic mapping of the bias, hence retains the parametric efficiency of the VAR class while matching those coverage guarantees, often resulting in narrower and more informative intervals, so long as dynamic misspecification is not large and highly persistent. It should be noted that the ability to construct valid inference for the local-to-zero parameters is intrinsic to the structure of this problem and cannot be exploited generally. Notably, the same recentering logic is used by Gánics, Inoue and Rossi (2019) to propose confidence sets for the bias in weak instrumental variable settings.

Simulation results show that the proposed methods attain coverage close to the nominal level across a variety of sample sizes, horizons of interest, and degrees of misspecification. I then revisit the widely studied price puzzle. In a standard three-variable VAR, estimated with a lag order that both the AIC and a Wald pre-test endorse, the confidence intervals for the misspecification bias in the response of prices to a contractionary monetary policy shock are positive and bounded away from zero at several horizons. The estimated price responses are therefore significantly biased upwards, by a magnitude the interval quantifies, and a share of the measured puzzle is accounted for without augmenting the cross-sectional information set, as the puzzle is typically resolved. More broadly, the framework equips applied researchers with a tool to evaluate whether an empirical anomaly reflects a genuine structural phenomenon or merely truncated system dynamics.

The remainder of the paper is organized as follows. Section 2 motivates the problem in an IRF application, while Section 3 lays out the general econometric framework. Section 4 characterizes the dynamic effects of misspecification, and Section 5 shows that conventional pre-tests cannot reliably detect them. Section 6 develops the robust inference procedures. Section 7 lays out the implementation steps and reports the numerical performance of the proposed inference tools, before applying them to re-evaluate the price puzzle in Section 8. Section 9 concludes. All the proofs, as well as additional analyses, are collected in the Appendix.

2 Illustration for IRF Analysis

In this section, I motivate the main results of the paper in a simplified setting applied to IRF estimation; see Appendix B for more details. Suppose that the data is generated according to the following process:

$$\begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} y_{t-1,1} \\ y_{t-1,2} \end{bmatrix} + \begin{bmatrix} \Delta_{11}/\sqrt{T} & \Delta_{12}/\sqrt{T} \\ \Delta_{21}/\sqrt{T} & \Delta_{22}/\sqrt{T} \end{bmatrix} \begin{bmatrix} y_{t-2,1} \\ y_{t-2,2} \end{bmatrix} + \begin{bmatrix} u_{t,1} \\ u_{t,2} \end{bmatrix}.$$

This data-generating process aims to capture a common concern among applied macroeconomists, whereby fitting a stylized VAR model to the data one may ignore small, yet important dynamic effects in finite samples. To illustrate the practical implications of this framework, let us posit that the object of interest is the impulse response of variable 2 to a one-unit reduced-form shock in variable 1 at horizon h.

The practitioner, unaware of the true data-generating process, might choose to estimate a VAR(1) model. When the terms $\Delta/\sqrt{T}$ are small, this choice would typically be supported by the usual model selection criteria. For instance, consistent criteria such as the BIC select the VAR(1) with probability approaching one, while criteria with fixed penalties such as the AIC do so with non-trivial probability; see Section 5.

In estimating a VAR(1), the practitioner underfits the data by omitting the second lag matrix, which gives

rise to asymptotic bias in the estimated parameters that extends to any dynamic effects that depend on them. This modeling choice, however, presents a classic bias-variance trade-off. While expanding the specification to a VAR(2) would completely eliminate this systematic bias, it forces the practitioner to estimate four additional parameters from a weak, shrinking second lag. In finite samples, this over-specification injects substantial sampling noise into the estimated IRFs. By restricting the model to a VAR(1), the researcher implicitly accepts a potential bias as a deliberate compromise to achieve a significantly tighter and more stable estimator.

The estimated IRF matrix at horizon h from the practitioner's VAR(1) model is given by $\hat{B}^h$. Thus, the estimated IRF of variable 2 to a one-unit reduced-form shock in variable 1 at horizon h can be compactly written as $e_2'\hat{B}^h e_1$, where $e_1 = (1,0)'$ and $e_2 = (0,1)'$ are selection vectors.

The asymptotic bias in the estimated IRF $e_2'\hat{B}^h e_1$ at horizon $h = 1$ can be shown to be

$$e_2'\Delta\Omega_{21}\Omega_{11}^{-1}e_1,$$

where Ω is the second-moment matrix of $z_{t-1} = (y_{t-1}', y_{t-2}')'$, partitioned into 2×2 blocks Ω_{ij} as in Section 3, while at horizon $h = 2$ it is given by

$$e_2'\left(B\Delta\Omega_{21}\Omega_{11}^{-1} + \Delta\Omega_{21}\Omega_{11}^{-1}B - \Delta\right)e_1.$$

The bias expression becomes increasingly convoluted for longer horizons because the initial misspecification error dynamically propagates through the autoregressive system and recursively compounds with the ongoing effects of the omitted lag at each subsequent step; see Section 4 for the general characterization and Appendix B for the expression in this example.

Because Δ cannot be consistently estimated, it is impossible to calculate the exact value of this distortion, or to remove it from the IRF estimate without leaving uncertainty behind. However, the degree of misspecification Δ proves to be crucial to assess the biases in the estimated IRFs, hence in balancing the mean-squared error rankings among different modeling choices or conducting valid inference. Fortunately, this paper demonstrates that one can still learn meaningful information about it.

It is worth asking first why the natural diagnostic does not settle the matter. A conventional F-test evaluating the significance of the second lag in a VAR(2), or equivalently the Wald test of Section 5, is based on the difference between the estimated misspecification $\hat{\Delta} = \sqrt{T}\hat{B}^{(2)}$ and zero, where $\hat{B}^{(2)}$ denotes the estimated coefficient matrix on the second lag of a VAR(2). This difference therefore retains information about the true, unestimable degree of misspecification Δ, leading to a non-pivotal limiting distribution. A failure to reject cannot then certify that the omitted lag is irrelevant.

The observation behind this paper is that the very same estimate is informative once it is compared to the right benchmark. In particular, it is possible to show that

$$\mathrm{vec}(\hat{\Delta}) \xrightarrow{d} N(\mathrm{vec}(\Delta), V),$$

where V is the matrix defined in Section 3. That is, $\hat{\Delta}$ is not a consistent estimate of Δ, but is instead an asymptotically normal random variable centered at Δ. Centering the statistic at the true degree of

misspecification rather than at zero, the unestimable drift perfectly cancels out:

$$\left(\text{vec}(\hat{\Delta}) - \text{vec}(\Delta)\right)' V^{-1} \left(\text{vec}(\hat{\Delta}) - \text{vec}(\Delta)\right) \xrightarrow{d} \chi^2_4,$$

where χ^2_4 denotes a chi-squared distribution with 4 degrees of freedom, the number of elements of Δ. Consequently, the limiting distribution does not depend on how close the true misspecification is to zero, yielding a pivotal distribution that permits standard, valid statistical inference. One conveniently obtains a valid confidence set for Δ by inverting it.

It might appear counterintuitive that we can obtain valid confidence intervals for the degree of misspecification despite the fact that it cannot be consistently estimated. The recentering above is what makes it possible. Because the asymptotic bias in the estimated IRFs is linear in Δ, a projection argument then guarantees that a valid confidence interval for it can be constructed from a confidence set for Δ. In fact, applying the same recentering directly to the estimated bias, which is a known linear combination of $\hat{\Delta}$, delivers an interval that is sharp rather than conservative, and this is the route recommended in Section 6. These are the crucial mechanisms behind the proposed robust inference method.

The main benefit of the proposed method is being able to retain the conveniently parsimonious VAR(1) while acknowledging any potential dynamic distortions due to the omitted second lag in a transparent manner. The practitioner reports the impulse response of the VAR(1) alongside an interval for how much it may be distorted by omitted dynamics. As a byproduct, subtracting the estimated bias from the reported response also delivers misspecification-robust confidence bands for the impulse response itself, whose width is comparable to that of the unrestricted VAR(2); see Section 6.4.

3 Framework

Based on the local-to-zero misspecification paradigm introduced by Lohmeyer et al. (2019), let the K-dimensional time series $\{\{y_t\}_{t=-\infty}^{T}\}_{T=1}^{\infty}$ constitute a vector triangular array generated by the VAR processes

$$y_t = B_1 y_{t-1} + \cdots + B_{p_1} y_{t-p_1} + \frac{\Delta_1}{\sqrt{T}} y_{t-p_1-1} + \cdots + \frac{\Delta_{p_2}}{\sqrt{T}} y_{t-p_1-p_2} + u_t, \tag{1}$$

where $\{B_i\}_{i=1}^{p_1}$ and $\{\Delta_i\}_{i=1}^{p_2}$ are $K \times K$ coefficient matrices.

This drifting data-generating process serves as a formal asymptotic thought experiment designed to capture the pervasive lag truncation problem in VARs, whether arising from an intentional pursuit of parameter parsimony or unintentionally induced by data-driven lag selection procedures. Specifically, it formalizes how standard rules of thumb or information criteria often truncate the system at a restricted lag order p_1, thereby introducing a post-model-selection vulnerability that omits higher-order terms which may appear statistically negligible in finite samples yet remain crucial for propagating the full structural dynamics over time. The drifting parameter analysis thus provides an analytical framework to evaluate the fundamental bias-variance tradeoffs inherent to low-order lag truncation or other forms of omitted relevant dynamic effects, such as local moving average dynamics or omitted weakly exogenous variables (see Appendices E and F; cf. Schorfheide, 2005). Throughout this paper, the observable data vector y_t is assumed to be mean-zero, which in empirical applications is standardly achieved by demeaning the series prior to estimation.

The window of p_2 higher order lags represents the source of misspecification. While the baseline specifica-

tion in (1) presumes a finite lag order p_2 for the misspecification drift, this framework readily accommodates some data-generating processes featuring a true VAR(∞) representation. This setting captures the standard empirical paradigm where only a finite number of omitted dynamic terms exert a first-order impact on the system. Beyond the threshold $p_1 + p_2$, the structural coefficients typically decay at a rapid rate, rendering their cumulative marginal contribution asymptotically negligible relative to the $O(1/\sqrt{T})$ local drift (cf. Claeskens and Hjort, 2003, Section 10.1). This environment directly comports with the linearized solutions of standard structural macroeconomic models, which yield reduced-form VARs whose coefficients decay exponentially as a function of the system's stable eigenvalues (Fernández-Villaverde et al., 2007; Ravenna, 2007).[1]

To simplify notation, I collect all the parameters in the following matrices:

$$B = \begin{bmatrix} B_1 & \cdots B_{p_1} \end{bmatrix}, \quad C_T = \Delta/\sqrt{T}, \quad \Delta = \begin{bmatrix} \Delta_1 & \cdots \Delta_{p_2} \end{bmatrix}, \quad \Theta_T = \begin{bmatrix} B & C_T \end{bmatrix}.$$

I also stack the observations over time to define, letting $p = p_1 + p_2$:

$$Y = (y_1, \ldots, y_T), \quad z_t = (y_t', \ldots, y_{t-p+1}')', \quad Z = (z_0, \ldots, z_{T-1}), \quad U = (u_1, \ldots, u_T).$$

The following regularity conditions will be imposed throughout.

Assumption 1 *The sequence $\{u_t\}$ of random K-vectors is an independent and identically distributed sequence with mean vector zero, a positive definite covariance matrix $\mathbb{E}[u_t u_t'] = \Sigma$, and there exists a $c > 0$ such that $\mathbb{E}[|u_{it}u_{jt}u_{kt}u_{mt}|] < c < \infty$ for $i, j, k, m = 1, 2, \ldots, K$.*

Assumption 1 establishes standard regularity conditions to apply a Central Limit Theorem. The i.i.d. assumption may in some cases be restrictive, as it rules out temporal dependence in volatility, such as conditional heteroskedasticity, yet can be relaxed. For instance, if $\{u_t\}$ is a martingale difference sequence, every asymptotic normality result below continues to hold, and the only change is that the variance V of the estimated drift acquires a sandwich form. Remark 1 in Appendix A gives that expression and the corresponding estimator, which specialize to the formulae used in the main text under conditional homoskedasticity.

Assumption 2 $\det\left(\mathrm{I}_K z^p - B_1 z^{p-1} - \cdots - B_{p_1} z^{p_2} - \frac{\Delta_1}{\sqrt{T}} z^{p_2-1} - \cdots - \frac{\Delta_{p_2}}{\sqrt{T}}\right) \neq 0$ *for all* $|z| \geqslant 1, T \in \mathbb{N}$.

Assumption 2 ensures that the underlying, true data-generating process remains stationary for all sample sizes. This condition is standard, and implies that the local perturbations are small enough relative to the baseline dynamics to prevent the system from crossing the boundary into non-stationarity. A violation of this condition would cause parameters to converge at non-standard rates, invalidating the framework employed here. Indeed, by construction, the misspecification drift term operates at the parametric rate.

Assumption 3 $\det\left(\mathrm{I}_K z^{p_1} - B_1 z^{p_1-1} - \cdots - B_{p_1}\right) \neq 0$ *for all* $|z| \geqslant 1$.

Parallel to Assumption 2, Assumption 3 requires that the true asymptotic model of order p_1 is also stable and stationary. This assumption is often mild when paired with Assumption 2. In practice, if the true high-order system is stable, a parsimonious lower-order truncation that captures the primary auto-covariance structure will routinely inherit this stability.

[1] See Appendix C for a more detailed discussion on the extension to VAR(∞).

The researcher chooses to estimate a VAR(m) model, for $p_1 \leqslant m \leqslant p$, and the analysis throughout treats m, p_1 and p_2 as fixed integers chosen by them. Throughout, I will use the convention that the baseline order p_1 is the specification the researcher would have reported in any case, arrived at by a rule of thumb, an information criterion, or a task-based criterion, while the window p_2 is the number of further lag blocks against which protection is sought. This comports with empirical practice where it is standard for researchers to condition their analysis on a data-driven lag order selection. The role of p_2 is analogous to the degree of lag augmentation in lag-augmented local projections: a user-specified window, not an estimated quantity, and the appropriate robustness check is to report the results for an alternative value of it.

Selection matrices are used to relate all estimators to the estimator using p lags. That is, for some integer m, such that $p_1 \leqslant m \leqslant p$,

$$S_0 = S_0^{(1)} \otimes \mathrm{I}_K, \text{ with } S_0^{(1)} = \begin{bmatrix} 0_{p_1 \times p_2} \\ \mathrm{I}_{p_2} \end{bmatrix}, \qquad S_m = S_m^{(1)} \otimes \mathrm{I}_K, \text{ with } S_m^{(1)} = \begin{bmatrix} \mathrm{I}_m \\ 0_{(p-m)\times m} \end{bmatrix},$$
$$\Pi_m' = \Pi_m^{'(1)} \otimes \mathrm{I}_K, \text{ with } \Pi_m^{'(1)} = \begin{bmatrix} \mathrm{I}_{m-p_1} \\ 0_{(p-m)\times(m-p_1)} \end{bmatrix}.$$

When the practitioner estimates a VAR(m), the regressor matrix satisfies $Z_m = S_m' Z$ and the population coefficients are $\Theta_{T,m} = \Theta_T S_m$. Thus, the $K \times Km$ estimated OLS coefficients can be shown to satisfy

$$\hat{\Theta}_{T,m} = \Theta_{T,m} + C_T(\mathrm{I}_{Kp_2} - \Pi_m'\Pi_m)S_0' Z Z_m'(Z_m Z_m')^{-1} + U Z_m'(Z_m Z_m')^{-1}. \tag{2}$$

To lighten notation, I henceforth write

$$\delta = \mathrm{vec}(\Delta) = \sqrt{T}\mathrm{vec}(\Theta_{T,p} S_0), \qquad \hat{\delta} = \mathrm{vec}(\hat{\Delta}) = \sqrt{T}\mathrm{vec}(\hat{\Theta}_{T,p} S_0),$$

for the degree of dynamic misspecification and its estimate obtained from an unrestricted VAR(p) model, respectively. Moreover, I will denote

$$\Omega = \mathrm{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} z_{t-1} z_{t-1}', \qquad V = [S_0'\Omega^{-1}S_0] \otimes \Sigma, \qquad S_Z = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(z_{t-1} \otimes u_t),$$

where Ω is partitioned into $K \times K$ blocks $\Omega_{ij} = \mathrm{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} y_{t-i} y_{t-j}'$, V is the asymptotic variance of $\hat{\delta}$ derived in Section 6, and S_Z is the standardized score of the unrestricted regression, which satisfies $S_Z \xrightarrow{d} N(0, \Omega \otimes \Sigma)$ under Assumption 1.

4 Dynamic Effects of Misspecification

The misspecification drift terms summarized by δ will affect any dynamic analysis of the model that depends on the autoregressive matrices, including forecasts and IRFs, inducing an asymptotic bias term in their expressions. The results in this section show that the bias in these objects is a particular linear function of δ, while their variances are unaffected by misspecification.

4.1 IRF Estimation

The impulse response at horizon h is defined as the hth coefficient matrix of the MA(∞) representation $y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}$, with $\Phi_0 = \mathrm{I}_K$ (e.g., Lütkepohl, 1990). Let A_T be the companion form representation of Θ_T, A_∞ the $Kp \times Kp$ companion matrix related to the process in which the misspecification coefficients have been set to zero, and $J = \begin{bmatrix} \mathrm{I}_K & 0_{K \times K(p-1)} \end{bmatrix}$ a selection matrix of dimension $K \times Kp$. The IRF matrices can be obtained from the VAR autoregressive matrices as $\Phi_i = J A_T^i J'$.

The estimated reduced-form IRFs are obtained via the companion matrix of the restricted model as $\hat{\Phi}_h = J_m \hat{A}_{T,m}^h J_m'$, where $\hat{A}_{T,m}$ is the $Km \times Km$ companion matrix constructed from the estimates of the VAR(m) specification, and $J_m = [\mathrm{I}_K, 0_{K \times K(m-1)}]$ is the selection matrix mapping the companion space back to the observational dimensions.

Lemma 1 *The estimated IRF at horizon h, $\hat{\Phi}_h$, admits the asymptotically linear representation*

$$\sqrt{T}\left(\mathrm{vec}(\hat{\Phi}_h) - \mathrm{vec}(\Phi_h)\right) = G_h C_m \delta + G_h P_m S_Z + o_p(1) \xrightarrow{d} N\left(G_h C_m \delta, G_h P_m (\Omega \otimes \Sigma) P_m G_h'\right),$$

where

$$G_h = \sum_{j=0}^{h-1} J (A_\infty')^{h-1-j} \otimes \Phi_j, \quad P_m = S_m [S_m' \Omega S_m]^{-1} S_m' \otimes I_K,$$
$$C_m = \left(S_m [S_m' \Omega S_m]^{-1} S_m' \Omega - I_{Kp}\right) S_0 \left(I_{Kp_2} - \Pi_m' \Pi_m\right) \otimes I_K.$$

The result in Lemma 1, derived in the corollary on page 769 of Lohmeyer et al. (2019),[2] characterizes the asymptotic distribution of reduced-form IRFs in the presence of local dynamic misspecification, where the distortions appear in the form of an asymptotic bias term. It should be noted that the coefficients Φ_j appearing inside G_h are those of the limit process in which the drift has been set to zero, $\Phi_j = J A_\infty^j J'$, whereas the Φ_h being estimated on the left-hand side is that of the drifting process (1). The two differ by $O(T^{-1/2})$, which is immaterial for the first-order result but fixes what the plug-in $\hat{G}_h$ should evaluate, namely the restricted estimates. Per the continuous mapping theorem, the next result follows.

Corollary 1 *Let $\Psi_h = \Phi_h P$ denote the structural impulse response matrix at horizon h, where P is a non-stochastic $K \times K$ shock identification matrix. The estimated structural IRF, $\hat{\Psi}_h = \hat{\Phi}_h P$, satisfies*

$$\sqrt{T}\left(\mathrm{vec}(\hat{\Psi}_h) - \mathrm{vec}(\Psi_h)\right) \xrightarrow{d} N\left(H_P G_h C_m \delta,\ H_P G_h P_m (\Omega \otimes \Sigma) P_m G_h' H_P'\right),$$

where $H_P = (P' \otimes \mathrm{I}_K)$.

Corollary 1 states the asymptotic distribution of the structural IRF matrix treating the identification matrix P as fixed. In applied practice, P is often estimated as a function of the data. For instance, one popular choice is the lower-triangular Cholesky factor of the reduced-form residual covariance matrix, where $\hat{P}\hat{P}' = \hat{\Sigma}$. Crucially, the inferential challenges addressed in this framework are entirely decoupled from this structural shock identification problem for schemes that solely rely on $\hat{\Sigma}$. Under the local misspecification paradigm, the contribution of the omitted higher-order lags to the residual covariance matrix Σ is of order $O(1/T)$ and vanishes asymptotically. Consequently, the dynamic misspecification distortions

[2] Lemma 1 fixes a typo in Lohmeyer et al. (2019), where A_∞ is missing a transpose.

isolated in this paper constitute a purely reduced-form phenomenon that operate orthogonally to, and do not overlap with, the structural identification strategy, hence the true asymptotic bias vector for the structural responses remains exactly $H_P G_h C_m \delta$. When alternative identification schemes are employed, Appendix D argues that dynamic misspecification may interfere in the recovery of structural shocks, leading to $O(1/\sqrt{T})$ distortions. For instance, when performing identification via sign restrictions over multiple horizons, the reduced-form bias dynamically deforms the boundaries of the admissible structural space, and for external instrument identification, dynamic misspecification introduces a compounding asymptotic bias into the structural impact vector whenever the instrument is not strictly lead/lag exogenous, which is exacerbated when the proxies are weak.

Therefore, when using identification schemes that only rely on $\hat{\Sigma}$, the replacement of P by its sample counterpart does not affect the validity of the proposed inference methods for the misspecification bias in Section 6.

4.2 Forecasting

To analyze the implications of dynamic misspecification on forecasts, it becomes helpful to write the data-generating process in companion form:

$$z_t = A_T z_{t-1} + J' u_t,$$

where, as before, A_T is the companion form representation of Θ_T. The expected value of y_{T+h} conditional on the information at time T is given by:

$$y_{T+h|T} = \mathbb{E}[y_{T+h}|\mathcal{F}_T] = J A_T^h z_T, \quad \mathcal{F}_t = \sigma(u_s, -\infty < s \leqslant t).$$

When the practitioner estimates a VAR(m) model, the implied forecast is $\hat{y}_{T+h|T} = J_m \hat{A}_{T,m}^h S_m' z_T$.

Lemma 2 *The forecast estimation error conditional on the final observation z_T admits the asymptotically linear representation*

$$\sqrt{T}\left(\hat{y}_{T+h|T} - y_{T+h|T}\right) = G_h C_m \delta + G_h P_m S_Z + o_p(1) \xrightarrow{d} N\left(G_h C_m \delta, G_h P_m (\Omega \otimes \Sigma) P_m G_h'\right),$$

where

$$G_h = \sum_{j=0}^{h-1} \left(z_T'(A_\infty')^j \otimes \Phi_{h-1-j}\right), \quad P_m = S_m[S_m'\Omega S_m]^{-1} S_m' \otimes I_K,$$
$$C_m = \left(S_m[S_m'\Omega S_m]^{-1} S_m'\Omega - I_{Kp}\right) S_0 \left(I_{Kp_2} - \Pi_m'\Pi_m\right) \otimes I_K.$$

The symbols G_h, P_m and C_m are deliberately reused across Lemmas 1 and 2 to allow Section 6 treating impulse responses and forecasts in one pass. The matrices P_m and C_m are the exact same objects in the two lemmas, as they depend only on the design of the estimation problem and not on the object of interest. However, the matrix G_h differs. It is the Jacobian of the target with respect to the autoregressive coefficients, so its row dimension is that of the target: $K^2 \times K^2 p$ for the vectorized impulse response matrix in Lemma 1, and $K \times K^2 p$ for the h-step forecast vector in Lemma 2. Its column dimension, $K^2 p$, is common to both,

which is precisely why the composition $G_h C_m \delta$ has the same form in the two cases. I henceforth let n generically denote the row dimension of G_h for the object at hand.

The result in Lemma 2, which follows as a special case of Theorem 2 in Lohmeyer et al. (2019), characterizes the asymptotic distribution of h-step-ahead forecasts subject to local dynamic misspecification. As with the impulse response functions, local dynamic misspecification merely introduces a deterministic location shift in the asymptotic distribution.

5 Pre-Testing Dangers

A natural safeguard against the distortions characterized in Section 4 would be to pre-test for dynamic misspecification, and to take these biases seriously only when the test rejects. However, standard tests for the null hypothesis of no dynamic misspecification are invalid in this environment, as the next proposition shows.[3]

Proposition 1 *Define the Wald statistic for testing whether the last p_2 lag coefficient blocks are zero as:*

$$W_T = \text{vec}(\hat{\Theta}_{T,p} S_0)' \left([S_0'(ZZ')^{-1} S_0] \otimes \hat{\Sigma} \right)^{-1} \text{vec}(\hat{\Theta}_{T,p} S_0),$$

where $\hat{\Sigma}$ is a consistent estimator of the innovation covariance matrix Σ.

Under Assumptions 1-2 and the locally misspecified data-generating process (1)*:*

(a) $W_T \xrightarrow{d} \chi^2_{K^2 p_2}(\lambda)$, *where* $\lambda = \delta' V^{-1} \delta$;

(b) $\lim_{T\to\infty} \mathbb{P}(W_T > \chi^2_{K^2 p_2, 1-\alpha}) < 1$.

In words, Proposition 1 shows that the Wald statistic (which is asymptotically equivalent to the F-statistic multiplied by the number of restrictions $K^2 p_2$) for testing the null of no dynamic misspecification has a non-pivotal distribution in this environment, and depends on a parameter that cannot be consistently estimated. As a consequence, such a test is invalid as a consistent tool for detecting small misspecification.

The result in Proposition 1 demonstrates that a failure to reject the null cannot certify negligible misspecification, and that binary pre-testing thereby leaves downstream inference exposed to local bias. Montiel Olea et al. (2026) reach the same conclusion for a Hausman test that contrasts the VAR estimate of an impulse response with a local projection. They read it as a reason not to rely on ex post specification tests and to either use local projections or lengthen the VAR until its confidence intervals are no longer narrower than those of a local projection. This paper provides an alternative, complementary view on the problem, targeted to researchers who want to remain in the VAR class and retain a parsimonious model. Section 6 employs a Hausman contrast of the same kind, taken within the VAR class, but instead inverts it into a confidence interval for the dynamic misspecification bias. This allows researchers to report the bias those specifications may carry in a transparent way.

The failure of conventional pre-tests extends naturally to data-driven lag selection via standard information criteria. As established in the foundational framework of Claeskens and Hjort (2003), penalized likelihood

[3] I denote by $\chi^2_d(\lambda)$ the noncentral chi-squared distribution with d degrees of freedom and noncentrality parameter λ. Similarly, I let $\chi^2_{d,1-\alpha}$ be the $1-\alpha$ quantile of a (central) chi-squared distribution with d degrees of freedom.

criteria fundamentally fail to correctly resolve model uncertainty under local-to-zero misspecification. For criteria with constant penalties, such as the AIC, this implies that the asymptotic probability of selecting either the restricted or the unrestricted model remains strictly bounded between zero and one. Consequently, post-selection estimators inherit a non-standard, highly non-pivotal distribution. Conversely, for consistent criteria with penalties proportional to $\ln(T)$, such as the BIC, the deterministic penalty strictly dominates the $O(1)$ likelihood improvement in large samples. As a result, the BIC selects the underfitted VAR(p_1) model with probability approaching one. In both cases, standard model selection algorithms structurally fail to incorporate the $O(1/\sqrt{T})$ dynamics, systematically exposing the econometrician to the asymptotic bias in downstream structural analysis. Failures of both pre-tests and selection criteria are visible in the simulations of Section 7.

6 Quantifying the Extent of Misspecification Bias

The misspecification bias plays a key role in the properties of parameter estimates, hence in shaping the path of dynamic effects implied by the VAR. However, δ is by construction too small to be consistently estimated, precluding from any costless analytical bias corrections.

In this section, I show that while the asymptotic bias cannot be consistently estimated, it is possible to construct valid inference for it. The starting point of the analysis is Lemma 3 below, yet I distinguish two cases. I then show how the same insight delivers misspecification-robust inference for the IRFs and forecasts themselves.

6.1 Basic Insight

From the decomposition in (2), we deduce the following result for the estimated $p_1 + 1$ through p lags (cf. Lohmeyer et al., 2019, Theorem 1(a)):[4]

Lemma 3 *Under Assumptions 1-2,*

$$\sqrt{T}\left(\text{vec}(\hat{\Theta}_{T,p}S_0) - \text{vec}(\Theta_{T,p}S_0)\right) \xrightarrow{d} N(0, V).$$

In other words, Lemma 3 shows that $\hat{\delta}$ is not a consistent estimate of δ. Rather, $\hat{\delta}$ is an asymptotically normal random variable centered at δ. This key observation is the foundation of the inference results that follow.

6.2 Dynamic Misspecification

The first object of interest in applied practice could be the omitted dynamic effects Δ. This object could serve to quantify ignored lagged effects left over by selection criteria. Let $d = K^2 p_2$ denote the dimension of δ. Despite the d-dimensional vector δ itself cannot be consistently estimated, Proposition 2 demonstrates that one can statistically bound its magnitude in a rigorous way.

[4]The results in Claeskens and Hjort (2003) imply that the score functions of all intermediate models are asymptotically linear combinations of the score function of the full unrestricted model. Consequently, the estimate of the omitted dynamics in any intermediate VAR(m) model, $\hat{\delta}_m$, is asymptotically just a deterministic linear projection of the unrestricted estimate from the unrestricted VAR(p) model, $\hat{\delta}$.

Proposition 2 *Under Assumptions 1-2, the confidence set*

$$\tilde{\mathcal{C}} = \left\{ \delta \in \mathbb{R}^d : (\hat{\delta} - \delta)' \hat{V}^{-1} (\hat{\delta} - \delta) \leqslant \chi^2_{d,1-\alpha} \right\},$$

where $\hat{V}$ is a consistent estimate of V, satisfies $\lim_{T\to\infty} \mathbb{P}(\delta \in \tilde{\mathcal{C}}) = 1 - \alpha$.

Proposition 2 guarantees that $\tilde{\mathcal{C}}$ is a valid joint $1-\alpha$ confidence region for the degree of dynamic misspecification δ. The estimate $\hat{V}$ can be obtained by replacing the population expectations Ω and Σ by sample averages, which are consistent despite the dynamic misspecification.

6.3 Target Parameters

Alternatively, one may only be interested in quantifying dynamic misspecification to the extent that it affects relevant target parameters. Let b denote the asymptotic bias of a specific quantity of interest, e.g., the h-step ahead forecast of the i-th variable, or a specific IRF coefficient. This special case is important because it is commonplace in economic applications.

From the derivations in Section 4, the bias vector is generally given by $G_h C_m \delta$, where recall that the specific expression for G_h, and in particular its row dimension n, depends on the object of interest: $n = K^2$ for impulse responses and $n = K$ for forecasts. Let us extract the bias of the target parameter using a selection matrix E of dimension $q \times n$, where q is the dimension of the quantity of interest:

$$b = E G_h C_m \delta.$$

To simplify notation, I define the matrix $w = E G_h C_m$. Thus, the bias is a linear combination of the misspecification parameters:

$$b = w\delta.$$

Importantly, observe that the matrix w can be consistently estimated from a VAR(m) by replacing the population objects with sample counterparts.

To formally construct valid confidence intervals for a target parameter, one could use the geometric property of projection for linear transformations. Since both the forecast and IRF biases can be represented as a linear mapping of δ, one could construct the desired confidence intervals by projecting the joint $1-\alpha$ confidence region for δ, $\tilde{\mathcal{C}}$, onto the respective dimensions of interest:

$$\mathcal{C}^{\text{pr}} = \left\{ \hat{w}\delta \in \mathbb{R}^q : \delta \in \tilde{\mathcal{C}} \right\}.$$

Because $\delta \mapsto \hat{w}\delta$ is a continuous linear map and $\tilde{\mathcal{C}}$ is a closed, bounded, convex set, its image is itself closed, bounded, and convex. That the image also inherits the right coverage is the content of Proposition 3.

Proposition 3 *Under Assumptions 1-3, assuming w to have full row rank, $\lim_{T\to\infty} \mathbb{P}(b \in \mathcal{C}^{\text{pr}}) \geqslant 1 - \alpha$.*

When the target parameter is scalar, the projected confidence set simplifies to

$$\mathcal{C}^{\text{pr}} = \left[\min_{\delta \in \tilde{\mathcal{C}}} \hat{w}\delta,\ \max_{\delta \in \tilde{\mathcal{C}}} \hat{w}\delta \right].$$

Applying the Cauchy-Schwarz inequality to the optimization problem yields a closed-form, data-driven tolerance radius for the asymptotic bias. Specifically, the scalar bias b is bounded within the interval:

$$b \in \left[\hat{w}\hat{\delta} - \mathfrak{r}_\alpha,\ \hat{w}\hat{\delta} + \mathfrak{r}_\alpha\right],$$

where the tolerance radius is defined as $\mathfrak{r}_\alpha = \sqrt{\chi^2_{d,1-\alpha}\hat{w}\hat{V}\hat{w}'}$.

The prevailing consensus on bias-aware VAR inference and worst-case mean-squared error analyses has often viewed bounding VAR distortions as practically intractable without imposing stringent a priori assumptions. As recently noted by Montiel Olea et al. (2026), such approaches typically require researchers to impose "an extremely tight bound on the noise-to-signal ratio" to be operationalized. Proposition 3 reframes this requirement. Given the maintained window p_2, the worst-case radius of the dynamic bias is data driven instead of being set by the researcher. The window itself, however, is a user choice, and it may be understood as an analogue of the degree of lag augmentation in a lag-augmented local projection. What the result buys is therefore a substitution of one judgment for another, from an unobservable magnitude to a count of lag blocks that applied researchers already reason about, and a robustness check that consists of varying that count.

There exist numerical methods to efficiently construct the general confidence set in Proposition 3 by exploiting the Kronecker form of the variance V under homoskedasticty. However, there is an inherent grid search over the space for δ, whose dimension is K^2p_2, hence the step may be nontrivial in higher dimensional models.

Proposition 4 shows that there exists a projection-free method that sidesteps the curse of dimensionality due to grid search, and it is sharp. The result is an immediate implication of Lemma 3 due to Slutsky's theorem.

Proposition 4 *Define the* $1-\alpha$ *confidence interval for the misspecification bias* $b = w\delta$ *as*

$$\mathcal{C}^{\text{sh}} = \left\{b \in \mathbb{R}^q : (\hat{w}\hat{\delta} - b)'(\hat{w}\hat{V}\hat{w}')^{-1}(\hat{w}\hat{\delta} - b) \leqslant \chi^2_{q,1-\alpha}\right\}.$$

Under Assumptions 1-3, and assuming w *to have full row rank,* $\lim_{T\to\infty}\mathbb{P}(b \in \mathcal{C}^{\text{sh}}) = 1-\alpha$.

It is crucial to distinguish the inferential scope of Proposition 4 from the confidence regions constructed via Proposition 3: $\mathcal{C}^{\text{pr}}$ is the projection of the joint region and satisfies a coverage inequality, while $\mathcal{C}^{\text{sh}}$ is the sharp set and satisfies a coverage equality. While Proposition 2 delivers a simultaneous joint confidence region for the entire local lag vector δ, its projection onto a low-dimensional structural target in Proposition 3 may often be conservative. Conversely, Proposition 4 circumvents this joint projection step by deriving the limiting distribution directly for the estimate of the target bias vector $b = w\delta$. Thus, it is the recommended approach to conduct target-specific inference, since it achieves exact asymptotic coverage, and the intervals will typically be dramatically tighter and far more informative than those in Proposition 3.[5] Indeed, when $q = 1$, the tolerance radius implied by Proposition 4 tightens to $\mathfrak{r}_\alpha = z_{1-\alpha/2}\sqrt{\hat{w}\hat{V}\hat{w}'}$,

[5]The rank conditions in Propositions 3 and 4 implicitly require $q \leqslant K^2(p-m)$. While this condition holds generally in empirical settings, it may fail, for instance, if a practitioner attempts to construct a joint confidence region for an entire path of impulse responses across many horizons. Although not possible in its current form, the projection logic from Proposition 3 can easily be adapted to allow for the case where $q > K^2(p-m)$, e.g., defining a rectangular band exploiting the row-wise closed-form radius $\mathfrak{r}_{\alpha,j} = \sqrt{\chi^2_{d,1-\alpha}\hat{w}_j\hat{V}\hat{w}_j'}$ (cf. Montiel Olea and Plagborg-Møller, 2019; Lütkepohl, Staszewska-Bystrova and Winker,

where z_α is the α quantile of a standard normal distribution.

The sharp set also admits a familiar interpretation. The estimated bias $\hat{w}\hat{\delta}$ is, to first order, the Hausman contrast between the VAR(m) and the VAR(p) estimates of a given target, and $\hat{w}\hat{V}\hat{w}'$ is the difference of their asymptotic variances. Thus, $\mathcal{C}^{\text{sh}}$ excludes zero exactly when a Hausman test (Hausman, 1978) for that particular target rejects at level α. In particular, the interval inverts a Hausman test tailored to the object of interest, with the variance of the contrast computed directly rather than as a difference of variance estimates, which keeps it valid under the homoskedasticity relaxation in Remark 1. This test should not be confused with the pre-test of Proposition 1, which coincides with the Hausman test on the retained autoregressive coefficients under a mild rank condition (Holly, 1982). The target-specific version concentrates the d degrees of freedom of the pre-test into q, and neither rejection implies the other. The bias interval can exclude zero while the pre-test is far from rejecting, because the latter spends its degrees of freedom on directions of the drift that leave the target unaffected, as in Section 8, and it can cover zero while the pre-test rejects. Both tests are nonetheless subject to the caveat in Proposition 1. The Hausman statistic is the pivotal statistic defining $\mathcal{C}^{\text{sh}}$ evaluated at $b = 0$ rather than at the true bias, so under the local drift its limit is a noncentral χ^2_q with finite noncentrality parameter $b'(wVw')^{-1}b$ and its rejection probability is bounded away from one—exactly as for the Hausman test of Montiel Olea et al. (2026)—whereas the interval retains the pivot by ranging over every value of the bias and therefore delivers a sign and a magnitude whether or not zero is excluded.

6.4 Misspecification-Robust Inference for IRFs and Forecasts

While the confidence sets above quantify the misspecification bias itself and are the main contribution of the paper, Lemma 3 can be used as a byproduct to purge that bias from inference on the dynamic objects of interest, delivering confidence bands for IRFs and forecasts that remain valid under local dynamic misspecification.

As discussed in Section 4, researchers conducting inference on structural IRFs must note that the asymptotic variance of $\hat{\Psi}_h$ will contain an additional additive term, $(\mathrm{I}_K \otimes \Phi_h)\Sigma_P(\mathrm{I}_K \otimes \Phi_h')$, to account for the sampling uncertainty of $\hat{P}$. This is summarized in the following proposition.

Proposition 5 *With some abuse of notation, let the estimated structural IRF be $\hat{\Psi}_h = \hat{\Phi}_h\hat{P}$ and let $\hat{H}_{\hat{P}} = (\hat{P}' \otimes \mathrm{I}_K)$ be the sample counterpart of H_P in Corollary 1. Let $\hat{G}_h, \hat{C}_m$ be consistent estimates of G_h, C_m, respectively. Moreover, suppose that $\hat{P}$ is asymptotically independent of the score S_Z and satisfies $\sqrt{T}\mathrm{vec}(\hat{P} - P) \xrightarrow{d} N(0, \Sigma_P)$. Under Assumptions 1-3,*

$$\sqrt{T}\left(\mathrm{vec}(\hat{\Psi}_h) - \mathrm{vec}(\Psi_h)\right) - \hat{H}_{\hat{P}}\hat{G}_h\hat{C}_m\hat{\delta} \xrightarrow{d} N\left(0, H_P G_h \Gamma_m(\Omega \otimes \Sigma)\Gamma_m' G_h' H_P' + (\mathrm{I}_K \otimes \Phi_h)\Sigma_P(\mathrm{I}_K \otimes \Phi_h')\right),$$

where $\Gamma_m = P_m - C_m(S_0'\Omega^{-1} \otimes \mathrm{I}_K)$.

Proposition 5 can be used to construct misspecification-robust inference for structural IRFs. The conditions on $\hat{P}$ are satisfied, for instance, by the Cholesky identification scheme (Lütkepohl, 1990). This is complementary to the robust confidence intervals for the bias proposed above and serves distinct empirical needs. The latter intervals provide a formal diagnostic tool to quantify exactly how much IRFs or forecasts are affected by dynamic misspecification, hence may become more useful during the model-building or diag-

2015). In contrast, the logic behind Proposition 4 crucially hinges on the rank condition.

nostic phase. In contrast, the former bands control for both downstream estimation error and potential dynamic misspecification, hence may be preferred when reporting the final results to policymakers.[6]

It is important to be clear about what the recentering does and does not buy relative to the more obvious benchmark, namely simply estimating the larger system. As shown at the end of the proof of Proposition 5, $\Gamma_m = \Omega^{-1} \otimes \mathrm{I}_K$. Thus, when calculating the variance of the misspecification-robust estimator by exploiting the Kronecker structure of the error variance, the resulting expression is exactly the first-order asymptotic variance derived via the delta method from an unrestricted VAR(p) model:

$$H_P G_h (\Omega^{-1} \otimes \Sigma) G_h' H_P' + (\mathrm{I}_K \otimes \Phi_h) \Sigma_P (\mathrm{I}_K \otimes \Phi_h').$$

Bias-corrected impulse responses from the restricted VAR(p_1) and ordinary impulse responses from the unrestricted VAR(p) have the same first-order influence function. Their difference is $o_p(T^{-1/2})$ under the maintained model, and so is the difference in the width of the two intervals. For the misspecification-robust bands, therefore, the recommendation to keep a small VAR and correct it is, to first order, the recommendation to estimate the VAR(p) and report it in the usual way. The benefit of correcting a parsimonious VAR is a finite sample improvement, as the simulations in Appendix F reveal. The correction further adds two practical advantages over the unrestricted VAR: it is applied to the parsimonious specification the researcher would report anyway, so the reported and the corrected object are the same; and the ingredients that produce it, $\hat{\delta}$ and $\hat{V}$, are exactly the ingredients that produce the confidence interval for the bias.

Section 7.3 and Appendix F compare these bands with those from lag-augmented local projections, which are a natural choice in a misspecified environment. Relative to direct multi-step approaches, the robust approach preserves the parametric efficiency of the VAR class, which can matter at longer horizons, where such projections are less precise. The robust interval accordingly inherits the width of an unrestricted VAR($p_1 + p_2$), which grows with p_2, whereas a lag-augmented local projection never estimates that system. The comparison ultimately reflects how large a class of departures one wishes to be protected against. Montiel Olea et al. (2026) and González-Casasús and Schorfheide (2026) study a deliberately demanding environment, in which misspecification is infinite dimensional, so that the omitted dynamics may be arbitrarily persistent. In that case, no finite lag augmentation can then bound their effect, and a VAR-based interval inherits robust coverage only as its width approaches that of a local projection. The framework here maintains a narrower class, confining the drift to p_2 lag blocks, which renders the induced distortion finite-dimensional and estimable, hence correctable. The price is that maintained window, made precise in Appendix C and illustrated for omitted moving average terms in Appendix F.

Similarly, one can perform misspecification-robust inference for forecasts based on the following result.

Proposition 6 *Let $\hat{G}_h, \hat{C}_m$ be consistent estimates of G_h, C_m, respectively. Under Assumptions 1-3,*

$$\sqrt{T}\left(\hat{y}_{T+h|T} - y_{T+h|T}\right) - \hat{G}_h \hat{C}_m \hat{\delta} \xrightarrow{d} N\left(0, G_h \Gamma_m (\Omega \otimes \Sigma) \Gamma_m' G_h'\right).$$

Proposition 6 controls for both downstream sampling noise and dynamic misspecification bias. By explicitly estimating and subtracting the dynamic bias, the procedure successfully recenters the asymptotic distribution at zero, and the modified asymptotic variance matrix reflects the additional sampling uncertainty

[6] Lohmeyer et al. (2019) propose similar confidence regions for a model-averaging estimator.

introduced by the bias correction term. The same comments as after Proposition 5 apply.

7 Monte Carlo Analysis

To evaluate the finite-sample performance of the proposed inference framework and illustrate the theoretical hazards of conventional specification pre-testing, this section conducts a series of Monte Carlo experiments. The simulation designs directly reflect the local-to-zero dynamic misspecification environment introduced in Section 3.

The primary objective of these numerical exercises is threefold: first, to demonstrate the empirical reality of the asymptotic bias embedded in underfitted dynamic models; second, to verify Proposition 1 by showing that standard Wald tests fail to provide valid size control under local dynamic misspecification; and third, to validate the proposed inference methods by examining how close the confidence intervals come to their nominal coverage in finite samples.

The simulation considers two distinct configurations: a univariate autoregressive model and a small-scale bivariate vector autoregressive model, which illustrate the framework for forecasts and for impulse responses, respectively.[7] In both environments, the true asymptotic data-generating process includes one lag, where the remaining lags are scaled by $1/\sqrt{T}$ to formalize the local-to-zero framework.[8] The practitioner, unaware of the true process generating the data, deliberately accepts a bias-variance trade-off and estimates a model with a single lag ($m = p_1 = 1$). Each experiment is simulated over 5,000 independent Monte Carlo replications. To eliminate initialization bias, a burn-in period of 100 observations is discarded from each generated sample path. The nominal significance level for hypothesis testing and confidence interval construction is set to $\alpha = 0.05$.

7.1 Implementation in Practice

Before turning to the numerical evidence, it is useful to summarize how the framework is implemented, since the steps below are exactly those carried out in the simulations of this section and in the empirical application of Section 8.

Step 1. Choose the working lag order p_1. This is the specification the researcher would report anyway, selected by whatever rule is standard for the application: a rule of thumb (four lags for quarterly data, twelve for monthly data), a conventional information criterion (AIC or BIC), or a task-based criterion targeted to the object of interest (González-Casasús and Schorfheide, 2026). No consistency property is required of the rule. As discussed in Section 3, a consistent criterion selects p_1 with probability approaching one under (1), and Section 5 shows that a non-rejected pre-test cannot certify that the discarded lags are irrelevant.

Step 2. Choose the robustness window p_2. This is the number of additional lag blocks against which the researcher wishes to be protected, and it is a deliberate user choice rather than an estimated quantity, much like the degree of lag augmentation in lag-augmented local projections. A natural default is

[7] Studying the performance in small-scale systems is important because they are more prone to suffering dynamic misspecification, as emphasized in Section 8 and illustrated in Appendix E.

[8] In the simulation code the drift is scaled by $1/\sqrt{T_p}$, with $T_p = T - p$ the effective sample of the unrestricted regression at which $\hat{\delta}$ is defined in Step 3 of Section 7.1, so that the true drift at that scale equals the design parameter (Δ or $\tilde{\delta}$ below) exactly. The difference from $1/\sqrt{T}$ is immaterial for the results.

one further year of dynamics ($p_2 = 4$ for quarterly data). Larger values guard against more distant dynamics, at a cost. In particular, the drift vector has dimension $d = K^2 p_2$, so the projection-based sets of Proposition 3 widen mechanically through the $\chi^2_{d,1-\alpha}$ quantile, the sharp intervals of Proposition 4 inherit the additional sampling uncertainty entering $\hat{w}\hat{V}\hat{w}'$, and the VAR(p) with $p = p_1 + p_2$ must remain estimable in the available sample. Reporting the results for an alternative p_2 is a cheap and informative robustness check.

Step 3. Estimate the two models. Estimate by OLS the restricted VAR(m), $m = p_1$, that is actually reported, and the unrestricted VAR(p). From the latter, collect the rescaled estimate of the last p_2 lag blocks, $\hat{\delta} = \sqrt{T}\mathrm{vec}(\hat{\Theta}_{T,p} S_0)$, and the variance estimate $\hat{V} = [S_0'\hat{\Omega}^{-1}S_0] \otimes \hat{\Sigma}$, where $\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} z_{t-1} z_{t-1}'$ and $\hat{\Sigma}$ is the residual covariance matrix.[9] Use Remark 1 to guard against violations of homoskedasticity when deemed appropriate. The two regressions are not run on the same data, since a VAR(m) conditions on m initial values, so the restricted model is estimated on $T_m = T - p_1$ usable observations and the unrestricted one on only $T_p = T - p$. The p_2 additional observations that the parsimonious specification retains are part of what makes it attractive in short samples, and they are exploited here by estimating each model on its own maximal sample and scaling every quantity by the effective sample of the regression that produces it. $\hat{\delta}$, $\hat{\Omega}$ and $\hat{\Sigma}$ are unrestricted objects computed on T_p observations, the restricted estimates and the standard interval built from them on T_m, and the misspecification-robust interval, which draws on both, scales each of its two variance components accordingly. Aligning the two regressions on the shorter sample is equally valid asymptotically, but discards precisely the finite-sample advantage that motivates the parsimonious specification.

Step 4. Form the map $\hat{w}$. The bias of the target parameter is $b = w\delta$ with $w = EG_hC_m$. Evaluate G_h at the restricted estimates using the expression in Lemma 1 for impulse responses or in Lemma 2 for forecasts, evaluate C_m at $\hat{\Omega}$, and let the selection matrix E pick the quantity of interest, such as the response of a given variable to a given shock at horizon h. For structural responses, premultiply by $\hat{H}_{\hat{P}} = (\hat{P}' \otimes \mathrm{I}_K)$ using the same identification scheme employed in the baseline analysis.

Step 5. Report the two intervals. The bias interval of Proposition 4 quantifies the distortion itself, divided by $\sqrt{T_p}$ to be read in the units of the object of interest. The misspecification-robust interval of Proposition 5, or of Proposition 6 for forecasts, recenters the baseline estimate by subtracting the estimated bias and widens it to reflect the additional sampling uncertainty of that correction.

7.2 Forecasting in a Scalar AR(2) Process

In the univariate setting ($K = 1$), the true process evolves according to:

$$y_t = B y_{t-1} + \frac{\Delta}{\sqrt{T}} y_{t-2} + u_t, \quad u_t \sim \text{i.i.d. } N(0, \sigma^2),$$

where the parameters are selected as $B = 0.5$ and $\sigma^2 = 1$. The practitioner fits an AR(1) specification and forecasts h periods ahead. Because the forecast estimand $y_{T+h|T}$ is conditional on the terminal state, all forecasts are computed from a common origin $z^* = (1, 1)'$, which keeps the target fixed across replications. The true target bias then follows from Lemma 2 as $b = w\delta$, with $w = G_hC_m$.

[9] One may alternatively estimate $\hat{\Sigma}$ from the restricted VAR, since the two are first-order equivalent.

Unlike an impulse response, the bias of a forecast depends on the state from which the forecast is made. At $h = 1$, the map is $w = Bz_1^* - z_2^*$, so the distortion vanishes at origins lying on the ray $z_2^* = Bz_1^*$ and grows as the two lags depart from it. At the chosen origin, $w = -0.5$, so that ignoring a positive second lag makes the AR(1) forecast fall short of the truth.

Figure I reports the results for $T = 250$, based on 5,000 Monte Carlo replications. The top row plots the confidence interval for the bias, averaged across replications and expressed in the units of the forecast, while the bottom row plots the forecast paths together with the standard and the misspecification-robust intervals. The wedge between the true forecast and the average AR(1) forecast is precisely the bias predicted by Lemma 2. At $(h, \Delta) = (3, 2)$, for instance, the true forecast is 0.3 while the average AR(1) forecast is 0.19, a difference of -0.11 against a predicted $b/\sqrt{T} = -0.11$.[10] The estimated bias tracks the true bias closely at every horizon, and the interval contains it.[11]

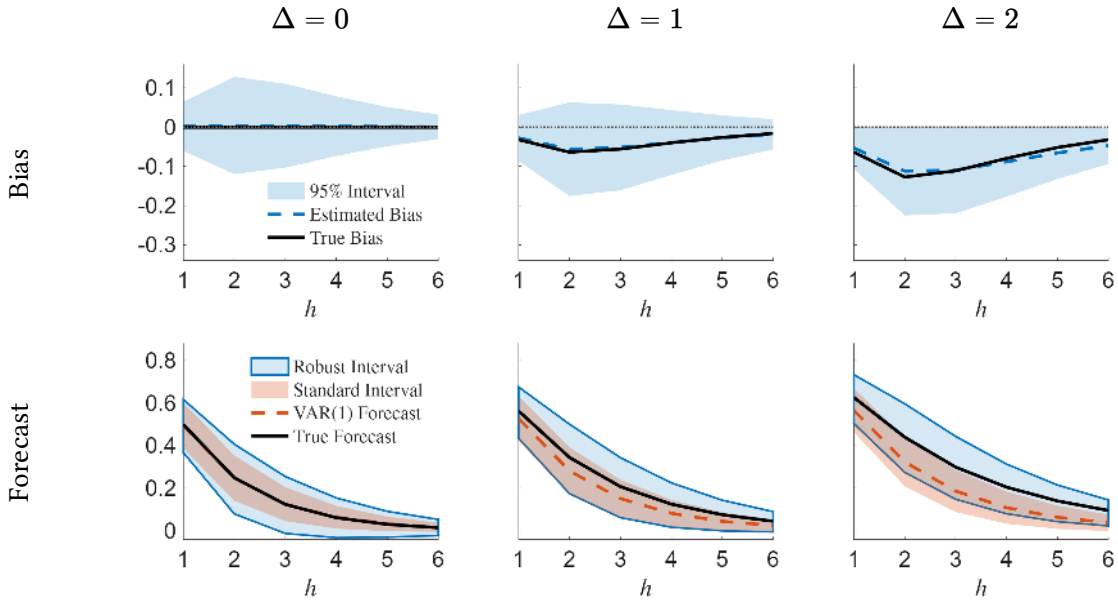


Figure I: Forecast Bias and Confidence Intervals

The bottom row shows what the bias does to inference. Under correct specification, the center of the two sets of bands essentially coincide, while the width of the robust one is slightly larger. As Δ grows, the standard bands drift away from the true forecast path, while the robust bands, which subtract the estimated bias and widen to reflect the uncertainty of that correction, continue to contain it.

Figure II makes the same point quantitatively, plotting empirical coverage against the degree of misspecification. At $\Delta = 2$, the standard interval covers 78%, 41% and 30% of the time at horizons 1, 3 and 5, respectively; the misspecification-robust interval covers 95%, 92% and 80%; and the interval for the bias covers 91%, 94% and 91%. The residual undercoverage of the robust bands at the longest horizon paired with the largest misspecification reflects the compounding of estimation error through the autoregressive system, a well-documented difficulty in small samples (Kilian, 1998, 2001). The interval for the bias, which

[10] With $B = 0.5$ and $z^* = (1, 1)'$, Lemma 2 gives $G_3 = B^2 z^{*\prime} + B(z_1^* B,\, z_1^*) + (z_1^* B^2,\, z_1^* B) = (0.75,\, 1.25)$ and $C_m = (\Omega_{12}/\Omega_{11}, -1)' = (B, -1)'$, the ratio being the first-order autocorrelation of the limit process. Hence $w = G_3 C_m = -0.875$ and $b = w\Delta = -1.75$ at the $\sqrt{T}$ scale, or -0.11 in the units of the forecast.

[11] At $\Delta = 2$, the upper end of the interval sits almost exactly at zero at every horizon. In this design $V = [S_0'\Omega^{-1}S_0] \otimes \Sigma = 1$, so the interval is $\hat{w}(\hat{\delta} \pm z_{1-\alpha/2})$ and excludes zero precisely when the Wald test rejects. Since $\Delta = 2 \approx z_{0.975} = 1.96$, that happens about half of the time, and the average interval ends at zero whatever the horizon.

is the object of primary interest, is the most stable of the three across the grid.

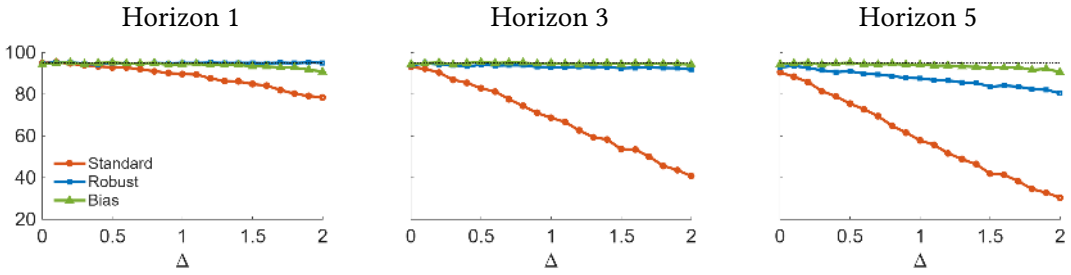


Figure II: Empirical Coverage of Forecast and Bias Intervals

Two further observations concern the sample size. First, the rejection rate of the Wald test is invariant to it, exactly as predicted by Proposition 1. Table I reports it across three sample sizes and the three degrees of misspecification. A consistent test would drive the rejection rate to one along each $\Delta \neq 0$ column as T grows, unlike what is observed in the table. At $\Delta = 1$, the test detects the misspecification in about one sample in six however long the sample is, and at $\Delta = 2$, which is roughly the point at which the drift equals two standard errors, in about one in two.

	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$
$T = 100$	5.1	15.2	46.2
$T = 250$	5.6	15.2	49.7
$T = 500$	4.9	16.0	50.6

Table I: Rejection Rates of the Wald Pre-Test

Second, the average length of the bias interval at the $\sqrt{T}$ scale is roughly constant at the three sample sizes considered, so in the units of the forecast the interval contracts at the parametric rate $O(T^{-1/2})$ while retaining its coverage. For instance, at $h = 1$ and $\Delta = 0$, it is 1.99, 1.97 and 1.97, respectively.[12]

It is worth being clear about what the interval for the bias does and does not deliver in this particular case. Because both the drift and the target are scalars in this design ($q = d = 1$), the interval excludes zero exactly when the Wald statistic rejects. What it adds is a magnitude. For instance, at $(h, \Delta) = (3, 2)$ the average bias interval runs from -0.219 to 0.001 in the units of the forecast, against a true forecast of 0.3. Thus, we can conclude that the AR(1) forecast may understate the truth by as much as three quarters of its own size. That bound is available precisely where the pre-test is uninformative, and no pre-test delivers it.

Overall, these findings demonstrate that while the local drift parameter Δ is inherently unestimable, the proposed robust methods provide researchers with a computationally simple and size-correct mechanism to bound the exact extent of misspecification bias in their forecasts.

Figure III illustrates the lag-truncation problem motivated in this paper from the lens of two popular information criteria. The figure plots the lag order selection frequencies based on the AIC and BIC for different degrees of misspecification, across the Monte Carlo repetitions. The results highlight several important observations. First, as it is common knowledge, we see that BIC selection is always more parsimonious

[12] The same holds over the whole grid of horizons and degrees of misspecification. Unreported results confirm that, for $\Delta \neq 0$, the average interval length varies by at most 20% at the $\sqrt{T}$ scale across the three sample sizes, since the plug-in $\hat{w}$ is evaluated at the biased AR(1) estimate and converges to w only as T grows. At $h = 1$, for instance, $|\hat{w}| = |\hat{B} - 1|$ shrinks when $\hat{B}$ is biased upward. In the units of the forecast, the ratio of the average length at $T = 100$ to that at $T = 500$ ranges from 1.9 to 2.7 across all horizons and degrees of misspecification considered, against the predicted $\sqrt{500/100} \approx 2.2$ for the parametric rate.

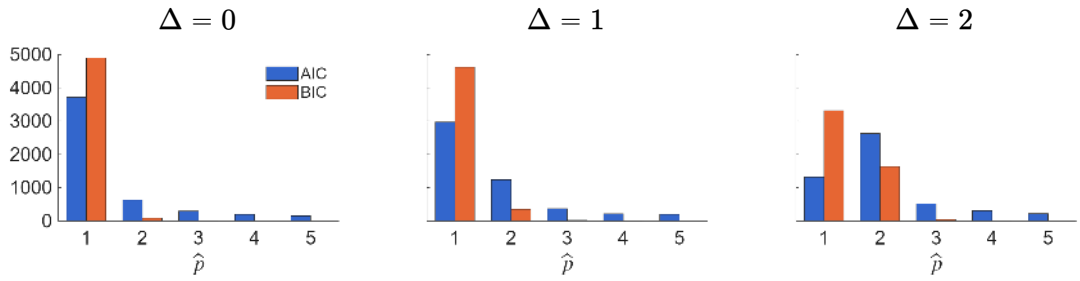


Figure III: AIC & BIC selection

than AIC selection, regardless of the degree of misspecification (e.g., Lütkepohl, 2005). When there is no misspecification ($\Delta = 0$), both criteria mostly pick the true (finite-sample and asymptotic) order $p = 1$. Under moderate misspecification ($\Delta = 1$), BIC remains largely favorable of a parsimonious AR(1). AIC still favors a more parsimonious AR(1) specification, but seems equally supportive of a (true) AR(2) model a non-negligible fraction of times. Under large misspecification ($\Delta = 2$), both methods pick both AR(1) and AR(2) specifications for an important number of series, with BIC favoring the more parsimonious specification and AIC often preferring the richer specification. These results illustrate that information criteria yield mixed results in practice, as they not unanimously pick the true finite-sample or asymptotic lag order of the process. Instead, these criteria yield a sample-dependent selection which has important consequences on the amount of dynamic misspecification that remains for post-selection inference.

7.3 Impulse Responses in a Bivariate VAR(3) Process

To assess the framework in a multivariate context, I simulate a bivariate process ($K = 2$) defined by:[13]

$$y_t = By_{t-1} + \frac{\tilde{\delta}}{\sqrt{T}}\begin{bmatrix} 0.4 & -0.2 \\ 0.2 & 0.3 \end{bmatrix} y_{t-2} + \frac{\tilde{\delta}}{\sqrt{T}}\begin{bmatrix} 0.2 & 0.1 \\ -0.1 & 0.2 \end{bmatrix} y_{t-3} + u_t, \quad u_t \sim \text{i.i.d. } N(0, \Sigma),$$

where the structural parameter matrices are selected as follows:

$$B = \begin{bmatrix} 0.6 & 0.2 \\ 0.1 & 0.7 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

The practitioner underfits the system by estimating a restricted VAR(1) model. In this design, the degree of misspecification is driven by the scalar $\tilde{\delta}$. The practitioner, unaware of this special structure, must allow for a bias drift of full dimension $d = K^2 p_2 = 8$. The quantity of interest is defined as the impulse response of variable 1 to a one-unit reduced-form shock in variable 1 at horizon h, isolated using the selection vector $e = (1, 0, 0, 0)'$ mapped to the vectorized companion matrix formulation, and again simplifies to a scalar target, making inference straightforward per Proposition 4.

Figure IV reveals that the confidence intervals for the bias have very good coverage properties for all degrees of misspecification. At the shortest horizon, the coverage stays within about two points of the nominal level (black dotted line) at every sample size, the minima being 92.8%, 94% and 94.1% at $T = 100$, 250 and 500. At longer horizons, finite-sample distortions appear, and they are asymmetric in the direction of misspecification because positive and negative $\tilde{\delta}$ imply different persistence in the data. For instance, at

[13] The highest eigenvalue modulus of the companion matrix across the entire parameter configurations is 0.97, which occurs exactly at the boundaries $\tilde{\delta} = 3$ and $T = 100$. Such a value matches that found in the empirical application in Section 8.

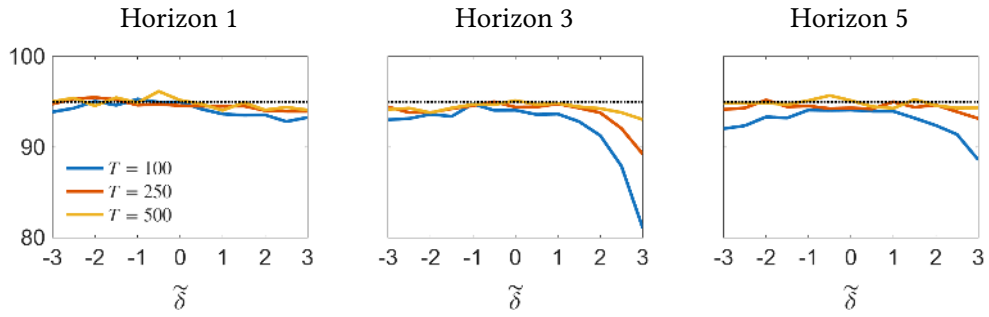


Figure IV: Coverage Frequencies of Bias Confidence Intervals

$(h, T) = (3, 100)$, the coverage falls to 81.1% at $\tilde{\delta} = 3$ while remaining above 92% at $\tilde{\delta} = -3$. These distortions are remedied as the sample size grows, to 89.2% at $T = 250$ and 93% at $T = 500$.

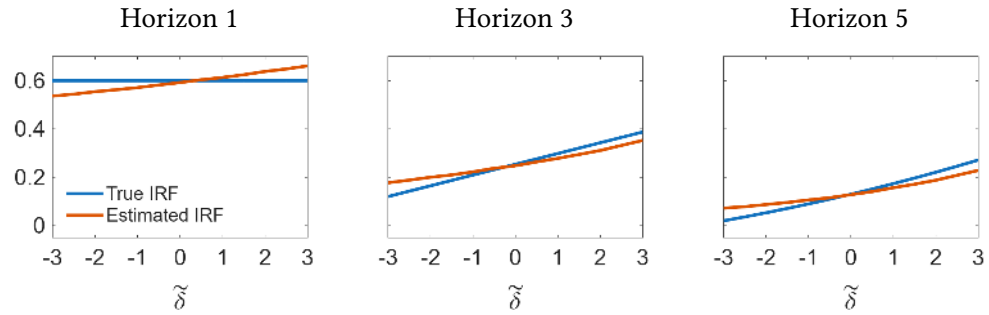


Figure V: True versus Average Estimated IRFs, $T = 250$

Figure V reports the true IRF against the average estimated IRF across different degrees of misspecification, fixing the sample size to $T = 250$. The difference between the red and blue lines is the measure of average dynamic misspecification bias. As expected, the lines always cross at $\tilde{\delta} = 0$ because under no misspecification the bias is zero, hence the true and estimated IRFs should agree. As $\tilde{\delta} \neq 0$, the two lines become increasingly more distant, revealing that the bias grows with the degree of misspecification. It does not, however, grow monotonically in the horizon, nor does it keep its sign: at $h = 1$ the true response is invariant to $\tilde{\delta}$ and the estimate overstates it for $\tilde{\delta} > 0$, whereas at $h = 3$ and $h = 5$ the true response moves with $\tilde{\delta}$ faster than the estimate does and the sign of the distortion reverses. Because the true degree of misspecification is unknown in practice, and its consequences are not signable by inspection, this figure emphasizes the need for misspecification-robust methods as derived in this paper.

The same design supports a direct comparison with lag-augmented local projections, the leading alternative for inference on impulse responses that is robust to dynamic misspecification (Montiel Olea et al., 2026). Four intervals for the same target are compared at $T = 250$: the standard delta-method interval from the restricted VAR(1); the misspecification-robust interval (MR-VAR) of Proposition 5, specialized to reduced-form responses so that the Σ_P term is absent; the lag-augmented local projection (LA-LP) of $y_{1,t+h}$ on y_t, controlling for two further lags and using Eicker-White standard errors (Montiel Olea and Plagborg-Møller, 2021); and the delta-method interval from the unrestricted VAR(3). Figure VI reports empirical coverage and average length over the $\tilde{\delta}$ grid.[14]

[14]Augmenting a third lag in the local projection leaves the results unchanged, as expected from González-Casasús and Schorfheide (2026) and Montiel Olea et al. (2026). Appendix F compares the same four procedures, at the same nominal level and for the same target, under omitted moving average dynamics.

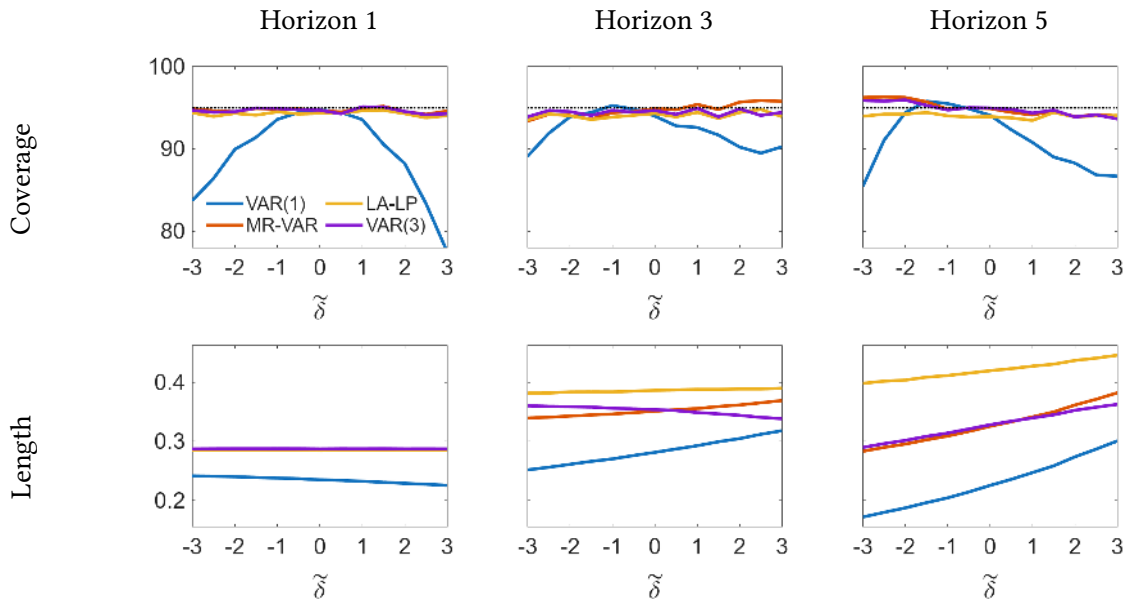


Figure VI: Standard, Misspecification-Robust, Lag-Augmented LP and Unrestricted VAR Inference, $T = 250$

The top row confirms that both robust procedures perform as expected. The standard VAR(1) interval undercovers as $|\tilde{\delta}|$ grows, at every horizon and asymmetrically in the direction of the drift: its coverage falls from 95% to 78% at $h = 1$, to 89% at $h = 3$ and to 85% at $h = 5$. The MR-VAR and LA-LP intervals stay within about two points of the nominal level across the entire grid, as does the unrestricted VAR(3).

The bottom row is where the two robust procedures separate. Sidestepping the bias by projecting directly onto the horizon of interest is expensive at distant horizons, and the LA-LP intervals inherit that cost. Averaging over the $\tilde{\delta}$ grid, the MR-VAR intervals are indistinguishable from LA-LP ones at $h = 1$, about 9% shorter at $h = 3$ and about 22% shorter at $h = 5$.[15] The MR-VAR interval instead widens the standard interval by only what the bias correction costs, and its width tracks that of the unrestricted VAR(3) almost exactly, as the first-order equivalence discussed in Section 6.4 anticipates. This is the sense in which the correction attains the robustness of a direct projection while retaining the parametric efficiency of the VAR class. The price, as Section 6 emphasized, is the maintained window p_2. Appendix F shows that the same pattern survives when the omitted dynamics are moving average rather than autoregressive, so that no finite p_2 captures them exactly.

8 Monetary Policy Transmission and the Price Puzzle

This section applies the proposed framework to a classic question in empirical macroeconomics: the structural transmission of monetary policy shocks. The setting is a natural laboratory for the problem studied in this paper. Quarterly macroeconomic samples are short, particularly when the analysis is confined to a structurally stable regime, such as the post-Volcker or post-Great Moderation eras, so applied researchers face strong incentives to estimate parsimonious models, routinely excluding higher-order lag blocks on the

[15] One might ask whether that gap is partly an artifact of the standard errors rather than of the procedures, since the local projection uses Eicker-White errors while the MR-VAR uses the homoskedasticity-only $\hat{V} = [S_0'\hat{\Omega}^{-1}S_0] \otimes \hat{\Sigma}$, in a design with homoskedastic Gaussian innovations. The results themselves answer it. At $h = 1$, the two intervals agree to within 0.9% of each other's length, so the Eicker-White correction imposes no systematic penalty in this design. The gap at $h = 3$ and $h = 5$ is therefore due to the multi-step projection variance and not the choice of variance formula. The same reading applies to the comparison in Appendix F, where the two are again indistinguishable at the horizons at which the projection is nearly one-step.

basis of information criteria or sequential Wald tests (Sims, 1980; Ivanov and Kilian, 2005; Lütkepohl, 2005).

In the macroeconometric literature, the price puzzle refers to the empirical finding that monetary policy shocks have surprisingly large and persistent effects on the price level, much larger than standard macroeconomic models predict (e.g., Sims, 1992; Eichenbaum, 1992; Christiano, Eichenbaum and Evans, 1999). Resolving the price puzzle has historically required expanding the cross-sectional dimension of the information set. This has been typically achieved either by introducing explicit forward-looking commodity price indices (Sims, 1992) or by moving to high-dimensional Factor-Augmented VARs that extract unobserved informational states from large panels of macroeconomic data (Bernanke, Boivin and Eliasz, 2005); see Ramey (2016) for a survey of the identification strategies that have since become standard, and Barth III and Ramey (2001); Hanson (2004); Castelnuovo and Surico (2010) for accounts that locate the puzzle outside the information set. While empirically effective, the proposed solutions often come at a steep econometric cost, as they induce severe parameter proliferation, exhaust finite-sample degrees of freedom, and often rely on sequential empirical experimentation, raising concerns over specification search biases and the selective reporting of successful ex post fixes (Stevenson, 2026).

The aforementioned cross-sectional expansions are rooted in the historical interpretation of the price puzzle as fundamentally an omitted variable bias. As originally argued by Sims (1992), small-scale VARs lacking explicit forward-looking informational indicators fail to fully capture the central bank's real-time information set, thereby conflating exogenous monetary shocks with endogenous responses to anticipated inflation. However, from a time-series perspective, this cross-sectional informational deficiency has a direct dynamic manifestation. By standard marginalization arguments, omitting a relevant state variable from a finite-order VAR transforms the data-generating process of the retained variables into an infinite-order VAR representation; see Appendix E for details. Therefore, the traditional omitted variable problem manifests within the small-scale reduced-form system precisely as a problem of omitted high-order dynamics. The question addressed below is how much of the price puzzle survives once the dynamic misspecification of a parsimonious, small-scale VAR is explicitly bounded and purged from structural inference, without expanding the cross-sectional information set.

8.1 Data, Model Specification, and the Pre-Testing Dilemma

I utilize a standard quarterly macroeconomic dataset spanning the pre-Volcker era from 1959:Q1 through 1979:Q4, aligned with the samples used in the early discussions around the price puzzle, and downloaded from the FRED database. The data consists of three foundational variables:

$$y_t = \begin{bmatrix} 100 \times [\ln(\text{Real GDP}_t) - \ln(\text{Real GDP}_{t-1})] \\ 100 \times [\ln(\text{GDP Deflator}_t) - \ln(\text{GDP Deflator}_{t-1})] \\ \text{Federal Funds Rate}_t \end{bmatrix}, \tag{3}$$

representing economic growth, inflation, and the monetary policy instrument, respectively. To keep the notation consistent with the theoretical results, the data are demeaned prior to estimation.

A researcher seeking to trace the dynamic impact of a monetary policy shock estimates a baseline VAR. To conserve degrees of freedom, the researcher must choose an optimal lag length. I replicate a standard applied paradigm by conducting a sequential specification pre-test. First, I employ the AIC criterion as a lag order selection tool as routinely done in the literature, which recommends including 4 lags. Next, I estimate

a VAR(8) specification and test the null hypothesis that the fifth through eighth lag blocks are exactly zero. A nesting order of $p = 8$ is conventional at quarterly frequency (see, e.g., Sims, 1980; Stock and Watson, 2002), since an additional one-year window captures annual cyclical dynamics, policy transmission lags, and residual seasonality or smoothing artifacts embedded in seasonally adjusted data. The standard Wald test statistic for this joint restriction yields a p-value of 0.61, so the null is not rejected at any conventional significance level.[16]

Following conventional pre-testing practice, the researcher would conclude that a parsimonious VAR(4) specification is fully sufficient, drop the fifth through eighth lag blocks entirely, and proceed to conduct structural inference using the restricted model.[17] Four lags was also the benchmark choice before the resolution of the puzzle (Christiano et al., 1999).

Despite the pre-testing evidence, I evaluate the restricted VAR(4) against an unrestricted VAR(8). The restricted model is appealing as it retains 80 effective observations against the 76 of the unrestricted model, and $K^2 p_2 = 36$ fewer autoregressive parameters. However, Proposition 1 implies that the non-rejection cannot certify that the discarded blocks are irrelevant, so treating the VAR(8) as the unrestricted benchmark is what allows structural inference to be robustified against the omissions the diagnostics fail to detect. The more comfortable the non-rejection, as in this case with a large p-value, the more the subsequent analysis rests on a restriction against which the data have little power, and the more is gained by attaching a magnitude to that restriction rather than a binary verdict.

Structural identification is achieved via a standard Cholesky decomposition with the ordering given by y_t, treating the Federal Funds Rate as the chronologically latest-moving policy variable. The policy shock is normalized to raise the Federal Funds Rate by 100 basis points on impact.

8.2 Applying the Robust Inference Framework

I operationalize the proposed framework to evaluate the robust validity of the inflation and output IRFs deduced from the VAR(4) model. Rather than conditioning inference on the problematic assumption that the fifth through eighth lag blocks are exactly zero, I treat the omitted lag block as an unestimable local drift parameter:

$$\Delta = \sqrt{T}\begin{bmatrix} B_5 & \cdots & B_8 \end{bmatrix}, \qquad p_1 = 4,\ p_2 = 4,\ d = K^2 p_2 = 36.$$

The scalar quantities of interest are the impulse responses of inflation and output to a policy shock at horizons in the following 3 years. For each, I construct the confidence interval for the misspecification

[16] The size of that number is not innocuous, and the reason is worth noting. The statistic is an asymptotic χ^2_{36} evaluated on 76 effective observations, and the unrestricted VAR(8) fits $Kp = 24$ regressors per equation, so the maximum-likelihood estimator $\hat{U}'\hat{U}/T_p$ understates Σ by roughly $1 - 52/76 \approx 32\%$ and inflates the Wald statistic by the reciprocal factor. Dividing instead by the residual degrees of freedom, $T_p - Kp = 52$, as is conventional for VAR inference in short samples (e.g., Lütkepohl, 2005) and exactly unbiased for Σ under Gaussian errors, moves the statistic from 48.4 to 33.1 and the p-value from 0.08 to 0.61. Thus, without the correction the same data would have looked like marginal evidence against the restriction. All the quantities reported below use the corrected estimator. The takeaway is that whenever the number of restrictions is large relative to the effective sample, as it is in any application of this framework with a short sample and a wide robustness window, the choice of divisor can reverse the reading of the pre-test.

[17] In practice, most researchers would typically perform only one of the two selection steps. Here, I employ this two-step approach to emphasize the dangers of post-model-selection. Although formally this is a sequential test with different size, it captures the prevailing practice of treating the lag order selected by information criteria as the baseline order for the analysis. BIC, for instance, recommends including only 1 lag.

bias from the pivotal distribution of Proposition 4, which maps the sampling uncertainty of the low-order VAR parameters and the data covariance structure into a projection-free bound without requiring Δ to be consistently estimated. I also compute the standard and misspecification-robust intervals for the IRFs. Table II reports these results at the 90% level, and Figure VII shades the 68%, 90% and 95% bias intervals.[18]

Table II: Robust IRF Inference for Monetary Transmission

Horizon	Estimated IRF	Bias CI	Standard CI	Robust CI
Inflation Response to FFR Shock				
2	0.103	[0.001, 0.067]	[0.060, 0.145]	[0.013, 0.124]
4	0.060	[-0.027, 0.069]	[0.008, 0.112]	[-0.033, 0.112]
8	0.043	[0.000, 0.132]	[-0.033, 0.119]	[-0.126, 0.080]
12	0.015	[0.001, 0.180]	[-0.061, 0.090]	[-0.195, 0.043]
Output Response to FFR Shock				
2	-0.195	[-0.081, 0.081]	[-0.297,-0.093]	[-0.325, -0.064]
4	-0.146	[-0.079, 0.124]	[-0.242,-0.051]	[-0.309, -0.029]
8	-0.056	[-0.132, 0.048]	[-0.157, 0.046]	[-0.149, 0.122]
12	0.034	[-0.123, 0.024]	[-0.060, 0.127]	[-0.037, 0.204]

Note: Responses are to a contractionary monetary policy shock normalized to raise the Federal Funds Rate by 100 basis points on impact. The nominal level is $\alpha = 0.1$ throughout. Bias CI reports the interval of Proposition 4 for the asymptotic bias b, in the units of the impulse response; Standard CI the delta method interval for the structural IRF from the VAR(p_1); and Robust CI the interval of Proposition 5.

The estimated IRFs replicate well-known anomalies found in the small-scale VAR literature. Following a contractionary shock to the FFR, the point estimates for the inflation response are positive across all horizons. This is the classic price puzzle, where a monetary tightening appears to induce a persistent increase in the price level. In contrast, the output response is appropriately negative at shorter horizons.

Under the standard inference approach, which ignores potential dynamic misspecification, an applied researcher would observe that these anomalous positive responses are statistically insignificant, as the intervals consistently contain zero at longer horizons; see the fourth column in Table II or the blue shaded region in the bottom panels of Figure VII. Standard practice would likely conclude that the small-scale model simply lacks the statistical power or informational capacity to identify a clear contractionary transmission mechanism.

The central object is the asymptotic bias interval in the third column of Table II (or, analogously, the normalized top row in Figure VII), which states, in the units of the impulse response, how distorted the reported VAR(4) response may be due to omitted dynamics. For output, the interval includes zero at every horizon: the estimated distortion is small relative to its standard error, and dynamic misspecification does not generate a statistically detectable bias in the output response. For inflation, the pattern differs. The interval is strictly positive at $h = 2$, and again at $h = 8$, $h = 11$ and $h = 12$.[19] The distortion is not static, but growing with the horizon as the omitted lag effects accumulate through the companion matrix (the point estimate of the bias is positive throughout and rises from 0.005 at $h = 1$ to 0.091 at $h = 12$), and its sign is estimated precisely enough to be resolved at several horizons: the VAR(4) overstates the price response.

[18] The 90% shaded region is that reported in the table. However, the figure is reported at the $\sqrt{T_p}$-scale on which $\hat{\delta}$ is defined, which multiplies the Bias CI column in the table by $\sqrt{T_p} = \sqrt{76}$.

[19] The lower bound of the inflation bias interval at $h = 8$ is positive yet rounds to 0.000 in Table II.

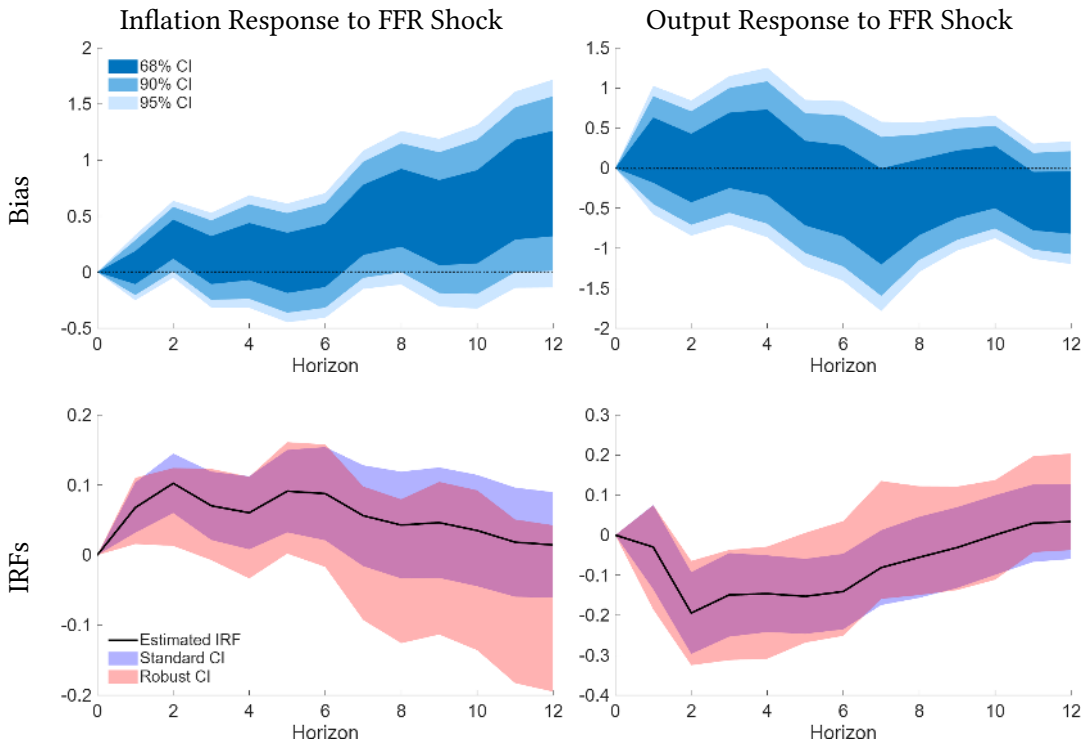


Figure VII: Inference on IRFs to a FFR shock

The final column in Table II nets that bias out; see, equivalently the red shaded region in Figure VII. As Section 6.4 showed, the variance of the misspecification-robust estimator coincides to first order with the delta method variance of an unrestricted VAR($p_1 + p_2$), so the interval is about as wide as a VAR(8) interval would have been, up to small order terms, and the correction moves its center. That shift is precisely what changes conclusions. Across the full path plotted in Figure VII, the standard interval declares a significantly positive price response at $h = 1$ through $h = 6$, whereas the robust interval does so at $h = 1$, $h = 2$ and, marginally, at $h = 5$. Three horizons therefore change verdict, at $h = 3$, $h = 4$ and $h = 6$, all in the same direction and all at the short-to-medium horizons at which the puzzle is usually diagnosed. At $h = 8$ and $h = 12$, the standard interval already contained zero, so the correction rules out nothing there that had been established. What it adds at those horizons is the magnitude in the bias column.

The empirical evidence suggests that dynamic misspecification accounts for a significant share of the price puzzle at short-to-medium horizons. Full resolution of the price puzzle often displays a stronger negative reaction of prices, especially upon impact. By construction, the present framework targets the downstream VAR-implied dynamics, hence is inherently limited to resolve impact distortions stemming from the contemporaneous identification scheme. The evidence presented then points to a dual structure in which the historical price puzzle is an artifact of both dynamic underfitting and residual omitted variable bias.[20]

These results carry a practical implication for applied work. Rather than over-parameterizing a model to approach an unattainable ideal of correct specification, or dropping higher-order lags after a non-significant pre-test and thereby equating a non-rejected null with a zero parameter, the researcher can retain the parsimonious system and report how much its implied dynamics may be distorted. Here that turns the price puzzle from a binary verdict into a decomposition, revealing that a measurable part of the anomaly is

[20] An alternative concern could be that $p_2 = 4$ is not large enough to capture the whole extent of true dynamic misspecification. I have explored the alternative choice $p_2 = 8$, which adds an extra year of dynamics, and the conclusions remain the same.

attributable to lag truncation. More generally, reporting robust intervals alongside the baseline responses allows a researcher to delineate which structural conclusions withstand potential dynamic misspecification and which are artifacts of underfitting, while retaining a parsimonious VAR model.

9 Conclusion

This paper takes the lag truncation problem in VARs as a fact of applied practice rather than an error to be corrected. Empirical researchers select parsimonious lag structures by information criteria or sequential pre-tests, implicitly treating the omitted higher-order dynamics as strictly zero, while theoretical macroeconomic models routinely imply long or infinite autoregressive structures whose distant coefficients are small yet collectively non-negligible. Formalizing this environment through a local-to-zero drifting parameter framework, I show that conventional pre-tests cannot safeguard downstream structural inference, yet what they fail to detect can nonetheless be quantified via valid confidence sets for the bias induced in IRFs or forecasts.

For applied work, this changes what a parsimonious VAR can be asked to report. Alongside the baseline IRFs or forecasts, the researcher can report an interval for the direction and magnitude of the distortion those responses may carry and as a byproduct, when presenting the final results, bands that net the estimated bias out. The exercise is a diagnostic rather than a specification search, saying how much of a given conclusion survives the dynamics that the model leaves out.

The value of the diagnostic is illustrated on the classic monetary policy transmission mechanism. In a small-scale VAR whose lag order is endorsed by both the AIC and a Wald pre-test, the bias in the price response to a contractionary policy shock is found to be significantly positive at several horizons, and its magnitude is a substantial fraction of the estimated response. The evidence suggests that a quantifiable share of the measured puzzle is due to dynamic misspecification, without augmenting the cross-sectional information set.

Two connections are worth noting. First, lag-augmented local projections are a natural choice for IRF inference under dynamic misspecification, and Montiel Olea et al. (2026) show that the degree of misspecification pins down the optimal way to combine local projections and VAR estimates. The framework developed here supplies a data-driven measure of that degree, hence a way to defend a multi-step inferential strategy, but also a way to remain within the VAR class with the bias made explicit. Second, Bayesian shrinkage is not a substitute for the proposed robustness approach. González-Casasús and Schorfheide (2026) stress that Bayesian lag order selection based on the marginal likelihood does not capture the systematic bias induced by dynamic misspecification, so that even highly regularized Bayesian VARs remain exposed to the distortions characterized in this paper. Because the diagnostic proposed here relies only on VAR-implied dynamics, it is independent of the prior specification and may equally be paired with a Bayesian approach.

References

Barth III, M. J. and Ramey, V. A. (2001), 'The cost channel of monetary transmission', *NBER Macroeconomics Annual* **16**, 199–240.

Baumeister, C. (2025), Discussion of" Local Projections or VARs? A Primer for Macroeconomists" by José

Luis Montiel Olea, Mikkel Plagborg-Møller, Eric Qian, and Christian K, Technical report, Wolf, Working paper, Prepared for the NBER Macroeconomics Annual.

Bernanke, B. S. (1983), 'Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression.', *American Economic Review* **73**(3).

Bernanke, B. S., Boivin, J. and Eliasz, P. (2005), 'Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach', *The Quarterly journal of economics* **120**(1), 387–422.

Blanchard, O. J. and Quah, D. (1993), 'The dynamic effects of aggregate demand and supply disturbances: Reply', *The American Economic Review* **83**(3), 653–658.

Castelnuovo, E. and Surico, P. (2010), 'Monetary policy, inflation expectations and the price puzzle', *The Economic Journal* **120**(549), 1262–1283.

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (1999), 'Monetary policy shocks: What have we learned and to what end?', *Handbook of macroeconomics* **1**, 65–148.

Claeskens, G. and Hjort, N. L. (2003), 'The focused information criterion', *Journal of the American Statistical Association* **98**(464), 900–916.

Das Gupta, S. and Perlman, M. D. (1974), 'Power of the noncentral F-test: Effect of additional variates on Hotelling's T^2-test', *Journal of the American Statistical Association* **69**(345), 174–180.

De Graeve, F. and Westermark, A. (2025), 'Long-lag VARs', *Journal of Monetary Economics* p. 103831.

Del Negro, M. and Schorfheide, F. (2004), 'Priors from general equilibrium models for VARs', *International Economic Review* **45**(2), 643–673.

Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R. (2007), 'On the fit of new Keynesian models', *Journal of Business & Economic Statistics* **25**(2), 123–143.

Eichenbaum, M. (1992), 'Comments on 'Interpreting the Time Series Facts: The Effects of Monetary Policy'by Christopher Sims', *European Economic Review* **36**(5), 1001–1011.

Evans, G. W. and Honkapohja, S. (2001), *Learning and expectations in macroeconomics*, Princeton University Press.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J. and Watson, M. W. (2007), 'ABCs (and Ds) of understanding VARs', *American economic review* **97**(3), 1021–1026.

Galí, J. (1999), 'Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations?', *American economic review* **89**(1), 249–271.

Gánics, G., Inoue, A. and Rossi, B. (2019), 'Confidence intervals for bias and size distortion in IV and local projections-IV models', *Journal of Business & Economic Statistics* pp. 1–18.

Giannone, D., Lenza, M. and Primiceri, G. E. (2015), 'Prior selection for vector autoregressions', *Review of Economics and Statistics* **97**(2), 436–451.

González-Casasús, O. and Schorfheide, F. (2026), Misspecification-robust shrinkage and selection for VAR forecasts and IRFs, Technical report, National Bureau of Economic Research.

Hanson, M. S. (2004), 'The "price puzzle" reconsidered', *Journal of Monetary Economics* **51**(7), 1385–1413.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica* **46**(6), 1251–1271.

Holly, A. (1982), 'The hausman specification test', *Econometrica* **50**(3), 749–759.

Inoue, A., Kuo, C.-H. and Rossi, B. (2020), 'Identifying the sources of model misspecification', *Journal of Monetary Economics* **110**, 1–18.

Ivanov, V. and Kilian, L. (2005), 'A practitioner's guide to lag order selection for VAR impulse response analysis', *Studies in Nonlinear Dynamics & Econometrics* **9**(1).

Jordà, Ò. (2005), 'Estimation and inference of impulse responses by local projections', *American economic review* **95**(1), 161–182.

Kilian, L. (1998), 'Small-sample confidence intervals for impulse response functions', *Review of economics and statistics* **80**(2), 218–230.

Kilian, L. (2001), 'Impulse response analysis in vector autoregressions with unknown lag order', *Journal of Forecasting* **20**(3), 161–179.

Leeb, H. and Pötscher, B. M. (2005), 'Model selection and inference: Facts and fiction', *Econometric Theory* **21**(1), 21–59.

Leeb, H. and Pötscher, B. M. (2008), 'Sparse estimators and the oracle property, or the return of Hodges' estimator', *Journal of Econometrics* **142**(1), 201–211.

Lewis, R. and Reinsel, G. C. (1985), 'Prediction of multivariate time series by autoregressive model fitting', *Journal of multivariate analysis* **16**(3), 393–411.

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

Lohmeyer, J., Palm, F., Reuvers, H. and Urbain, J.-P. (2019), 'Focused information criterion for locally misspecified vector autoregressive models', *Econometric Reviews* **38**(7), 763–792.

Ludwig, J. F. (2024), 'Local projections are var predictions of increasing order', *Available at SSRN 4882149* .

Lütkepohl, H. (1985), 'Comparison of criteria for estimating the order of a vector autoregressive process', *Journal of time series analysis* **6**(1), 35–52.

Lütkepohl, H. (1990), 'Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models', *The review of economics and statistics* pp. 116–125.

Lütkepohl, H. (2005), *New Introduction to multiple time series analysis*, Springer Berlin, Heidelberg.

Lütkepohl, H., Staszewska-Bystrova, A. and Winker, P. (2015), 'Comparison of methods for constructing joint confidence bands for impulse response functions', *International Journal of Forecasting* **31**(3), 782–798.

Milani, F. (2007), 'Expectations, learning and macroeconomic persistence', *Journal of monetary Economics* **54**(7), 2065–2082.

Montiel Olea, J. L. and Plagborg-Møller, M. (2019), 'Simultaneous confidence bands: Theory, implementation, and an application to svars', *Journal of Applied Econometrics* **34**(1), 1–17.

Montiel Olea, J. L. and Plagborg-Møller, M. (2021), 'Local projection inference is simpler and more robust than you think', *Econometrica* **89**(4), 1789–1823.

Montiel Olea, J. L., Plagborg-Møller, M., Qian, E. and Wolf, C. K. (2025), Local projections or VARs? a primer for macroeconomists, Technical report, National Bureau of Economic Research.

Montiel Olea, J. L., Plagborg-Møller, M., Qian, E. and Wolf, C. K. (2026), 'Double robustness of local projections and some unpleasant varithmetic', *Econometrica* **94**(4), 1313–1343.

Montiel Olea, J. L., Stock, J. H. and Watson, M. W. (2021), 'Inference in structural vector autoregressions identified with an external instrument', *Journal of Econometrics* **225**(1), 74–87.

Plagborg-Møller, M. and Wolf, C. K. (2021), 'Local projections and VARs estimate the same impulse responses', *Econometrica* **89**(2), 955–980.

Ramey, V. A. (2016), Macroeconomic shocks and their propagation, *in* 'Handbook of Macroeconomics', Vol. 2, Elsevier, pp. 71–162.

Ravenna, F. (2007), 'Vector autoregressions and reduced form representations of DSGE models', *Journal of monetary economics* **54**(7), 2048–2064.

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

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica: journal of the Econometric Society* pp. 1–48.

Sims, C. A. (1992), 'Interpreting the macroeconomic time series facts: The effects of monetary policy', *European economic review* **36**(5), 975–1000.

Stevenson, M. T. (2026), 'Humans in the Loop: The Next Frontier in the Credibility Revolution', *Journal of Economic Literature* . Forthcoming.

Stock, J. H. and Watson, M. W. (2001), 'Vector autoregressions', *Journal of Economic perspectives* **15**(4), 101–115.

Stock, J. H. and Watson, M. W. (2002), 'Has the business cycle changed and why?', *NBER macroeconomics annual* **17**, 159–218.

Weiss, A. A. (1991), 'Multi-step estimation and forecasting in dynamic models', *Journal of Econometrics* **48**(1-2), 135–149.

Appendix

Quantifying Dynamic Misspecification Biases in VARs

Oriol González-Casasús

A Proofs **33**

B Illustrative Example **38**

C Extension to VAR(∞) **41**

D Alternative Identification Schemes **42**

D.1 Identification via Sign Restrictions 43

D.2 Identification via External Proxies 43

E Omitted Variables As Moving Average Terms **44**

E.1 Analytical Explanation 44

E.2 Numerical Illustration 46

F Performance Under Other Forms of Dynamic Misspecification **47**

F.1 Few Omitted MA Terms 49

F.2 Many Omitted MA Terms 51

A Proofs

Proof of Lemma 3. Multiplying equation (2) by $\sqrt{T}$ on both sides, we get

$$\sqrt{T}(\hat{\Theta}_{T,m} - \Theta_{T,m}) = \sqrt{T}C_T \left(\mathrm{I}_{Kp_2} - \Pi_m'\Pi_m\right) S_0' \left(\frac{1}{T}ZZ'\right) S_m \left[S_m' \left(\frac{1}{T}ZZ'\right) S_m\right]^{-1} + \left(\frac{1}{\sqrt{T}}UZ'\right) S_m \left[S_m' \left(\frac{1}{T}ZZ'\right) S_m\right]^{-1}.$$

Let $m = p$. Then $S_m = \mathrm{I}_{Kp}$ and $\Pi_m' = \mathrm{I}_{Kp_2}$, hence $\Pi_m'\Pi_m = \mathrm{I}_{Kp_2}$. It thus follows that

$$\sqrt{T}(\hat{\Theta}_{T,p} - \Theta_{T,p}) = \left(\frac{1}{\sqrt{T}}UZ'\right)\left(\frac{1}{T}ZZ'\right)^{-1}.$$

Post-multiplying by S_0 on both sides and applying the vectorization operator yields

$$\sqrt{T}(\mathrm{vec}(\hat{\Theta}_{T,p}S_0) - \mathrm{vec}(\Theta_{T,p}S_0)) = \left(S_0'\left(\frac{1}{T}ZZ'\right)^{-1} \otimes \mathrm{I}_K\right) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathrm{vec}(u_t z_{t-1}').$$

Because $\frac{1}{T}ZZ' \xrightarrow{p} \Omega$ and $S_Z = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathrm{vec}(u_t z_{t-1}') \xrightarrow{d} N(0, \Omega \otimes \Sigma)$ under Assumptions 1-2 (see the proof of Theorem 1(a) in Lohmeyer et al. (2019) for more details), Slutsky's theorem implies the result. ■

Remark 1 (Martingale difference innovations) *Suppose Assumption 1 is weakened to: $\{u_t, \mathcal{F}_t\}$ is a martingale difference sequence with $\mathbb{E}[u_t u_t'] = \Sigma$ positive definite, $\mathbb{E}[|u_{it}u_{jt}u_{kt}u_{mt}|] < c < \infty$, and the array $\{(z_{t-1}z_{t-1}') \otimes (u_t u_t')\}$ obeys a law of large numbers with*

$$\Xi = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[(z_{t-1}z_{t-1}') \otimes (u_t u_t')\right]$$

finite and positive definite. Conditional heteroskedasticity of any form consistent with these moments is permitted. The proof above is unchanged up to the central limit theorem applied to S_Z, which now delivers $S_Z \xrightarrow{d} N(0, \Xi)$ rather than $S_Z \xrightarrow{d} N(0, \Omega \otimes \Sigma)$. Since $\hat{\delta} = (S_0'\Omega^{-1} \otimes \mathrm{I}_K)S_Z + \delta + o_p(1)$, Lemma 3 continues to hold with

$$V^{\mathrm{mds}} = \left(S_0'\Omega^{-1} \otimes \mathrm{I}_K\right) \Xi \left(\Omega^{-1}S_0 \otimes \mathrm{I}_K\right), \tag{A.1}$$

consistently estimated by the sandwich $\hat{V}^{\mathrm{mds}} = (S_0'\hat{\Omega}^{-1} \otimes \mathrm{I}_K)\hat{\Xi}(\hat{\Omega}^{-1}S_0 \otimes \mathrm{I}_K)$ with $\hat{\Xi} = \frac{1}{T}\sum_{t=1}^{T}(z_{t-1}z_{t-1}') \otimes (\hat{u}_t\hat{u}_t')$, the residuals being those of the unrestricted VAR(p). Under conditional homoskedasticity, $\mathbb{E}[u_t u_t' \mid \mathcal{F}_{t-1}] = \Sigma$, the law of iterated expectations gives $\Xi = \Omega \otimes \Sigma$ and (A.1) collapses to $V = [S_0'\Omega^{-1}S_0] \otimes \Sigma$. Every statement in Sections 5 and 6 then goes through with V replaced by V^{mds} and $\hat{V}$ by $\hat{V}^{\mathrm{mds}}$, since each of them uses only Lemma 3 together with the consistency of the variance estimator. What does not survive the weakening is the simplification $\Gamma_m(\Omega \otimes \Sigma)\Gamma_m' = \Omega^{-1} \otimes \Sigma$ at the end of the proof of Proposition 5. The algebraic cancellation there uses the Kronecker form of the score variance, so under (A.1) the robust bands retain the sandwich $\Gamma_m\Xi\Gamma_m' = (\Omega^{-1} \otimes \mathrm{I}_K)\Xi(\Omega^{-1} \otimes \mathrm{I}_K)$, which no longer coincides with the homoskedastic delta-method variance of an unrestricted VAR(p) except in the limit.

Proof of Lemma 1. See the corollary on page 769 of Lohmeyer et al. (2019). The corrected typo noted in footnote 2 can be derived by following the same steps as in Lemma 2 applied to $\mu(\theta) = JA_T(\theta)^h J'$. ■

Proof of Lemma 2. First, to relate the forecast from a VAR(m) directly to the Kp-dimensional space, I pad the VAR(m) parameter estimates with zeros to create a $Kp \times Kp$ companion matrix $\hat{A}_{T,(m)}$. The top K rows of $\hat{A}_{T,(m)}$ are exactly $\hat{\Theta}_{T,m} S_m'$. This allows us to write the forecast identically as:

$$\hat{y}_{T+h|T} = J \hat{A}_{T,(m)}^h z_T.$$

Recall that the analysis is conditional on z_T. Let $\mu(\theta) = y_{T+h|T} = J A_T(\theta)^h z_T$, where I have made explicit that $\theta = \text{vec}(\Theta)$ corresponds to the parameters in the first K rows of the companion matrix A_T. Let θ_∞ denote the parameters obtained by taking $\theta_{T,p}$ but setting $\Delta_1 = \cdots = \Delta_{p_2} = 0$.

The goal is to compute $D_\theta = \partial \mu(\theta_\infty)/\partial \theta'$. Applying the product rule for matrix powers, the differential of A^h is:

$$\text{d}(A_T(\theta)^h) = \sum_{j=0}^{h-1} A_T(\theta)^{h-1-j} (\text{d}A_T(\theta)) A_T(\theta)^j = \sum_{j=0}^{h-1} A_T(\theta)^{h-1-j} J' (\text{d}\Theta) A_T(\theta)^j.$$

Pre-multiplying by J and post-multiplying by z_T, we have the differential of the forecast:

$$\text{d}(J A_T(\theta)^h z_T) = \sum_{j=0}^{h-1} J A_T(\theta)^{h-1-j} J' (\text{d}\Theta) A_T(\theta)^j z_T.$$

Applying the vectorization identity $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ yields:

$$\text{vec}(\text{d}(J A_T(\theta)^h z_T)) = \sum_{j=0}^{h-1} \left((A_T(\theta)^j z_T)' \otimes J A_T(\theta)^{h-1-j} J' \right) \text{vec}(\text{d}\Theta).$$

Notice that $J A_T(\theta_\infty)^k J' = J A_\infty^k J'$ is exactly the k-th horizon true IRF coefficient matrix, Φ_k. Thus, evaluating the derivative at the limit θ_∞, the $K \times K^2 p$ Jacobian matrix D_θ is:

$$D_\theta = \frac{\partial \mu(\theta_\infty)}{\partial \theta'} = \sum_{j=0}^{h-1} \left(z_T' (A_\infty')^j \otimes \Phi_{h-1-j} \right).$$

The result then follows per Theorem 2 in Lohmeyer et al. (2019), whose notation maps into the present one by setting the derivative with respect to the autoregressive parameters to $D_\theta = G_h$ and the derivative with respect to the innovation covariance parameters to zero. ■

Proof of Proposition 1. Part (a) is an immediate consequence of Lemma 3, $ZZ'/T \xrightarrow{p} \Omega$, $\hat{\Sigma} \xrightarrow{p} \Sigma$ and Slutsky's theorem. Part (b) follows because the asymptotic rejection probability under the true local alternative is $\lim_{T\to\infty} \mathbb{P}(W_T > \chi^2_{K^2 p_2, 1-\alpha}) = \mathbb{P}(\chi^2_{K^2 p_2}(\lambda) > \chi^2_{K^2 p_2, 1-\alpha})$. Because λ is a finite non-negative constant, the probability distribution function of a non-central χ^2 random variable places strictly positive mass below any finite threshold. Therefore, $\mathbb{P}(\chi^2_{K^2 p_2}(\lambda) > \chi^2_{K^2 p_2, 1-\alpha}) < 1$. ■

Proof of Proposition 2. By Lemma 3 and the fact that $\hat{V}$ is a consistent estimate of V, Slutsky's theorem implies that

$$(\hat{\delta} - \delta)' \hat{V}^{-1} (\hat{\delta} - \delta) \xrightarrow{d} \xi' V^{-1} \xi,$$

where $\xi \sim N(0, V)$. Thus, $\xi' V^{-1} \xi \sim \chi^2_d$. Let $F_{\chi^2_d}(\cdot)$ denote the cdf of a chi-squared random variable with d degrees of freedom. Then,

$$\lim_{T\to\infty} \mathbb{P}(\delta \in \tilde{\mathcal{C}}) = \lim_{T\to\infty} \mathbb{P}((\hat{\delta} - \delta)' \hat{V}^{-1} (\hat{\delta} - \delta) \leqslant \chi^2_{d,1-\alpha}) = F_{\chi^2_d}(\chi^2_{d,1-\alpha}) = 1 - \alpha,$$

showing the result. ■

Proof of Proposition 3. We must prove that $\mathcal{C}^{\text{pr}}$ is a valid $1 - \alpha$ confidence set for the true asymptotic bias b. Let us start by defining the confidence set

$$\mathcal{C}^0 = \left\{ w\delta \in \mathbb{R}^q : \delta \in \tilde{\mathcal{C}} \right\},$$

which replaces $\hat{w}$ by w. By construction, if the true parameter vector δ lies within the confidence region $\tilde{\mathcal{C}}$, then the true bias $w\delta$ must lie within the mapped set $\mathcal{C}^0$. Therefore, the event $\{\delta \in \tilde{\mathcal{C}}\}$ is a subset of the event $\{w\delta \in \mathcal{C}^0\}$. By the monotonicity of probability measures,

$$\mathbb{P}(b \in \mathcal{C}^0) = \mathbb{P}(w\delta \in \mathcal{C}^0) \geqslant \mathbb{P}(\delta \in \tilde{\mathcal{C}}).$$

Taking the limit as $T \to \infty$ on both sides, Proposition 2 implies that

$$\lim_{T\to\infty} \mathbb{P}(b \in \mathcal{C}^0) \geqslant 1 - \alpha.$$

Since $\hat{V}$ is positive definite, $\tilde{\mathcal{C}}$ is an ellipsoid centered at $\hat{\delta}$. Because the image of an ellipsoid under a surjective linear map is itself an ellipsoid (with probability approaching one when $\hat{w}$ is involved):

$$\mathcal{C}^{\text{pr}} = \left\{ b \in \mathbb{R}^q : (b - \hat{w}\hat{\delta})' (\hat{w}\hat{V}\hat{w}')^{-1} (b - \hat{w}\hat{\delta}) \leqslant \chi^2_{d,1-\alpha} \right\},$$
$$\mathcal{C}^0 = \left\{ b \in \mathbb{R}^q : (b - w\hat{\delta})' (w\hat{V}w')^{-1} (b - w\hat{\delta}) \leqslant \chi^2_{d,1-\alpha} \right\}.$$

Let $D_T = (b - \hat{w}\hat{\delta})' (\hat{w}\hat{V}\hat{w}')^{-1} (b - \hat{w}\hat{\delta})$ and $D_T^0 = (b - w\hat{\delta})' (w\hat{V}w')^{-1} (b - w\hat{\delta})$, so that $b \in \mathcal{C}^{\text{pr}} \iff D_T \leqslant \chi^2_{d,1-\alpha}$ and $b \in \mathcal{C}^0 \iff D_T^0 \leqslant \chi^2_{d,1-\alpha}$.

By standard arguments, $D_T^0 \xrightarrow{d} \chi^2_q$. Because $\hat{w} \xrightarrow{p} w$ and $\hat{\delta} = O_p(1)$, $D_T \xrightarrow{d} \chi^2_q$ and $D_T - D_T^0 \xrightarrow{p} 0$. Since the χ^2_q distribution is continuous, $\chi^2_{d,1-\alpha}$ is a continuity point of its cdf, and therefore

$$\lim_{T\to\infty} \mathbb{P}(b \in \mathcal{C}^{\text{pr}}) = \lim_{T\to\infty} \mathbb{P}(D_T \leqslant \chi^2_{d,1-\alpha}) = \lim_{T\to\infty} \mathbb{P}(D_T^0 \leqslant \chi^2_{d,1-\alpha}) = \lim_{T\to\infty} \mathbb{P}(b \in \mathcal{C}^0) \geqslant 1 - \alpha,$$

completing the proof. ■

Proof of Proposition 4. Note that

$$\hat{w}\hat{\delta} - w\delta = (\hat{w} - w)\hat{\delta} + w(\hat{\delta} - \delta).$$

The consistency of $\hat{w}$ and Lemma 3 imply, by Slutsky's theorem,

$$\hat{w}\hat{\delta} \xrightarrow{d} N(w\delta, w\{[S_0' \Omega^{-1} S_0] \otimes \Sigma\} w').$$

Moreover, $\hat{w}\hat{V}\hat{w}' \xrightarrow{p} wVw'$, which is invertible by the rank condition on w. The result then follows by standard arguments (cf. the proof of Proposition 2). ■

Proof of Proposition 5. Because $\hat{P}$, $\hat{G}_h$ and $\hat{C}_m$ are consistent estimators of P, G_h, and C_m, respectively, and $\hat{\delta} = O_p(1)$, Slutsky's theorem implies that

$$\sqrt{T}\left(\text{vec}(\hat{\Psi}_h) - \text{vec}(\Psi_h)\right) - \hat{H}_{\hat{P}}\hat{G}_h\hat{C}_m\hat{\delta} = \sqrt{T}\left(\text{vec}(\hat{\Psi}_h) - \text{vec}(\Psi_h)\right) - H_P G_h C_m \hat{\delta} + o_p(1).$$

Next, I expand the estimation error of the structural IRF, $\hat{\Psi}_h = \hat{\Phi}_h\hat{P}$, using the delta method:

$$\hat{\Psi}_h - \Psi_h = \hat{\Phi}_h\hat{P} - \Phi_h P = (\hat{\Phi}_h - \Phi_h)P + \Phi_h(\hat{P} - P) + o_p(T^{-1/2}).$$

Using the property of vectorization $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, and multiplying by $\sqrt{T}$,

$$\sqrt{T}\text{vec}(\hat{\Psi}_h - \Psi_h) = (P' \otimes \mathrm{I}_K)\sqrt{T}\text{vec}(\hat{\Phi}_h - \Phi_h) + (\mathrm{I}_K \otimes \Phi_h)\sqrt{T}\text{vec}(\hat{P} - P) + o_p(1).$$

It then follows that

$$\begin{aligned}
&\sqrt{T}\left(\text{vec}(\hat{\Psi}_h) - \text{vec}(\Psi_h)\right) - \hat{H}_{\hat{P}}\hat{G}_h\hat{C}_m\hat{\delta} \\
&\quad = (P' \otimes \mathrm{I}_K)\underbrace{\left[\sqrt{T}\text{vec}(\hat{\Phi}_h - \Phi_h) - G_h C_m \hat{\delta}\right]}_{=:M} + (\mathrm{I}_K \otimes \Phi_h)\sqrt{T}\text{vec}(\hat{P} - P) + o_p(1).
\end{aligned}$$

Because $\hat{P}$ is asymptotically independent of S_Z, and both $\hat{\Phi}_h$ and $\hat{\delta}$ are linear in S_Z up to $o_p(1)$ terms, the estimation error of the autoregressive parameters and the estimation error of the structural shock matrix are asymptotically independent, so the first and second terms will be independent (Lütkepohl, 1990; Lohmeyer et al., 2019). What remains is then to characterize the marginal asymptotic distributions.

By assumption and the continuous mapping theorem, the second term satisfies

$$(\mathrm{I}_K \otimes \Phi_h)\sqrt{T}\text{vec}(\hat{P} - P) \xrightarrow{d} N(0, (\mathrm{I}_K \otimes \Phi_h)\Sigma_P(\mathrm{I}_K \otimes \Phi_h')).$$

For instance, when $\hat{P}$ is the Cholesky factor of $\hat{\Sigma}$, this holds under Assumptions 1-3 with $\Sigma_P = J_P\Omega_\Sigma J_P'$, where Ω_Σ is the $K^2 \times K^2$ fourth-moment covariance matrix of the VAR residuals and $J_P = L_K' F_P^{-1} L_K$, with L_K the $K(K+1)/2 \times K^2$ elimination matrix that extracts the lower triangular elements such that $\text{vech}(A) = L_K\text{vec}(A)$, and the $K(K+1)/2 \times K(K+1)/2$ matrix $F_P = L_K(\mathrm{I}_{K^2} + K_{KK})(P \otimes \mathrm{I}_K)L_K'$ where the $K^2 \times K^2$ commutation matrix K_{KK} has the property $K_{KK}\text{vec}(A) = \text{vec}(A')$ (Lütkepohl, 1990).

To find the distribution of M, we must establish the joint asymptotic behavior of $\hat{\Phi}_h$ and $\hat{\delta}$. Both estimators are linear functions of the same underlying OLS score vector. Under Assumption 1,

$$\sqrt{T}\text{vec}(\hat{\Theta}_{T,p} - \Theta_{T,p}) = (\Omega^{-1} \otimes \mathrm{I}_K)S_Z + o_p(1), \quad S_Z \xrightarrow{d} N(0, \Omega \otimes \Sigma).$$

Because $\hat{\delta} = \sqrt{T}\text{vec}(\hat{\Theta}_{T,p}S_0)$, we can isolate the asymptotic distribution of the estimated drift:

$$\hat{\delta} = (S_0' \otimes \mathrm{I}_K)\sqrt{T}\text{vec}(\hat{\Theta}_{T,p} - \Theta_{T,p}) + \sqrt{T}\text{vec}(\Theta_{T,p}S_0) = (S_0'\Omega^{-1} \otimes \mathrm{I}_K)S_Z + \delta + o_p(1).$$

Moreover, per equation (2),

$$\sqrt{T}\text{vec}(\hat{\Theta}_{T,m}S_m' - \Theta_{T,p}) = C_m\delta + P_m S_Z + o_p(1),$$

so that Lemma 1 implies

$$\sqrt{T}\text{vec}(\hat{\Phi}_h - \Phi_h) = G_h C_m \delta + G_h P_m S_Z + o_p(1).$$

It then follows that

$$M = \{G_h C_m \delta + G_h P_m S_Z\} - G_h C_m \left\{(S_0'\Omega^{-1} \otimes \mathrm{I}_K)S_Z + \delta\right\} + o_p(1).$$

The asymptotic biases cancel out and we obtain

$$M = G_h \Gamma_m S_Z + o_p(1), \quad \Gamma_m = P_m - C_m(S_0'\Omega^{-1} \otimes \mathrm{I}_K),$$

hence

$$(P' \otimes \mathrm{I}_K)M \xrightarrow{d} N(0,\ (P' \otimes \mathrm{I}_K)G_h\Gamma_m(\Omega \otimes \Sigma)\Gamma_m' G_h'(P \otimes \mathrm{I}_K)).$$

We then conclude that

$$\sqrt{T}\left(\text{vec}(\hat{\Psi}_h) - \text{vec}(\Psi_h)\right) - \hat{H}_{\hat{P}}\hat{G}_h\hat{C}_m\hat{\delta} \xrightarrow{d} N(0, \tilde{V}),$$

where $\tilde{V} = H_P G_h \Gamma_m(\Omega \otimes \Sigma)\Gamma_m' G_h' H_P' + (\mathrm{I}_K \otimes \Phi_h)\Sigma_P(\mathrm{I}_K \otimes \Phi_h')$.

Finally, Γ_m simplifies considerably. Because $\Pi_m'\Pi_m$ selects the first $K(m-p_1)$ coordinates of the drift block, $S_0(\mathrm{I}_{Kp_2} - \Pi_m'\Pi_m)S_0' = \mathrm{I}_{Kp} - S_m S_m'$. Letting $Q_m = S_m[S_m'\Omega S_m]^{-1}S_m'$, so that $P_m = Q_m \otimes \mathrm{I}_K$,

$$\begin{aligned}
\Gamma_m &= P_m - C_m(S_0'\Omega^{-1} \otimes \mathrm{I}_K) = \left[Q_m - (Q_m\Omega - \mathrm{I}_{Kp})S_0(\mathrm{I}_{Kp_2} - \Pi_m'\Pi_m)S_0'\Omega^{-1}\right] \otimes \mathrm{I}_K \\
&= \left[Q_m - (Q_m\Omega - \mathrm{I}_{Kp})(\mathrm{I}_{Kp} - S_m S_m')\Omega^{-1}\right] \otimes \mathrm{I}_K \\
&= \left[Q_m - Q_m + Q_m\Omega S_m S_m'\Omega^{-1} + \Omega^{-1} - S_m S_m'\Omega^{-1}\right] \otimes \mathrm{I}_K = \Omega^{-1} \otimes \mathrm{I}_K,
\end{aligned}$$

where the last equality uses $Q_m\Omega S_m = S_m$. Consequently, $\Gamma_m(\Omega \otimes \Sigma)\Gamma_m' = \Omega^{-1} \otimes \Sigma$, which delivers the variance expression discussed in Section 6.4. ■

Proof of Proposition 6. The proof is similar to that of Proposition 5, so I purposely skip some details.

Because $\hat{G}_h$ and $\hat{C}_m$ are consistent estimators of G_h and C_m, respectively, and $\hat{\delta} = O_p(1)$, Slutsky's theorem implies that

$$\sqrt{T}\left(\hat{y}_{T+h|T} - y_{T+h|T}\right) - \hat{G}_h\hat{C}_m\hat{\delta} = \sqrt{T}\left(\hat{y}_{T+h|T} - y_{T+h|T}\right) - G_h C_m \hat{\delta} + o_p(1).$$

We must then establish the joint asymptotic behavior of $\hat{y}_{T+h|T}$ and $\hat{\delta}$. Both estimators are linear functions of the same underlying OLS score vector. Per the same arguments as in the proof of Proposition 5,

$$\hat{\delta} = (S_0'\Omega^{-1} \otimes \mathrm{I}_K)S_Z + \delta + o_p(1).$$

By Lemma 2,

$$\sqrt{T}(\hat{y}_{T+h|T} - y_{T+h|T}) = G_h C_m \delta + G_h P_m S_Z + o_p(1).$$

It then follows that

$$\sqrt{T}\left(\hat{y}_{T+h|T} - y_{T+h|T}\right) - \hat{G}_h \hat{C}_m \hat{\delta} = G_h \Gamma_m S_Z + o_p(1), \quad \Gamma_m = P_m - C_m(S_0'\Omega^{-1} \otimes \mathrm{I}_K).$$

We then conclude that

$$\sqrt{T}\left(\hat{y}_{T+h|T} - y_{T+h|T}\right) - \hat{G}_h \hat{C}_m \hat{\delta} \xrightarrow{d} N(0,\ G_h \Gamma_m (\Omega \otimes \Sigma) \Gamma_m' G_h').$$

■

B Illustrative Example

Consider the following data generating process

$$\begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} y_{t-1,1} \\ y_{t-1,2} \end{bmatrix} + \begin{bmatrix} \Delta_{11}/\sqrt{T} & \Delta_{12}/\sqrt{T} \\ \Delta_{21}/\sqrt{T} & \Delta_{22}/\sqrt{T} \end{bmatrix} \begin{bmatrix} y_{t-2,1} \\ y_{t-2,2} \end{bmatrix} + \begin{bmatrix} u_{t,1} \\ u_{t,2} \end{bmatrix}.$$

In this system, the true IRF matrix is given by:

- Horizon $h = 0$: $\Phi_0 = \mathrm{I}_2$;
- Horizon $h = 1$: $\Phi_1 = B\Phi_0 = B$;
- Horizon $h \geqslant 2$: $\Phi_h = B\Phi_{h-1} + \frac{\Delta}{\sqrt{T}}\Phi_{h-2}$.

Suppose that the object of interest is the IRF of a shock to variable 1 on variable 2. The response of variable 2 to a one-unit reduced-form shock in variable 1 at horizon h is given by the $(2,1)$-element of the matrix Φ_h. Let $e_1 = (1,0)'$ and $e_2 = (0,1)'$. The target IRF coefficient is then:

$$\Phi_{h,21} = e_2'\Phi_h e_1.$$

Letting $e = (e_1 \otimes e_2)$, we can rewrite the IRF in vectorized form as $\Phi_{h,21} = e'\mathrm{vec}(\Phi_h)$.

An asymptotic bias arises when the econometrician underfits the model by omitting the second lag matrix (which contains the local-to-zero drift). Thus, the econometrician estimates a VAR(m) with $m = p_1 = 1$. I specialize the components of Lemma 1 under $K = 2, p_1 = 1, p_2 = 1, p = 2$, and $m = 1$.

The lag cross-product matrix Ω can be partitioned into 2×2 blocks corresponding to the two lags:

$$\Omega = \operatorname{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_{t-1} z_{t-1}' = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix},$$

where $\Omega_{ij} = \operatorname{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} y_{t-i} y_{t-j}'$. The selection matrices for $m = 1$ are:

$$S_1^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \implies S_1 = S_1^{(1)} \otimes \mathrm{I}_2 = \begin{bmatrix} \mathrm{I}_2 \\ 0_{2\times 2} \end{bmatrix},$$

$$S_0^{(1)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \implies S_0 = S_0^{(1)} \otimes \mathrm{I}_2 = \begin{bmatrix} 0_{2\times 2} \\ \mathrm{I}_2 \end{bmatrix}.$$

Since $m = p_1$, the matrix $\Pi_1 = 0$. Moreover,

$$S_1' \Omega S_1 = \begin{bmatrix} \mathrm{I}_2 & 0 \end{bmatrix} \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \begin{bmatrix} \mathrm{I}_2 \\ 0 \end{bmatrix} = \Omega_{11},$$

and thus

$$S_1 [S_1' \Omega S_1]^{-1} S_1' \Omega = \begin{bmatrix} \mathrm{I}_2 \\ 0 \end{bmatrix} \Omega_{11}^{-1} \begin{bmatrix} \Omega_{11} & \Omega_{12} \end{bmatrix} = \begin{bmatrix} \mathrm{I}_2 & \Omega_{11}^{-1}\Omega_{12} \\ 0 & 0 \end{bmatrix}.$$

Subtracting I_4 and multiplying by S_0:

$$\left(S_1 [S_1' \Omega S_1]^{-1} S_1' \Omega - \mathrm{I}_4 \right) S_0 = \begin{bmatrix} 0 & \Omega_{11}^{-1}\Omega_{12} \\ 0 & -\mathrm{I}_2 \end{bmatrix} \begin{bmatrix} 0 \\ \mathrm{I}_2 \end{bmatrix} = \begin{bmatrix} \Omega_{11}^{-1}\Omega_{12} \\ -\mathrm{I}_2 \end{bmatrix}.$$

Because $\Pi_1 = 0$, the term $(\mathrm{I}_2 - \Pi_1'\Pi_1) = \mathrm{I}_2$. Thus, via the Kronecker formulation:

$$C_1 = \begin{bmatrix} \Omega_{11}^{-1}\Omega_{12} \\ -\mathrm{I}_2 \end{bmatrix} \otimes \mathrm{I}_2 = \begin{bmatrix} (\Omega_{11}^{-1}\Omega_{12}) \otimes \mathrm{I}_2 \\ -\mathrm{I}_4 \end{bmatrix}.$$

The companion matrix A_∞ eliminates the local-to-zero misspecification coefficients, reducing the framework to a standard VAR(1) limit process in companion form:

$$A_\infty = \begin{bmatrix} B & 0 \\ \mathrm{I}_2 & 0 \end{bmatrix}, \quad J = \begin{bmatrix} \mathrm{I}_2 & 0 \end{bmatrix}.$$

By induction, for any power $k \geqslant 1$, $J(A_\infty')^k = \begin{bmatrix} (B')^k & (B')^{k-1} \end{bmatrix}$. In the limit local-to-zero process, the

moving average coefficients inside G_h equal $\Phi_j = B^j$. Thus:

$$\begin{aligned} G_h &= \sum_{j=0}^{h-1} \left(J(A'_\infty)^{h-1-j} \otimes \Phi_j \right) \\ &= J \otimes B^{h-1} + \sum_{j=0}^{h-2} \left(\begin{bmatrix} (B')^{h-1-j} & (B')^{h-2-j} \end{bmatrix} \otimes B^j \right) \\ &= J \otimes B^{h-1} + \sum_{j=0}^{h-2} \begin{bmatrix} (B')^{h-1-j} \otimes B^j & (B')^{h-2-j} \otimes B^j \end{bmatrix}. \end{aligned}$$

From Lemma 1, the asymptotic bias vector of $\text{vec}(\hat{\Phi}_h)$ is given by $G_h C_1 \delta$, where $\delta = \text{vec}(\Delta)$. Multiplying G_h and C_1 found above:

$$\begin{aligned} G_h C_1 &= \left\{ J \otimes B^{h-1} + \sum_{j=0}^{h-2} \begin{bmatrix} (B')^{h-1-j} \otimes B^j & (B')^{h-2-j} \otimes B^j \end{bmatrix} \right\} \begin{bmatrix} (\Omega_{11}^{-1}\Omega_{12}) \otimes \mathrm{I}_2 \\ -\mathrm{I}_4 \end{bmatrix} \\ &= (\mathrm{I}_2 \otimes B^{h-1}) \left((\Omega_{11}^{-1}\Omega_{12}) \otimes \mathrm{I}_2 \right) + \sum_{j=0}^{h-2} \left\{ \left((B')^{h-1-j} \otimes B^j \right) \left((\Omega_{11}^{-1}\Omega_{12}) \otimes \mathrm{I}_2 \right) - \left((B')^{h-2-j} \otimes B^j \right) \mathrm{I}_4 \right\}. \end{aligned}$$

Applying the Kronecker product property $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$:

$$G_h C_1 = (\Omega_{11}^{-1}\Omega_{12}) \otimes B^{h-1} + \sum_{j=0}^{h-2} \left\{ \left((B')^{h-1-j} \Omega_{11}^{-1}\Omega_{12} \right) \otimes B^j - (B')^{h-2-j} \otimes B^j \right\}.$$

Because $(A + B) \otimes C = A \otimes C + B \otimes C$, we may write:

$$G_h C_1 = (\Omega_{11}^{-1}\Omega_{12}) \otimes B^{h-1} + \sum_{j=0}^{h-2} \left((B')^{h-1-j} \Omega_{11}^{-1}\Omega_{12} - (B')^{h-2-j} \right) \otimes B^j.$$

Hence, the asymptotic bias for the IRF of a shock to variable 1 on variable 2 is given by:

$$e' G_h C_1 \delta = (e_1' \otimes e_2') \left[(\Omega_{11}^{-1}\Omega_{12}) \otimes B^{h-1} + \sum_{j=0}^{h-2} \left((B')^{h-1-j} \Omega_{11}^{-1}\Omega_{12} - (B')^{h-2-j} \right) \otimes B^j \right] \delta.$$

Using again the mixed-product rule of Kronecker products, we can simplify the bias expression:

$$\left[(e_1' \Omega_{11}^{-1}\Omega_{12}) \otimes \left(e_2' B^{h-1} \right) + \sum_{j=0}^{h-2} \left(e_1' (B')^{h-1-j} \Omega_{11}^{-1}\Omega_{12} - e_1' (B')^{h-2-j} \right) \otimes \left(e_2' B^j \right) \right] \delta.$$

Alternatively, we may map this back to standard matrix form using the identity $(M \otimes N)\text{vec}(X) = \text{vec}(NXM')$, in which case the bias writes as

$$e_2' B^{h-1} \Delta \Omega_{21} \Omega_{11}^{-1} e_1 + e_2' \left\{ \sum_{j=0}^{h-2} B^j \Delta \left(\Omega_{21}\Omega_{11}^{-1} B^{h-1-j} - B^{h-2-j} \right) \right\} e_1.$$

For instance, at horizon $h = 1$, the second term drops out[21] and the asymptotic bias is $e_2' \left(\Delta\Omega_{21}\Omega_{11}^{-1}\right) e_1$, i.e., the (2,1)-element of the matrix $\Delta\Omega_{21}\Omega_{11}^{-1}$.

C Extension to VAR(∞)

The framework in the main text assumes a finite-order VAR data-generating process. In this section I discuss how the results may extend to infinite-order VAR data-generating processes.

If remote omitted lags outside the p_2 window exert a local drift of order $O(T^{-1/2})$, their omission could introduce an unmodeled, sign-dependent bias that invalidates the coverage guarantees of the proposed misspecification-robust tools. To anchor the framework natively within a VAR(∞) environment, it suffices to impose an explicit tail condition on the autoregressive coefficients, requiring the cumulative effect of lags beyond $p_1 + p_2$ to vanish faster than $O(T^{-1/2})$. Because the local-to-zero paradigm already describes the data as a triangular array, it is natural to let the whole coefficient sequence, and not only the drift block, be indexed by the sample size.

Assumption C.1 *Let the data be generated by the triangular array of VAR(∞) processes*

$$y_t = \sum_{j=1}^{\infty} B_{j,T}\, y_{t-j} + u_t,$$

where, for a given choice of baseline lag length p_1 and local lag block p_2,

$$B_{j,T} = B_j \quad \text{for } j \leqslant p_1, \qquad B_{j,T} = \frac{\Delta_j}{\sqrt{T}} \quad \text{for } j = p_1 + 1, \ldots, p_1 + p_2,$$

with $\{B_j\}_{j\leqslant p_1}$ and $\{\Delta_j\}_{j\leqslant p_2}$ fixed and bounded, and the remaining tail coefficients satisfy

$$\lim_{T\to\infty} \sqrt{T} \sum_{j=p_1+p_2+1}^{\infty} \|B_{j,T}\| = 0,$$

where $\|\cdot\|$ denotes the spectral norm. Moreover, $\sup_T \sum_{j=1}^{\infty} \|B_{j,T}\| < \infty$ and the roots of $\det\left(\mathrm{I}_K - \sum_{j=1}^{\infty} B_{j,T} z^j\right)$ lie outside the unit circle uniformly in T.

Assumption C.1 establishes that while the dynamic misspecification within the block $\{p_1+1, \ldots, p_1+p_2\}$ is local-to-zero and exerts a first-order asymptotic effect on the sampling distributions, the remote dynamics beyond $p_1 + p_2$ decay fast enough to be asymptotically negligible. Under this condition, none of the results in the main text is affected.

Proposition C.1 *Under Assumptions 1 and C.1, Lemmas 1, 2 and 3, and Propositions 1-6, continue to hold, with Ω and Σ defined as in the main text for the limiting process.*

Proof Sketch of Proposition C.1. Write the data-generating process as $y_t = \sum_{j=1}^{p} B_{j,T} y_{t-j} + u_t + r_t$, where $r_t = \sum_{j>p} B_{j,T}\, y_{t-j}$ is the truncation residual. Because the array is uniformly stationary with

[21] By standard mathematical convention, a summation where the lower bound exceeds the upper bound is empty and evaluates to zero.

uniformly summable coefficients, $\mathbb{E}[\|y_t\|^2]$ is bounded uniformly in T, so

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\operatorname{vec}(r_t z_{t-1}') = O_p\left(\sqrt{T}\sum_{j>p}\|B_{j,T}\|\right) = o_p(1),$$

by Assumption C.1. The estimating equations behind Lemma 3 are therefore unchanged up to $o_p(1)$, and so are the probability limits of the sample second moments $\frac{1}{T}ZZ'$. The same bound applies to the estimands. The impulse responses and forecasts implied by the truncated system differ from their VAR(∞) counterparts by $O\left(\sum_{j>p}\|B_{j,T}\|\right) = o(T^{-1/2})$, hence vanish after scaling by $\sqrt{T}$. All the limits in the statement are thus unaffected. ■

Assumption C.1 makes precise the sense in which lags beyond $p_1 + p_2$ may be ignored. They are allowed to be non-zero, but their cumulative contribution must vanish faster than the $O(T^{-1/2})$ drift the framework is designed to bound.

The relaxation closest to the framework of the main text places the entire tail on the scale of the drift. Suppose that

$$B_{j,T} = B_j \quad (j \leqslant p_1), \qquad B_{j,T} = \frac{\Delta_j}{\sqrt{T}} \quad (j > p_1), \qquad \sum_{j>p_1}\|\Delta_j\| < \infty,$$

so that the limiting process is still the VAR(p_1) but the whole omitted structure, and not only its first p_2 blocks, is local to zero, and let $r(p_2) = \sum_{j>p_1+p_2}\|\Delta_j\|$ denote the part of the drift left outside the window. Then $\sqrt{T}\sum_{j>p_1+p_2}\|B_{j,T}\| = r(p_2)$ does not vanish, but neither does it grow with T, and the argument in the proof of Proposition C.1 goes through with $o_p(1)$ replaced by $O_p(r(p_2))$ in both the estimating equations and the estimands. The limit distributions of Section 6 are then shifted by a deterministic term of order $r(p_2)$, and the asymptotic coverage of the confidence sets is $1 - \alpha + O(r(p_2))$; Assumption C.1 is the special case $r(p_2) = 0$. When the drift decays geometrically, $\|\Delta_j\| \leqslant C\rho^j$ with $0 < \rho < 1$, as for the reduced forms discussed in Section 3, $r(p_2) \leqslant \rho^{p_2} C\rho^{p_1+1}/(1-\rho)$, a fraction ρ^{p_2} of the bound on the total drift: about 3% for $\rho = 0.65$ and $p_2 = 8$, but 43% for $\rho = 0.9$. The approximation is thus accurate for moderately persistent omitted dynamics and deteriorates only when the omitted dynamics are themselves persistent, which is the configuration in which the robustness of a direct projection is worth its variance cost. A fully general treatment would let p_2 grow with the sample size and the drift dimension diverge, in the manner of sieve autoregressive approximations (Lewis and Reinsel, 1985). The scalar intervals of Proposition 4 are the natural candidates to survive such an extension, since the weights w remain summable, whereas the projection sets of Proposition 3 degrade as the χ^2_d critical value grows with d. I leave this to future work.

D Alternative Identification Schemes

Section 4 showed that identification strategies relying strictly on the contemporaneous residual covariance matrix $\hat{\Sigma}$ are asymptotically insulated from the local drift, because the contribution of the omitted lags to Σ is $O(1/T)$. This section records what happens when that insulation fails, for the two most common alternatives, although the content is deliberately limited. What is established formally is the first-order bias of the relevant estimators, which follows from Lemma 1 and equation (2) by the same steps as in the main text. What is offered as a construction rather than as a result is how one might correct for it. A coverage

statement in either case would require carrying the sampling error of the identification step through the argument, which is beyond the scope of this appendix and is left to future work.

D.1 Identification via Sign Restrictions

Under set identification, any valid impact matrix can be written $P = \tilde{P}Q$, with $\tilde{P}$ a baseline decomposition satisfying $\tilde{P}\tilde{P}' = \Sigma$ and $Q \in \mathcal{O}_K$ an orthogonal rotation. Let $\mathcal{R}(\cdot)$ evaluate the structural responses against a set of sign restrictions imposed over the horizons collected in $\mathcal{J}$, so that the true identified set at horizon h is

$$S_h = \Big\{ \Phi_h \tilde{P} Q : Q \in \mathcal{O}_K,\ \mathcal{R}(\Phi_j \tilde{P} Q) \geqslant 0 \ \ \forall j \in \mathcal{J} \Big\}.$$

Because $\hat{\tilde{P}}$ converges to $\tilde{P}$ free of local distortions while $\hat{\Phi}_j$ does not, extending Corollary 1 gives the asymptotic bias of the candidate response $\hat{\Psi}_h(Q) = \hat{\Phi}_h \hat{\tilde{P}} Q$ as $(Q'\tilde{P}' \otimes \mathrm{I}_K) G_h C_m \delta$. The consequence is specific to set identification: since the admissibility of Q is decided by the empirical evaluation $\mathcal{R}(\hat{\Phi}_j \hat{\tilde{P}} Q) \geqslant 0$, the bias in $\hat{\Phi}_j$ does not merely displace the estimated set but deforms its boundary, so that a rotation valid under the true process may be discarded and an invalid one retained. The distortion therefore survives even where the sign restrictions are uninformative about the level of the response.

A natural correction inverts the bias mapping over the confidence region $\tilde{\mathcal{C}}$ of Proposition 2. For $\delta \in \tilde{\mathcal{C}}$, define the bias-corrected reduced-form response $\hat{\Phi}_h^c(\delta) = \hat{\Phi}_h - T^{-1/2}\mathrm{vec}^{-1}(\hat{G}_h \hat{C}_m \delta)$, where $\mathbb{R}^{mn} \ni x \mapsto \mathrm{vec}^{-1}(x) = (\mathrm{vec}(\mathrm{I}_n)' \otimes \mathrm{I}_m)(\mathrm{I}_n \otimes x) \in \mathbb{R}^{m \times n}$, and take

$$\hat{S}_h^{\mathrm{rob}} = \bigcup_{\delta \in \tilde{\mathcal{C}}} \Big\{ \hat{\Phi}_h^c(\delta) \hat{\tilde{P}} Q : Q \in \mathcal{O}_K,\ \mathcal{R}\big(\hat{\Phi}_j^c(\delta) \hat{\tilde{P}} Q\big) \geqslant 0 \ \ \forall j \in \mathcal{J} \Big\}.$$

This removes the systematic deformation, and by construction contains the set that would have been obtained at the true drift with probability at least $1 - \alpha$ in the limit if $\hat{\Phi}_h$ and $\hat{\tilde{P}}$ were known; it is the sampling error in those two that a coverage statement would have to absorb.

D.2 Identification via External Proxies

In a proxy-VAR, a structural shock ϵ_{1t}, ordered first without loss of generality, is identified by an external instrument M_t satisfying $\mathbb{E}[M_t \epsilon_{1t}] = \kappa \neq 0$ and $\mathbb{E}[M_t \epsilon_{jt}] = 0$ for $j \neq 1$. Let P_1 be the first column of P. The unnormalized impact vector $\gamma = P_1 \kappa$ is estimated by $\hat{\gamma} = T^{-1} \sum_t \hat{u}_{m,t} M_t$, with $\hat{u}_{m,t}$ the residuals of the restricted VAR(m), and is typically normalized as $\hat{P}_1 = \hat{\gamma}/\hat{\gamma}_1$.

Under (1) those residuals satisfy

$$\hat{u}_{m,t} = u_t + T^{-1/2} \Delta (\mathrm{I}_{Kp_2} - \Pi_m' \Pi_m) S_0' z_{t-1} - (\hat{\Theta}_{T,m} - \Theta_{T,m}) S_m' z_{t-1},$$

so that, writing $\rho_0 = \mathbb{E}[M_t S_0' z_{t-1}]$ and $\rho_m = \mathbb{E}[M_t S_m' z_{t-1}]$ for the cross-moments of the instrument with

the omitted and the retained lags,

$$\sqrt{T}(\hat{\gamma} - P_1\kappa) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t M_t - P_1\kappa) + \Delta(\mathrm{I}_{Kp_2} - \Pi_m'\Pi_m)\left(\frac{1}{T}\sum_{t=1}^{T} M_t S_0' z_{t-1}\right) - \sqrt{T}(\hat{\Theta}_{T,m} - \Theta_{T,m})\left(\frac{1}{T}\sum_{t=1}^{T} M_t S_m' z_{t-1}\right).$$

Strict lead/lag exogeneity, the requirement that the instrument be unpredictable from past aggregates, is exactly the condition $\rho_0 = \rho_m = 0$, under which the last two terms vanish and the proxy shields the impact vector from dynamic misspecification entirely. If it fails, which is the relevant case when the instrument was constructed from an information set that itself omitted the higher-order lags $S_0' z_{t-1}$, then using $\sqrt{T}(\hat{\Theta}_{T,m} - \Theta_{T,m}) \xrightarrow{d} \Delta W_m + Z_U$ from (2), with $W_m = (\mathrm{I}_{Kp_2} - \Pi_m'\Pi_m)S_0'\Omega S_m[S_m'\Omega S_m]^{-1}$ and Z_U centered Gaussian, the impact estimator inherits the first-order bias $\Delta\Lambda$, or equivalently $(\Lambda' \otimes \mathrm{I}_K)\delta$, where $\Lambda = (\mathrm{I}_{Kp_2} - \Pi_m'\Pi_m)\rho_0 - W_m\rho_m$. The estimated structural response $\hat{\Psi}_h = \hat{\Phi}_h\hat{\gamma}$ then carries

$$\Big[(\gamma' \otimes \mathrm{I}_K)G_h C_m + \Phi_h(\Lambda' \otimes \mathrm{I}_K)\Big]\delta,$$

the first term being the familiar propagation bias and the second the contamination of the proxy itself, which vanishes only under orthogonality to the omitted dynamics.

The two channels become inseparable under weak identification. With local-to-zero relevance $\kappa = \tilde{\kappa}/\sqrt{T}$, the same derivations give the asymptotic bias of $\hat{\gamma}$ as $P_1\tilde{\kappa} + (\Lambda' \otimes \mathrm{I}_K)\delta$: the structural signal $\Phi_h P_1\tilde{\kappa}$ and the misspecification distortion $\Phi_h(\Lambda' \otimes \mathrm{I}_K)\delta$ enter the limit additively and at the same rate, so neither is recoverable from $\hat{\gamma}$ alone. Weak-instrument procedures that maintain a correctly specified dynamic model (e.g., Montiel Olea, Stock and Watson, 2021) will therefore attribute the lag-truncation distortion to the structural shock. As with sign restrictions, one could partial the bias out over $\delta \in \tilde{\mathcal{C}}$, yet the same caveat about coverage applies.

E Omitted Variables As Moving Average Terms

This section shows how the traditional omitted variable problem manifests within small-scale systems as a problem of omitted high-order dynamics. I first proceed analytically, and then illustrate the claim using a numerical experiment.

E.1 Analytical Explanation

To prove this formally, I rely on standard marginalization arguments. For simplicity, suppose the true data-generating process is a stationary VAR(1) that includes both the observed macroeconomic variable and an omitted forward-looking state variable. Let the full system be partitioned as:

$$\begin{bmatrix} Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ u_{z,t} \end{bmatrix},$$

where Y_t is the retained variable and Z_t is the omitted variable.

We can rewrite this, using the lag operator L, as:

$$\begin{cases} (1-\Phi_{11}L)Y_t - \Phi_{12}LZ_t = u_{y,t} \\ -\Phi_{21}LY_t + (1-\Phi_{22}L)Z_t = u_{z,t} \end{cases}$$

To find the univariate representation of the retained variable Y_t, multiply the first equation by $(1-\Phi_{22}L)$ and the second by $\Phi_{12}L$, then add them to eliminate the omitted variable Z_t:

$$[(1-\Phi_{22}L)(1-\Phi_{11}L) - \Phi_{12}\Phi_{21}L^2]Y_t = (1-\Phi_{22}L)u_{y,t} + \Phi_{12}Lu_{z,t}.$$

Assume for simplicity that the structural innovations $u_{y,t}$ and $u_{z,t}$ are orthogonal white noise with variances σ_y^2 and σ_z^2, respectively. Then, the right hand side is an MA(1) process. By Wold's Theorem, it has a fundamental representation

$$(1-\Phi_{22}L)u_{y,t} + \Phi_{12}Lu_{z,t} = (1-\theta L)v_t,$$

where v_t is white noise with variance σ_v^2. Tedious algebra gives

$$\theta = \frac{c - \sqrt{c^2-4}}{2}, \quad c = \frac{1+\Phi_{22}^2}{\Phi_{22}} + \frac{\Phi_{12}^2\sigma_z^2}{\Phi_{22}\sigma_y^2}.$$

The derivation above shows that omitting a relevant state variable from a bivariate VAR(1) leads to an ARMA(2,1) in the retained variable. Using standard arguments, it is then clear that the omission of a state variable then translates into an AR(∞).

We can further use this to show that higher-order lags are of smaller order under some conditions, replicating the asymptotic paradigm used in this paper. I formalize the weak omitted variable assumption by setting the cross-equation feedbacks Φ_{12} and Φ_{21} to scale as $O(T^{-1/4})$. This implies that

$$c = \Phi_{22}^{-1} + \Phi_{22} + \omega, \quad \omega = \frac{\Phi_{12}^2\sigma_z^2}{\Phi_{22}\sigma_y^2},$$

where by the assumed rates $\omega = O(T^{-1/2})$. Then,

$$c^2 - 4 = (\Phi_{22}^{-1} - \Phi_{22})^2 + 2\omega(\Phi_{22}^{-1} + \Phi_{22}) + \omega^2.$$

By a Taylor expansion,

$$\sqrt{c^2-4} = (\Phi_{22}^{-1} - \Phi_{22}) + \omega\left(\frac{1+\Phi_{22}^2}{1-\Phi_{22}^2}\right) + O(T^{-1}).$$

This implies that

$$\theta = \Phi_{22} - \omega\left(\frac{\Phi_{22}^2}{1-\Phi_{22}^2}\right) + O(T^{-1}),$$

or, in other words, $\theta = \Phi_{22} + O(T^{-1/2})$.

Recall that omitting a state variable from a bivariate VAR(1) leads to an ARMA(2,1). Mapping the ARMA into an AR(∞) representation, generally written as

$$Y_t = \sum_{j=1}^{\infty} A_j Y_{t-j} + v_t,$$

implies

$$\begin{aligned} A_1 &= \Phi_{11} + \Phi_{22} - \theta = \Phi_{11} + O(T^{-1/2}) \\ A_k &= \theta^{k-2}\left[(\Phi_{11} + \Phi_{22})\theta + (\Phi_{12}\Phi_{21} - \Phi_{11}\Phi_{22}) - \theta^2\right] = O(T^{-1/2}) \text{ for } k \geqslant 2. \end{aligned}$$

This perfectly reconstructs the paper's environment where the first $p_1 = 1$ lags capture the main $O(1)$ dynamics, and the subsequent p_2 lags encapsulate the $O(T^{-1/2})$ unestimable dynamic drift caused by the omitted variable. Strictly speaking, the structural marginalization yields an infinite-order autoregressive process which theoretically implies $p_2 \to \infty$. However, within the formal boundaries of the local misspecification framework, p_2 is deliberately kept finite, since the multiplier θ^{k-2} decays geometrically toward zero. The analysis in this section easily extends to a general p_1 by changing the order of the underlying VAR(1) data-generating process.

E.2 Numerical Illustration

To numerically illustrate the claim that omitted variables introduce MA terms that fundamentally alter the dynamic order of the observable system, I conduct a stylized Monte Carlo experiment.

I simulate a stationary, bivariate VAR(1) data-generating process of length $T = 10,000$ (to isolate the asymptotic theoretical properties). The system consists of a retained variable Y_t and an omitted variable Z_t. Following the notation in the earlier subsection, the autoregressive persistence parameters are set to $\Phi_{11} = 0.6$ and $\Phi_{22} = 0.8$. I then vary the cross-equation feedback terms (Φ_{12} and Φ_{21}) across three scenarios to visualize how spillovers from the omitted variable reshape the univariate representation of Y_t: (i) no spillover, with $\Phi_{12} = \Phi_{21} = 0$; (ii) weak spillover (local-to-zero), with $\Phi_{12} = \Phi_{21} = 0.15$; (iii) strong spillover, with $\Phi_{12} = \Phi_{21} = 0.25$. For each scenario, I marginalize out the omitted state Z_t and compute the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the retained variable Y_t. As formally derived in the previous subsection, the PACF is the informative diagnostic here: a pure finite-order AR process will exhibit a strict cutoff, whereas an ARMA process (caused by the omitted state) will yield an infinite-order AR representation characterized by a trailing decay in the PACF.

The results of this experiment are plotted in Figure E.VIII, with the ACF on the top row, the PACF on the bottom row, and the columns corresponding to the no, weak, and strong spillover scenarios, respectively.

Let us start by analyzing the no spillover case. When the cross-terms are exactly zero, the omitted variable Z_t exerts no feedback on Y_t. The PACF exhibits a large spike at the first lag and immediately cuts off to zero for all subsequent lags. This is what the analytical derivation predicts. When $\Phi_{12} = 0$ we have $\omega = 0$, hence $\theta = \Phi_{22}$ exactly, so the moving average root cancels the autoregressive root contributed by the omitted state and the retained variable follows a finite-order AR(1) process with no omitted variable bias.

Under weak spillovers, which capture the local-to-zero feedback, the strict lag-1 cutoff is broken. While

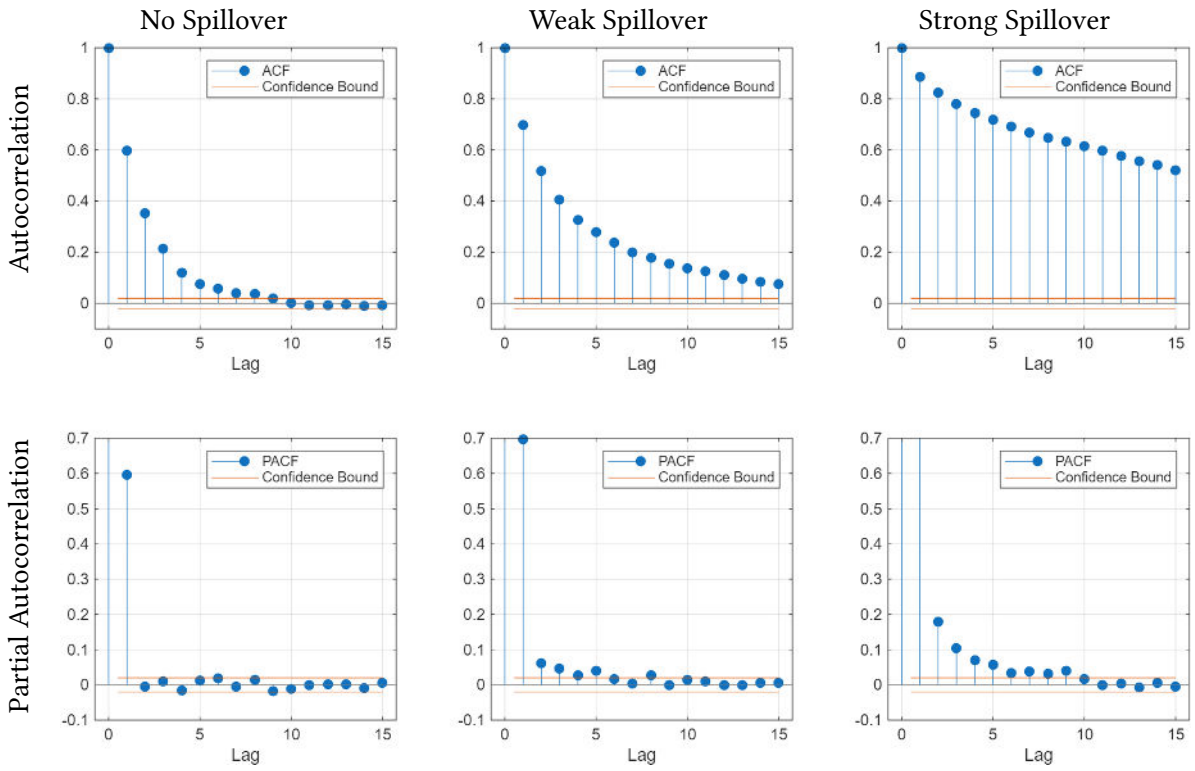


Figure E.VIII: Autocorrelation Functions For Different Spillover Levels

the first lag remains dominant, the PACF now exhibits a small, geometric decay across lags 2 through 5 and beyond. This trailing tail is consistent with the $O(T^{-1/2})$ dynamic misspecification drift derived above, although a single sample size cannot of course reveal a rate. In small samples, standard information criteria are highly likely to overlook these subtle trailing lags, leading an econometrician to erroneously fit a parsimonious VAR(1) and inadvertently absorb this MA-induced bias.

Finally, under strong spillovers where the cross-equation feedback grows, the infinite-order autoregressive nature of the marginalized process becomes apparent to the eye. The ACF shows heavily amplified persistence governed by the hidden state ($\Phi_{22} = 0.8$). Correspondingly, the PACF shows a prominent, deeply persistent tail. The observable system remains the same ARMA(2,1) process as in the weak-spillover case, but the two roots are now far from cancelling, so the autoregressive tail is long.

In summary, Figure E.VIII is consistent with the analytical framework. Omitting a persistent state variable explicitly generates moving average dynamics. When these unobserved spillovers are weak, they map directly into the $O(T^{-1/2})$ local-to-zero omitted lags that the proposed inference strategy is specifically designed to robustly bound.

F Performance Under Other Forms of Dynamic Misspecification

In the main text, the central reason behind dynamic misspecification has been motivated to be a lag truncation problem with respect to a larger finite-order VAR. In practice, finite-order VARs may suffer from dynamic misspecification due to other reasons. For instance, in Section 8 I argued that the omitted variable problem leading to the price puzzle could be understood as dynamic misspecification, using the arguments in Appendix E. See, more generally, the arguments in Schorfheide (2005); González-Casasús and Schorfheide

(2026); Montiel Olea et al. (2026).

In this section, I numerically evaluate the performance of the proposed misspecification-robust approach when the true data-generating process is a VARMA local to a VAR. As such, dynamic misspecification appears in the form of an omitted moving average structure.

In particular, consider the following data-generating process:

$$y_t = \sum_{k=1}^{K} B_k y_{t-k} + u_t + \frac{\tilde{\delta}}{\sqrt{T}} \sum_{j=K+1}^{\infty} C_j u_{t-j}.$$

As in Section 7.3, the scalar $\tilde{\delta}$ indexes the degree of misspecification. We can equivalently write the system as:

$$\left(\mathrm{I} - \sum_{k=1}^{K} B_k L^k \right) y_t = \left(\mathrm{I} + \frac{\tilde{\delta}}{\sqrt{T}} \sum_{j=K+1}^{\infty} C_j L^j \right) u_t. \tag{F.1}$$

To fit the model into our problem, we seek a VAR(∞) representation of the form:

$$\left(\mathrm{I} - \sum_{k=1}^{\infty} \Pi_k L^k \right) y_t = u_t. \tag{F.2}$$

By equating the two expressions for u_t in (F.1) and (F.2), we establish the core identity:

$$\left(\mathrm{I} + \frac{\tilde{\delta}}{\sqrt{T}} \sum_{j=K+1}^{\infty} C_j L^j \right) \left(\mathrm{I} - \sum_{k=1}^{\infty} \Pi_k L^k \right) = \mathrm{I} - \sum_{k=1}^{K} B_k L^k.$$

To find the specific autoregressive matrices Π_k, I expand the polynomial multiplication and match coefficients:

$$\mathrm{I} - \sum_{k=1}^{\infty} \Pi_k L^k + \frac{\tilde{\delta}}{\sqrt{T}} \sum_{j=K+1}^{\infty} C_j L^j - \frac{\tilde{\delta}}{\sqrt{T}} \sum_{j=K+1}^{\infty} \sum_{k=1}^{\infty} C_j \Pi_k L^{j+k} = \mathrm{I} - \sum_{k=1}^{K} B_k L^k.$$

Discarding terms of order $o(1/\sqrt{T})$, which are asymptotically negligible in the sense made precise by Assumption C.1, we can thus write the VAR(∞) as

$$y_t = \sum_{k=1}^{K} B_k y_{t-k} + \frac{\tilde{\delta}}{\sqrt{T}} \sum_{j=K+1}^{\infty} \left\{ C_j - \sum_{s=1}^{\min\{K, j-K-1\}} C_{j-s} B_s \right\} y_{t-j} + u_t.$$

Suppose that $C_j = 0$ for all $j > R > K$, so that the true data-generating process is a VARMA(K,R). The autoregressive representation then terminates exactly at lag $K+R$, so the framework applies with $p_1 = K$ and $p_2 = R$, since in that case

$$y_t = \sum_{k=1}^{K} B_k y_{t-k} + \frac{\tilde{\delta}}{\sqrt{T}} \sum_{j=K+1}^{R+K} \left\{ C_j - \sum_{s=1}^{\min\{K, j-K-1\}} C_{j-s} B_s \right\} y_{t-j} + u_t.$$

F.1 Few Omitted MA Terms

I first specialize the results to the following data-generating process with $K = 1$ and $R = 2$:

$$y_t = By_{t-1} + u_t + \frac{\tilde{\delta}}{\sqrt{T}} C u_{t-2}, \quad B = \begin{bmatrix} 0.6 & 0.2 \\ 0.1 & 0.7 \end{bmatrix}, \quad u_t \sim N(0, \mathrm{I}_2),$$

where the drift matrix C is drawn once from a standard normal distribution, normalized to unit spectral norm and held fixed across replications, so that the estimand is non-random and $\tilde{\delta}$ measures the magnitude of the omitted moving average component on a scale comparable to the experiments of Section 7.[22] The simulation uses 2,000 replications, a burn-in of 100 observations, and the same nominal level of 95% and target as in Section 7.3, namely the response of y_1 to a unit reduced-form shock in y_1. It compares the same four procedures: the standard VAR(1) interval, the MR-VAR interval, the lag-augmented local projection with two augmenting lags and Eicker-White standard errors, and the delta-method interval from the unrestricted VAR(3). Under the VARMA(1,2) process above, the true IRF matrix at horizon h is given by

$$\Phi_h = \begin{cases} B & \text{if } h = 1, \\ B^h + \frac{\tilde{\delta}}{\sqrt{T}} B^{h-2} C & \text{if } h \geqslant 2. \end{cases}$$

Using the earlier arguments, the VARMA(1,2) can be written, up to $o(1/\sqrt{T})$ terms, as:

$$y_t = By_{t-1} + \frac{\tilde{\delta}}{\sqrt{T}} C y_{t-2} - \frac{\tilde{\delta}}{\sqrt{T}} CB y_{t-3} + u_t.$$

Thus, I apply the local misspecification framework with $p_1 = 1$ and $p_2 = 2$.

Figure F.IX reports the empirical coverage of 95% confidence intervals for the asymptotic bias of IRFs from a VAR(1), as well as the ratio between the bias and the true IRF in percentage terms. The results are evaluated across varying degrees of misspecification ($\tilde{\delta}$) and horizons of interest (h).

We can observe from the results in Figure F.IX that the omitted moving average term induces significant relative biases into the IRFs derived from the VAR(1) (bottom row), yet the proposed inference approach accurately captures their magnitude (top row). The bands maintain coverage close to the nominal level even in small samples. At horizon 1 the true IRF is not affected by the moving average drift, which enters at the second lag, so the relative bias is exactly proportional to $\tilde{\delta}$, reaching about 10% at $\tilde{\delta} = \pm 3$ for $T = 250$. From horizon 2 onwards the true response itself varies with $\tilde{\delta}$, which makes the relative bias asymmetric and largest at those configurations where the true response is closest to zero. Its profile across horizons is therefore not monotone: at $T = 250$ and $\tilde{\delta} = -3$ it runs -10%, 60%, 13% and -1% at horizons 1, 2, 4 and 6, and it is roughly twice as large at $T = 100$, where the same drift is a larger perturbation of the process. Both the absolute bias and the response die out at long horizons, and here their ratio does too, but nothing forces it to.

Figures F.X and F.XI report the empirical coverage probabilities and average lengths of nominal 95% con-

[22] Without the normalization the results are qualitatively unchanged but the design leaves the local neighborhood: the realized draw is dominated by its $(1, 1)$ element, and at $T = 100$ the true own response at $h = 2$ crosses zero inside the $\tilde{\delta}$ grid, which makes the relative bias below diverge.

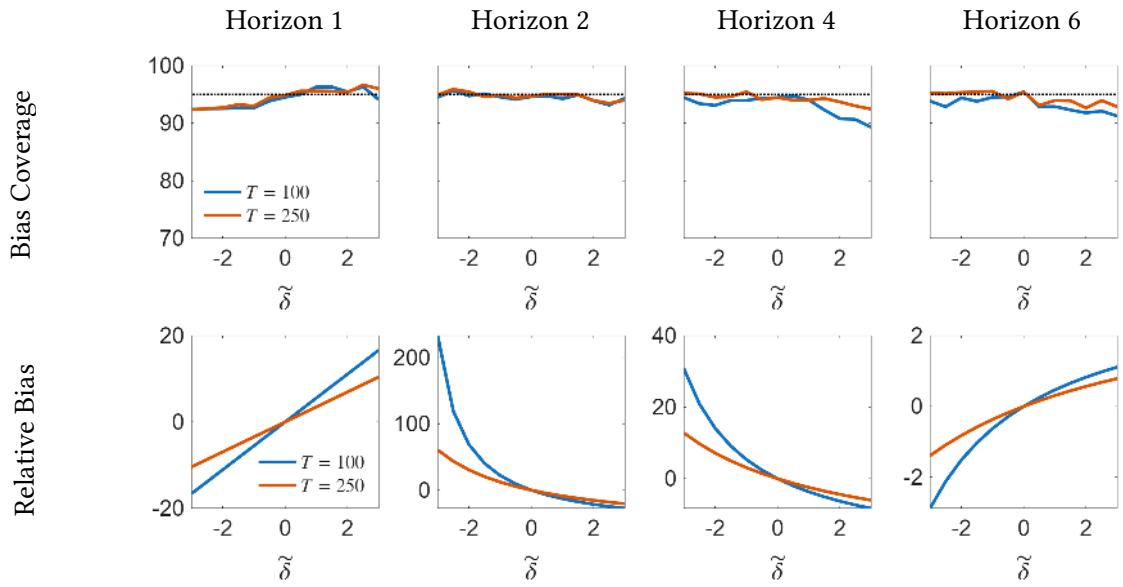


Figure F.IX: Dynamic Misspecification Bias Inference

fidence intervals for the IRFs with sample sizes $T = 100$ and $T = 250$, respectively.

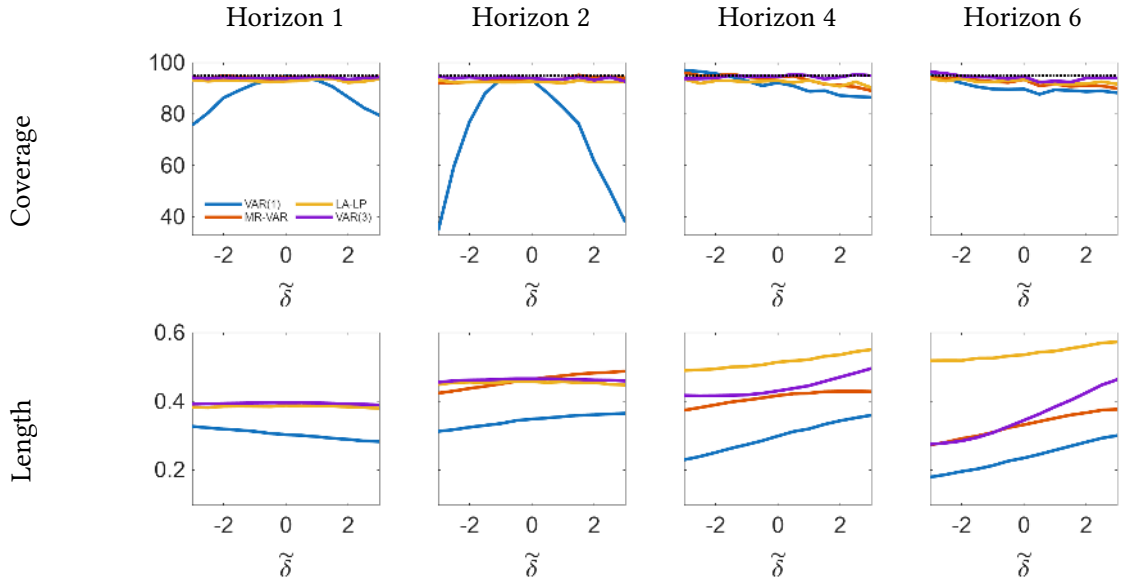


Figure F.X: Standard, Misspecification-Robust, Lag-Augmented LP and Unrestricted VAR Inference, $T = 100$

The results indicate that standard inference for a VAR(1), which ignores the dynamic misspecification due to the omitted moving average component, breaks down at the horizons where that misspecification distorts the response the most: at $T = 250$ its coverage falls to 74% at horizon 1 and to 42% at horizon 2, against a nominal 95%. By horizons 4 and 6 the bias has decayed together with the response, and standard inference is close to nominal again. This tracks the magnitudes in Figure F.IX. In contrast, the MR-VAR and LA-LP approaches, and the unrestricted VAR(3), all achieve valid inference across the entire grid. The three then separate on length. The MR-VAR and LA-LP intervals are indistinguishable at the first two horizons and diverge beyond them: averaging over the $\tilde{\delta}$ grid at $T = 250$, the MR-VAR intervals are 19% shorter at horizon 4 and 38% shorter at horizon 6, reflecting the serial correlation in the residuals of the multi-step

projection. The overparametrized VAR(3) remains an effective solution to control for dynamic misspecification, and its length is close to that of the MR-VAR intervals, which are shorter by about 6% at the two longest horizons when $T = 100$ and by 2% when $T = 250$. The relative benefits of MR-VAR over an overparametrized VAR thus shrink as T grows.

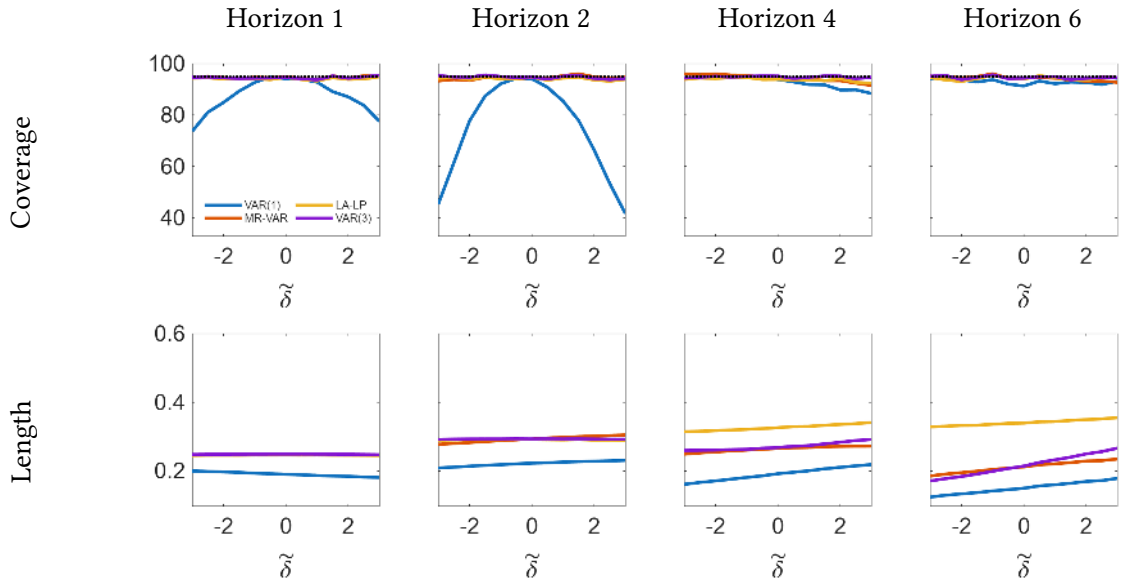


Figure F.XI: Standard, Misspecification-Robust, Lag-Augmented LP and Unrestricted VAR Inference, $T = 250$

F.2 Many Omitted MA Terms

Some macroeconomic series may contain more persistent moving average dynamics. I thus also specialize the results to the following data-generating process with $K = 1$ and $R = 8$:

$$y_t = By_{t-1} + u_t + \frac{\tilde{\delta}}{\sqrt{T}} \sum_{j=2}^{8} \rho^{j-2} C_j u_{t-j}, \quad B = \begin{bmatrix} 0.6 & 0.2 \\ 0.1 & 0.7 \end{bmatrix}, \quad \rho = 0.65, \quad u_t \sim N(0, \mathrm{I}_2).$$

The drift matrices $\{C_j\}$ are drawn once from a standard normal distribution, normalized to unit spectral norm and held fixed across replications, and the parameter ρ controls the strength of dependence on distant moving average lags. The estimation procedures, the target, the number of replications and the nominal level are as in the previous subsection, except that the unrestricted benchmark is now a VAR(9) and the framework is implemented with $p_1 = 1$ and $p_2 = 8$.

The results in Figure F.XII, specialized to $T = 250$, confirm that the MR-VAR approach remains valid when dynamic misspecification appears in the form of a persistent moving average component. The proposed confidence intervals for the bias maintain coverage between 91% and 96% across all horizons, and the MR-VAR bands for IRFs match the coverage of LA-LPs. What changes relative to the previous subsection is the length comparison, which is now essentially equalized: averaging over the $\tilde{\delta}$ grid, the two differ by at most about 2% at every horizon. The standard VAR(1) interval still undercovers substantially, falling to 79% at horizon 1 and 74% at horizon 2, over a $\tilde{\delta}$ grid that is narrower here, $[-2, 2]$ rather than $[-3, 3]$ for presentation purposes.

It is worth stressing some important limitations of the MR-VAR approach in this case. To accommodate

misspecification in the form of omitted moving average dynamics, I mapped a VARMA into a VAR to illustrate how to implement the MR-VAR approach. This mapping ignored $o(1/\sqrt{T})$ terms that were negligible in our asymptotic framework. However, these terms may be important in finite samples and cause some distortions for small T. In this case, the top row in Figure F.XII shows that such smaller order terms do not affect the validity of the robust inference approach. Furthermore, to obtain a clean mapping from a VARMA into a VAR, I imposed a restricted moving average structure in which the moving average coefficients for lags smaller than the autoregressive order are zero. This restriction is a presentational convenience rather than a requirement. Without it, the local moving average coefficients also perturb the first p_1 autoregressive matrices, which become $B_{k,T} = B_k + O(T^{-1/2})$ rather than B_k. The results are unaffected nevertheless. The data are then still a triangular array of the form covered by Assumption C.1, with an $O(1)$ limit for the first p_1 blocks, while the quantity $\delta = \sqrt{T}\text{vec}(\Theta_{T,p}S_0)$ continues to collect the last p_2 blocks alone, so Lemma 3 and the confidence sets built from it are unchanged. As for the bias, $\Theta_T(\mathrm{I}_{Kp} - S_m S_m')$ involves only the omitted blocks, so the $O(T^{-1/2})$ perturbation of the retained coefficients enters the estimator and the estimand in the same way and cancels from their difference, leaving the asymptotic bias $G_h C_m \delta$ of Lemma 1 intact. What the restriction buys is a closed-form mapping between the VARMA parameters and δ, which is what makes the true bias in this experiment computable. When the VARMA structure is unknown, that mapping is not available, even though inference on the bias remains valid. By contrast, the robustness properties of the LA-LP approach, as shown in Montiel Olea et al. (2026), do not require any such mapping.

There is a further and more consequential asymmetry, which the comparisons in this paper do not capture because the designs are built to isolate dynamic misspecification. Lag-augmented local projections are robust along three dimensions at once. First, dynamic misspecification: both procedures deliver valid inference, and the comparisons above concern only the price each pays in width. Second, persistence: the lag-augmented projection of Montiel Olea and Plagborg-Møller (2021) attains uniformly valid inference over persistence configurations that include a unit root and local-to-unity roots, whereas Assumptions 2-3 confine the analysis here to processes that are stationary for every T, and the $\sqrt{T}$-asymptotics behind Lemma 3 break down as the largest root approaches one. Third, conditional heteroskedasticity: the local projection is routinely paired with Eicker-White standard errors and inherits their robustness directly, while the variance $V = [S_0'\Omega^{-1}S_0] \otimes \Sigma$ used throughout is the homoskedastic one, and relaxing that requires the sandwich form of Remark 1, which is available but is not what the main text reports. The second of these dimensions is often the binding one in practice. In the application of Section 8, the largest companion eigenvalue of both the estimated VAR(4) and VAR(8) is 0.97, close enough to the unit circle that the stationary asymptotics should be regarded as an approximation. A researcher whose primary concern is near-unit-root persistence rather than lag truncation has good reason to prefer a lag-augmented projection, and the results of this paper are not an argument against doing so. What the framework here supplies, and what a local projection does not, is the interval for the misspecification bias of the parsimonious specification itself.

Finally, the derivations in the beginning of this section show that having a more persistent moving average structure requires controlling for a longer autoregressive structure, i.e., adding a larger number of p_2 terms. At that point, the parametric efficiency of VARs, namely from the MR-VAR approach, becomes futile compared to the LA-LP approach, and both approaches are roughly equivalent. This is precisely the class-of-departures trade-off discussed in Section 6.4. This limit equivalence as p_2 grows is also argued by

Montiel Olea et al. (2026). Nevertheless, the proposed confidence intervals for the bias remain a useful diagnostic tool.

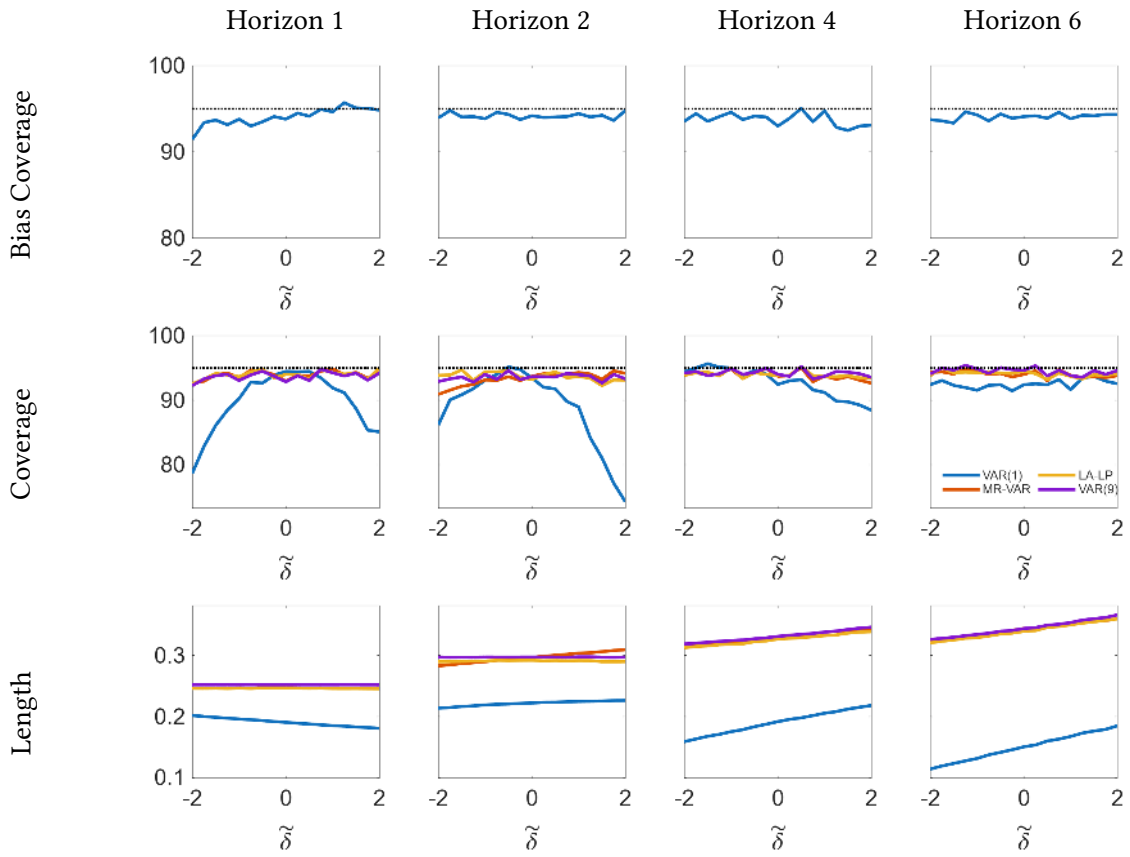


Figure F.XII: Inference Under Many Omitted Moving Average Terms, $T = 250$