

WORKING PAPER

GLS ESTIMATION OF LOCAL PROJECTIONS: TRADING ROBUSTNESS FOR EFFICIENCY

Ignace De Vos
Gerdie Everaert

May 2025
2024/1095

GLS Estimation of Local Projections: Trading Robustness for Efficiency

Ignace De Vos^{a,b} and Gerdie Everaert^{c,*}

^aVU Amsterdam, Department of Econometrics and Data Science

^bTinbergen Institute

^cGhent University, Department of Economics

Original version: October 10, 2024

This version: May 26, 2025

Abstract

Local projections (LPs) are widely used for estimating impulse responses and are often considered more robust to model misspecification than forward-iterated IRs from dynamic models such as VARs. However, this robustness comes at the cost of higher variance, particularly at longer horizons. To mitigate this trade-off, several GLS transformations of LPs have been proposed. This paper analyzes two broad strands of such GLS-type LP estimators: those that condition on residuals from an auxiliary VAR, and those that condition on residuals from previous-horizon LPs. We show that the former impose a VAR structure and reproduce VAR impulse responses, while the latter preserve the unrestricted nature of LPs and return the original LP OLS estimates. Consequently, the intended efficiency gains are either not achieved or come at the expense of the robustness that motivates the use of LPs in the first place, leaving the bias–variance trade-off unresolved.

JEL-codes: C22, C13, C53

Keywords: Impulse response functions, local projections, dynamic models, generalized least squares, efficiency, robustness

1 Introduction

Since the seminal work of Jordà (2005), local projections (LPs) have become a widely adopted method for estimating impulse responses (IRs). Unlike traditional VAR-based approaches that

*Corresponding author. E-mail: gerdie.everaert@ugent.be

We thank the associate editor and two anonymous referees for their constructive comments and suggestions, which greatly improved the paper. We also benefited from helpful feedback during presentations at various workshops and seminars. All remaining errors are our own.

extrapolate IRs from a limited number of sample autocovariances, LPs estimate them directly at each forecast horizon. This provides greater flexibility, as fewer restrictions are imposed on the data's dynamic structure. Jordà (2005) argues that this flexibility enhances the robustness of LPs to misspecification of the underlying data-generating process (DGP) compared to VAR-based methods. However, the extent of this robustness has been the subject of continued investigation in the literature.

Olea et al. (2024) offer a theoretical justification for Jordà's claim by showing that, within a general framework consistent with most linearized structural macroeconomic models, conventional LP confidence intervals retain correct asymptotic coverage even in the presence of substantial misspecification. In contrast, even minor deviations from the true DGP can lead to significant undercoverage in finite-order VAR-based confidence intervals. Kolesár and Plagborg-Møller (2024) further demonstrate that linear LPs retain a degree of robustness even when the true underlying model is nonlinear.

Nevertheless, LP estimates exhibit greater variability than VAR IRs and can sometimes appear erratic (Ramey, 2016). This variability arises because the LP method estimates each IR coefficient separately using ordinary least squares (OLS), thereby imposing fewer restrictions than the VAR-based approach, which extrapolates short-run dynamics across the entire IR horizon. Moreover, as the projection horizon increases, forecast errors accumulate, resulting in a moving average (MA) autocorrelation structure in the LP errors and thereby inflating their variance. This, in turn, amplifies uncertainty. Consequently, while LPs are often more robust to model misspecification, this robustness typically comes at the cost of reduced precision.

The trade-off between robustness and variability reflects a broader bias–variance spectrum in IR estimation. Plagborg-Møller and Wolf (2021) show that in population, LPs and VARs yield identical IRs when the VAR includes a sufficiently long lag length, implying equal robustness to misspecification in theory. In finite samples, however, lag selection reintroduces a trade-off: low-order VARs tend to have lower variance but can suffer from substantial bias if the lag length is insufficient, while LPs are more robust to misspecification but typically exhibit higher variance. Olea et al. (2024) similarly emphasize that robustness in VAR-based inference requires including enough lags to approximate LP behavior; reducing lag length may yield tighter confidence intervals, but at the expense of robustness. More broadly, LPs and VARs—along with regularized variants such as shrinkage estimators—can be viewed as points along a spectrum of methods that estimate the same IRs but differ in how they trade off bias and variance in finite samples.

Simulation studies offer practical insight into these theoretical trade-offs. Kilian and Kim (2011) show that when the data follow a finite-order VAR process, LP confidence intervals tend to be excessively wide, while bias-adjusted VAR bootstrap intervals are considerably narrower, making VAR-based methods preferable in such settings. Li et al. (2024) extend this analysis through large-scale simulations across thousands of DGPs and identification schemes. Their results confirm that LPs generally exhibit lower bias but higher variance than VAR-based estimators—particularly at

longer horizons—implying that LPs are preferable when minimizing bias is the primary concern, whereas VARs are more attractive when estimation precision is a priority.

Concerns over the finite-sample variability of LPs have motivated refinements aimed at improving their efficiency. In his seminal 2005 paper, Jordà already suggested that accuracy could potentially be enhanced by recursively incorporating previous-horizon projection errors as regressors in the current horizon projection. While the specifics were left for future work, this idea later inspired the development of Generalized Least Squares (GLS) transformations for LPs. Conditional on the Wold representation of the data being invertible into a VAR process, Lusompa (2023) shows that the autocorrelation pattern of LP errors can be traced to the dynamic structure of the VAR, motivating the use of GLS transformations to improve efficiency. Building on a general time series framework, Perron and González-Coya (2024) and Baillie et al. (2024) propose related approaches that, when applied to LPs, can be viewed as approximations of the method introduced by Lusompa (2023). Similarly, Breitung and Brüggemann (2023) propose a distinct GLS-style transformation that, while differing in implementation from Lusompa (2023), likewise seeks to improve the efficiency of the conventional LP OLS estimator.

However, since the GLS transformations proposed by Lusompa (2023) and Breitung and Brüggemann (2023) rely on a correctly specified auxiliary VAR—in the sense that the VAR projection errors are serially uncorrelated—the practical value of LP GLS remains uncertain. In an ideal scenario with no misspecification, it would be optimal to compute IR estimates directly from the correctly specified VAR model or to use a GLS estimator specifically designed to replicate those IRs. In that sense, Breitung and Brüggemann (2023) argue that one of their proposed GLS-type LP estimators is asymptotically equivalent to the iterative VAR approach. Despite this, Lusompa’s method has already found application in the recent work of Clark et al. (2024). Yet, in empirical contexts where misspecification is likely, the trade-off between lower variance and potential loss of robustness due to conditioning on a misspecified auxiliary VAR remains not fully understood.

This paper addresses this gap by evaluating how different versions of LP GLS perform relative to benchmark LPs and VAR-based IRs. To ensure that our results remain relevant across a wide range of empirical settings, we impose only weak regularity conditions, which allow for both local and global forms of misspecification. Accounting for such misspecification is essential to assess the practical value of GLS-based LP estimators. While earlier work has examined the absolute robustness of LPs under specific forms of misspecification, our focus is on the relative performance of LP GLS estimators, providing new insights into their effectiveness in managing the bias–variance trade-off.

We analyze two distinct strands of LP GLS estimators, each derived under a different assumption about whether the auxiliary VAR correctly represents the underlying DGP. The first strand exploits the fact that, under correct specification, LP errors follow a Vector Moving Average (VMA) process in terms of VAR residuals and impulse responses (see Lusompa, 2023). This motivates GLS transformations that use residuals from an auxiliary VAR to improve efficiency.

When fully implemented—meaning they condition on all available VAR residuals—these estimators reproduce forward-iterated VAR impulse responses, regardless of whether the VAR is correctly specified. This occurs because conditioning on all available VAR residuals effectively imposes the VAR’s full dynamic structure on the impulse response estimates. Accordingly, such GLS estimators should not be viewed as enhancing LP efficiency while preserving robustness, but rather as rebranded implementations of the VAR model.

The second strand is derived for settings where the auxiliary VAR is misspecified. In such cases, we show that LP errors follow a VMA process involving iteratively re-centered VAR projection errors and pseudo-true impulse responses. This motivates an alternative class of GLS estimators that instead condition on LP residuals from previous horizons, as originally suggested by Jordà. However, we show that these estimators coincide with LP OLS, either in large samples or numerically, and therefore do not deliver efficiency gains. Thus, these GLS estimators retain LP flexibility but fail to reduce estimation variance.

Importantly, these equivalence results hold under general nonparametric conditions and do not rely on a correct specification of the VAR. The limiting behavior of LP GLS estimators is driven by the structure of the residuals used in the GLS transformation. Conditioning on VAR residuals mechanically aligns the LP GLS estimator with VAR IRs, improving precision but sacrificing robustness—even when the VAR is misspecified. Conditioning on LP residuals, by contrast, leads to convergence to LP OLS, preserving robustness but forgoing efficiency gains. In this sense, LP GLS estimators replicate the performance of the benchmark estimator whose residuals they rely on, and therefore cannot outperform it in terms of both bias and variance. Accordingly, LP GLS fails to resolve the bias–variance trade-off.

The estimators proposed by Lusompa (2023) and Breitung and Brüggemann (2023) both rely on VAR residuals but differ in their implementation by excluding a specific residual from the full GLS transformation discussed above. Nevertheless, the estimator of Breitung and Brüggemann (2023) remains equivalent to the VAR IRs, as the excluded horizon-1 residual is orthogonal to the LP regressors. By contrast, the estimator of Lusompa (2023) excludes the current-horizon residual, which is not orthogonal to the regressors. This makes the estimator generally asymptotically distinct from both VAR and LP OLS, with its own asymptotic bias and variance. More broadly, the extent and nature of conditioning on VAR residuals determine where an estimator lies along the bias–variance spectrum: omitting all residuals recovers LP OLS; including all yields the VAR; and partial conditioning produces intermediate estimators, whose properties depend on which residuals are included, the DGP, and the forecast horizon.

While our general nonparametric framework ensures broad applicability, it leaves the projection error covariances unrestricted, rendering closed-form asymptotic distributions analytically intractable and of limited practical insight. For this reason, we complement the general results with structured DGPs that allow for sharper analytical and simulation-based comparisons. First, we analyze a stylized example with shrinking local misspecification—that is, a model where the

deviation from correct specification vanishes at rate $T^{-1/2}$ —which allows us to derive explicit expressions for asymptotic bias and variance. Second, we simulate data from the empirically calibrated DSGE model of Smets and Wouters (2005). Across these settings, the GLS estimator of Lusompa (2023) typically lies between LP OLS and VAR IRs along the bias–variance frontier, though it often tends to align more closely with the latter. However, it rarely dominates either benchmark in terms of weighted root mean squared error, suggesting limited benefits in applications. To illustrate how these insights translate to applied work, we conclude with an empirical application on the transmission of U.S. monetary policy shocks, using the external instrument and macroeconomic specification of Gertler and Karadi (2015).

The remainder of the paper is structured as follows. Section 2 introduces the general framework, defines benchmark VAR and LP estimators, and derives the autocorrelation structure of LP errors that underpins the construction of the various GLS transformations. Section 3 analyzes the bias–variance trade-off, showing how LP GLS estimators align with or deviate from the benchmark methods. Section 4 provides illustrative evidence from a stylized example, a DSGE-based simulation, and an empirical application to U.S. monetary policy shocks. Section 5 concludes. Proofs not included in the main text are provided in Appendix A and Appendix B.

2 Assumptions, Autocorrelation Processes, and Estimator Construction

This section establishes the framework for evaluating LP and VAR impulse response estimators. We begin by introducing the assumptions and the Wold representation of the data. We then define the benchmark VAR and LP estimators and analyze the autocorrelation structure of LP errors, building on the results of Lusompa (2023) for correctly specified VARs. These results motivate the first strand of LP GLS estimators, which condition on VAR residuals to mitigate the accumulation of projection errors. Finally, we extend the analysis to misspecified VARs and introduce a new theorem that justifies a second strand of LP GLS estimators based on LP residuals from previous horizons.

2.1 Assumptions and Wold Representation

Let \mathbf{y}_t denote a $(k \times 1)$ observed data vector. As in Plagborg-Møller and Wolf (2021), we impose the following nonparametric regularity condition:

Assumption 1. The data $\{\mathbf{y}_t\}$ are covariance stationary and purely non-deterministic, with an everywhere nonsingular spectral density matrix, absolutely summable Wold representation coefficients, and finite fourth moments. For notational convenience, we proceed as if $\{\mathbf{y}_t\}$ were a (strictly stationary) jointly Gaussian vector time series.

Note that the Gaussianity assumption is made without loss of generality. It allows us to replace linear projection operators with conditional expectations, but all results remain valid under Assumption 1 even in the absence of Gaussianity.

Under Assumption 1, \mathbf{y}_t has a Wold representation:

$$\mathbf{y}_t = \boldsymbol{\epsilon}_t + \sum_{j=1}^{\infty} \boldsymbol{\Theta}_j \boldsymbol{\epsilon}_{t-j}, \quad (1)$$

where $\boldsymbol{\epsilon}_t$ is a $(k \times 1)$ vector white noise process with $\mathbb{E}(\boldsymbol{\epsilon}_t) = \mathbf{0}_{k \times 1}$ and a positive definite covariance matrix $\boldsymbol{\Sigma}_\epsilon$.

The goal is to estimate IRs, which quantify the dynamic effects of each shock in $\boldsymbol{\epsilon}_t$ on \mathbf{y}_{t+h} , over the horizons $h = 1, \dots, H$, where H is fixed, finite, and satisfies $H < T$. Based on the Wold representation in eq.(1), the true IRs at these horizons are given by the sequence of $(k \times k)$ matrices $\{\boldsymbol{\Theta}_h\}_{h=1}^H$.

Remark 1. Assumption 1 imposes only mild regularity conditions and does not constrain the underlying causal structure of the data. Consequently, the IRs $\boldsymbol{\Theta}_h$ derived from the Wold representation in eq.(1) capture the effects of reduced-form shocks and, in general, do not correspond to responses to structural shocks. As noted by Plagborg-Møller and Wolf (2021), structural identification is inherently a population concept and is logically distinct from the choice of finite-sample estimation techniques. Therefore, our analysis mainly focuses on the relative performance of estimators for the reduced-form IRs. When structural IRs are of interest, they can often be obtained by post-multiplying the reduced-form IRs $\boldsymbol{\Theta}_h$ by an appropriate identification matrix \mathbf{B}^{-1} , such that the structural IRs are given by $\boldsymbol{\Theta}_h \mathbf{B}^{-1}$. The choice of \mathbf{B} depends on the identification strategy employed. Specific identification strategies are discussed in Plagborg-Møller and Wolf (2021), some of which are implemented in the empirical illustrations in Section 4.

2.2 Benchmark estimators: VAR and LP OLS

VAR. Consider the VAR model:

$$\mathbf{y}_{t+1} = \mathbf{A} \mathbf{y}_t + \boldsymbol{\epsilon}_{t+1}, \quad \text{for } t = 1, \dots, T-1, \quad (2)$$

where \mathbf{A} is a $(k \times k)$ parameter matrix, and $\boldsymbol{\epsilon}_{t+1} \equiv \mathbf{y}_{t+1} - \mathbb{E}[\mathbf{y}_{t+1} | \mathbf{y}_t]$ is a $(k \times 1)$ vector of projection errors.

The OLS estimator for \mathbf{A} in eq.(2), based on a sample of $T - a$ observations, is defined as:

$$\widehat{\mathbf{A}}_{(-a)} = \left(\sum_{t=1}^{T-a} \mathbf{y}_{t+1} \mathbf{y}_t' \right) \left(\sum_{t=1}^{T-a} \mathbf{y}_t \mathbf{y}_t' \right)^{-1}, \quad \text{for } 1 \leq a \leq T-1, \quad (3)$$

such that the corresponding VAR IR estimator for $\boldsymbol{\Theta}_h$ is given by $\widehat{\mathbf{A}}_{(-a)}^h$. Note that, to align with

some of the LP estimators introduced below, we allow for estimation over a reduced sample by excluding the last a observations. When using the maximum available sample ($a = 1$), we simplify the notation to $\widehat{A}_{(-1)} = \widehat{A}$.

Remark 2. The use of a VAR(1) model is purely for notational simplicity and does not restrict the lag length, as any finite-order VAR(p) can be expressed in its companion form as a VAR(1) with appropriately defined state variables. For this reason, we can generally refer to eq.(2) as a VAR model rather than specifically a VAR(1).

LP OLS. Local projections estimate the IRs Θ_h directly at each horizon through separate regressions:

$$\mathbf{y}_{t+h} = \mathbf{B}_h \mathbf{y}_t + \mathbf{e}_{t+h,h}, \quad \text{for } h = 1, \dots, H, \quad (4)$$

where \mathbf{B}_h represents the coefficients of the best linear projection of \mathbf{y}_{t+h} onto \mathbf{y}_t , and $\mathbf{e}_{t+h,h} = \mathbf{y}_{t+h} - \mathbb{E}[\mathbf{y}_{t+h} | \mathbf{y}_t] = \mathbf{y}_{t+h} - \mathbf{B}_h \mathbf{y}_t$ denotes the h -step-ahead projection error.

The OLS estimator for \mathbf{B}_h in eq.(4), based on a sample of $T - a$ observations, is given by:

$$\widehat{\mathbf{B}}_{h,(-a)} = \left(\sum_{t=1}^{T-a} \mathbf{y}_{t+h} \mathbf{y}_t' \right) \left(\sum_{t=1}^{T-a} \mathbf{y}_t \mathbf{y}_t' \right)^{-1}, \quad \text{for } h \leq a \leq T - 1. \quad (5)$$

In similar fashion to the VAR estimator above, we accommodate a flexible use of the sample and indicate it by the a subscript. Setting $a = h$ employs the maximum available sample at each horizon, in which case we simplify the notation to $\widehat{\mathbf{B}}_{h,(-h)} = \widehat{\mathbf{B}}_h$. Alternatively, setting $a = H$ uses the same dataset for the explanatory variable \mathbf{y}_t (i.e., $\mathbf{y}_1, \dots, \mathbf{y}_{T-H}$) across all horizons $h = 1, \dots, H$. This approach is often used in practice as it ensures a uniform sample size and composition, reducing variability that may arise from differing sample periods at each horizon.

2.3 Autocorrelation Structures of LP Errors

The accumulation of projection errors in $\mathbf{e}_{t+h,h}$ leads to serial correlation and causes the variance of the LP OLS estimator to increase with the horizon h . This phenomenon is reflected in the limiting distribution of the LP OLS estimator, as derived in Bhansali (1997) and Lusompa (2023), which shows that the variance grows with h . As such, understanding the autocorrelation structure of LP errors is crucial for analyzing the properties of LP estimators and developing GLS transformations aimed at avoiding the error accumulation process.

To that end, note that forward iteration of the VAR model in eq.(2) yields:

$$\mathbf{y}_{t+h} = \mathbf{A}^h \mathbf{y}_t + \sum_{j=1}^h \mathbf{A}^{h-j} \boldsymbol{\varepsilon}_{t+j}. \quad (6)$$

With eq.(6) we can analyze how the VMA structure in the term $\sum_{j=1}^h A^{h-j} \varepsilon_{t+j}$ shapes the autocorrelation structure of the LP errors. We consider two scenarios: when the VAR model is correctly specified and when it is misspecified.

2.3.1 Correctly Specified VAR

To formalize the conditions for correct specification of the VAR model in eq.(2), we introduce the following assumption:

Assumption 2. The Wold representation in eq.(1) is invertible.

This assumption ensures that the process \mathbf{y}_t admits a VAR representation, which forms the basis for defining correct specification in the VAR-based framework. Specifically, a correctly specified VAR model requires that the true DGP either follows a finite-order VAR or is well approximated by a truncated version of the infinite-order VAR obtained by inverting the Wold representation in eq.(1). When this condition holds, the VAR projection errors ε_t coincide with the Wold innovations ϵ_t .

Under Assumptions 1 and 2, the LP error $e_{t+h,h}$ can be expressed as a VMA process of VAR projection errors ε_{t+j} and IRs, where the VAR projection errors coincide with the Wold innovations ϵ_{t+j} . This representation, originally established in Theorem 1 of Lusompa (2023), takes the form:

$$e_{t+h,h}^{\text{inv}} = \Theta_{h-1} \varepsilon_{t+1} + \Theta_{h-2} \varepsilon_{t+2} + \cdots + \Theta_1 \varepsilon_{t+h-1} + \varepsilon_{t+h}. \quad (7)$$

Here, the superscript ‘inv’ indicates that this representation relies on the invertibility assumption (Assumption 2). This result provides the theoretical foundation for the GLS transformations of LPs based on VAR residuals, as presented in Subsection 2.4.1.

2.3.2 Misspecified VAR

A VAR is misspecified when it fails to fully capture the Wold representation of the underlying DGP. Two main types of misspecification can be distinguished: local and global.

Local misspecification arises when the VAR lag length is insufficient to approximate the true, possibly infinite-order VAR implied by an invertible Wold representation. As the sample size increases and the lag length grows accordingly, the approximation improves and the misspecification vanishes asymptotically. This type of misspecification is common when the DGP is approximated by a finite-order VAR due to degrees-of-freedom constraints or when the true process is MA-type and approximated by a VAR (see, e.g., Braun and Mitnik, 1993). Another example is measurement error: when the true series follows a finite-order VAR, classical i.i.d. measurement error induces an MA component in the observed data, leading to local misspecification that diminishes with longer lags (see, e.g., Stock and Watson, 2018).

Global misspecification, on the other hand, arises when the dynamics of the system cannot be recovered from the available data, even asymptotically. This typically occurs when the system is not invertible or when the number of structural shocks exceeds the number of observable variables. In such cases, no VAR based on observables, regardless of lag length, can recover the true structural dynamics. A prominent example involves anticipated shocks or ‘news shocks’, where agents respond to future policy before it materializes, leading to non-invertible MA structures, which prevent recovery of the structural shocks using only observable data (see, e.g., Sims, 2012; Leeper et al., 2013; Forni et al., 2019). Similarly, in DSGE models like Smets and Wouters (2005), omitting variables or shocks that drive the system can induce global misspecification because the remaining observables do not capture the full structure of the DGP.

In both cases, the VAR residuals ε_t exhibit serial correlation and differ from the Wold errors ϵ_t . Misspecification further implies that the LP parameters B_h generally do not coincide with the true IRs Θ_h (i.e., $B_h \neq \Theta_h$). However, following Galvao and Kato (2014), B_h can still be interpreted as a *pseudo-true* IR, defined as the best linear projection of \mathbf{y}_{t+h} onto \mathbf{y}_t . In this sense, B_h provides the optimal linear approximation to the true h -period-ahead response Θ_h , even if the underlying model is misspecified. Nonetheless, it is important to recognize that \hat{B}_h is not necessarily closer to Θ_h than \hat{A}^h , as discussed in Kilian and Kim (2011). The relative accuracy of these estimators depends on the degree and nature of misspecification, meaning that LPs do not universally dominate VAR-based estimators in terms of bias.

The following lemma provides an explicit expression for the pseudo-true IRs B_h :

Lemma 1. *Under Assumption 1, pseudo-true IRs B_h are given by:*

$$B_h = A^h + \sum_{j=1}^h A^{h-j} C_j,$$

where $C_j = \boldsymbol{\phi}_j \boldsymbol{\Gamma}^{-1}$, $\boldsymbol{\phi}_j = \mathbb{E}[\varepsilon_{t+j} \mathbf{y}_t']$ and $\boldsymbol{\Gamma} = \mathbb{E}[\mathbf{y}_t \mathbf{y}_t']$.

Proof. See Appendix A.

Lemma 1 shows that the LP coefficients B_h deviate from the VAR IRs A^h whenever $C_j \neq 0$ for some $j \leq h$. This occurs under VAR misspecification, where future VAR errors ε_{t+j} are correlated with current regressors \mathbf{y}_t . These forward dependence terms $\boldsymbol{\phi}_j = \mathbb{E}[\varepsilon_{t+j} \mathbf{y}_t']$ give rise to misspecification terms $C_j = \boldsymbol{\phi}_j \boldsymbol{\Gamma}^{-1}$, which accumulate over $j \leq h$ and generate a wedge between the LP-based pseudo-true IRs and the VAR-implied IRs. Intuitively, the deviation arises because LPs impose orthogonality at each horizon, whereas VARs impose it only through their finite lag structure.

An implication of Lemma 1 is that the LP errors $e_{t+h,h}$ cannot be written purely as a VMA process of Wold innovations and IRs but instead follow a more complex autocorrelation structure. The following theorem formalizes this structure under misspecified VARs.

Theorem 1. *Under Assumption 1, the horizon- h LP errors $e_{t+h,h}$ follow a VMA process of order $(h - 1)$,*

expressed as:

$$\mathbf{e}_{t+h,h} = \mathbf{B}_{h-1}\mathbf{v}_{t+1,1} + \mathbf{B}_{h-2}\mathbf{v}_{t+2,2} + \dots + \mathbf{B}_1\mathbf{v}_{t+h-1,h-1} + \mathbf{v}_{t+h,h}, \quad (8)$$

where the recursively defined re-centered VAR projection errors $\mathbf{v}_{s,j}$ are given by:

$$\mathbf{v}_{s,j} = \boldsymbol{\varepsilon}_s - \mathbf{C}_j\mathbf{y}_{s-j} - \sum_{\ell=1}^{j-1} \mathbf{C}_{j-\ell}\mathbf{v}_{s-j+\ell,\ell}. \quad (9)$$

Proof. We begin with substituting the expression for \mathbf{B}_h from Lemma 1 into the definition of the LP errors. This gives

$$\mathbf{e}_{t+h,h} = \mathbf{y}_{t+h} - \mathbf{B}_h\mathbf{y}_t = \mathbf{y}_{t+h} - \left(\mathbf{A}^h + \sum_{j=1}^h \mathbf{A}^{h-j}\mathbf{C}_j \right) \mathbf{y}_t = \sum_{j=1}^h \mathbf{A}^{h-j}(\boldsymbol{\varepsilon}_{t+j} - \mathbf{C}_j\mathbf{y}_t), \quad (10)$$

where the final equality makes use of eq.(6) to substitute \mathbf{y}_{t+h} . Using Lemma 1, \mathbf{A}^{h-j} can be expressed dynamically as

$$\mathbf{A}^{h-j} = \mathbf{B}_{h-j} - \sum_{\ell=j+1}^h \mathbf{A}^{h-\ell}\mathbf{C}_{\ell-j}. \quad (11)$$

Accordingly, setting $j = 1$ in eq.(11) allows substitution of \mathbf{A}^{h-1} into eq.(10) to obtain

$$\mathbf{e}_{t+h,h} = \mathbf{B}_{h-1}\mathbf{v}_{t+1,1} + \sum_{j=2}^h \mathbf{A}^{h-j}(\boldsymbol{\varepsilon}_{t+j} - \mathbf{C}_j\mathbf{y}_t - \mathbf{C}_{j-1}\mathbf{v}_{t+1,1}),$$

where $\mathbf{v}_{t+1,1} = \boldsymbol{\varepsilon}_{t+1} - \mathbf{C}_1\mathbf{y}_t$. In similar fashion, substituting \mathbf{A}^{h-2} using eq.(11) and continuing recursively in this way yields

$$\mathbf{e}_{t+h,h} = \mathbf{B}_{h-1}\mathbf{v}_{t+1,1} + \mathbf{B}_{h-2}\mathbf{v}_{t+2,2} + \dots + \mathbf{B}_1\mathbf{v}_{t+h-1,h-1} + \mathbf{v}_{t+h,h},$$

where each $\mathbf{v}_{t+j,j}$ is recursively defined as:

$$\mathbf{v}_{t+j,j} = \boldsymbol{\varepsilon}_{t+j} - \mathbf{C}_j\mathbf{y}_t - \sum_{\ell=1}^{j-1} \mathbf{C}_{j-\ell}\mathbf{v}_{t+\ell,\ell}.$$

□

Theorem 1 provides the basis for an alternative class of GLS estimators that condition on LP residuals from previous horizons, as originally suggested by Jordà (2005) and presented in Section 2.4.2.

2.4 LP GLS Estimators

In this section, we present the various GLS transformations of LPs. Since these estimators share the same structure and differ only in whether they condition on VAR or recursive LP residuals, they are nested in the following general expression:

$$\widehat{\mathbf{B}}_{h,(-a)}^{\text{GLS}} = \left(\sum_{t=1}^{T-a} (\mathbf{y}_{t+h} - \mathbf{\Psi}_{t,h}) \mathbf{y}'_t \right) \left(\sum_{t=1}^{T-a} \mathbf{y}_t \mathbf{y}'_t \right)^{-1}, \quad \text{for } h \leq a \leq T-1, \quad (12)$$

initialized with $\widehat{\mathbf{B}}_{1,(-a)}^{\text{GLS}} = \widehat{\mathbf{B}}_{1,(-a)}$ and maintaining the convention $\widehat{\mathbf{B}}_{0,(-a)}^{\text{GLS}} = \mathbf{I}_k$. The specification of the transformation term $\mathbf{\Psi}_{t,h}$ is what distinguishes the various GLS estimators, and we will consider several alternatives for it below. As before, the subscript $(-a)$ indicates that estimation is performed over a reduced sample of $T-a$ observations, and we again simplify notation to $\widehat{\mathbf{B}}_{h,(-h)}^{\text{GLS}} = \widehat{\mathbf{B}}_h^{\text{GLS}}$ when the maximum available sample is used at each horizon (such that $a = h$).

2.4.1 LP GLS Estimators Based on VAR Residuals

When the VAR is correctly specified, the VMA expression in eq.(7) allows the LP in eq.(4) to be written as:

$$\mathbf{y}_{t+h}^{\text{inv}} = \mathbf{B}_h \mathbf{y}_t + \mathbf{e}_{t+h,h}^{\text{inv}} = \mathbf{B}_h \mathbf{y}_t + \sum_{j=1}^h \mathbf{\Theta}_{h-j} \boldsymbol{\varepsilon}_{t+j},$$

with $\mathbf{\Theta}_0 = \mathbf{I}_k$. Estimates of $\boldsymbol{\varepsilon}_{t+j}$ are given by $\widehat{\boldsymbol{\varepsilon}}_{t+j,(-a)} = \mathbf{y}_{t+j} - \widehat{\mathbf{A}}_{(-a)} \mathbf{y}_{t+j-1}$ and are readily available from the VAR in eq.(2), while the IRs $(\mathbf{\Theta}_{h-1}, \dots, \mathbf{\Theta}_1)$ can be substituted with previous horizon LP estimates. This makes GLS transformations based on eq.(7) feasible. Multiple implementations are possible, each using a different conditioning set.

Lusompa (2023) proposes conditioning on the VAR projection errors $(\boldsymbol{\varepsilon}_{t+1}, \dots, \boldsymbol{\varepsilon}_{t+h-1})$ at horizon h , while excluding $\boldsymbol{\varepsilon}_{t+h}$. The corresponding feasible GLS estimator, $\widehat{\mathbf{B}}_{h,(-a)}^{\text{Lu}}$, is constructed iteratively by setting $\mathbf{\Psi}_{t,h} = \sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j,(-a)}^{\text{Lu}} \widehat{\boldsymbol{\varepsilon}}_{t+j,(-a)}$ in eq.(12).

Breitung and Brüggemann (2023) alternatively propose conditioning on $(\boldsymbol{\varepsilon}_{t+2}, \dots, \boldsymbol{\varepsilon}_{t+h})$, thereby excluding $\boldsymbol{\varepsilon}_{t+1}$. The corresponding feasible GLS estimator, $\widehat{\mathbf{B}}_{h,(-a)}^{\text{BB}}$, is constructed iteratively by setting $\mathbf{\Psi}_{t,h} = \sum_{j=2}^h \widehat{\mathbf{B}}_{h-j,(-a)}^{\text{BB}} \widehat{\boldsymbol{\varepsilon}}_{t+j,(-a)}$.¹

Since there is no compelling reason to exclude either $\boldsymbol{\varepsilon}_{t+1}$ or $\boldsymbol{\varepsilon}_{t+h}$ from the conditioning set at horizon h , we also consider an extended LP GLS estimator that conditions on the full set of VAR residuals $(\boldsymbol{\varepsilon}_{t+1}, \dots, \boldsymbol{\varepsilon}_{t+h})$. The corresponding feasible GLS estimator, $\widehat{\mathbf{B}}_{h,(-a)}^{\text{E}}$, is constructed

¹Breitung and Brüggemann (2023) propose transforming eq.(4) by moving $\widehat{\boldsymbol{\varepsilon}}_{t+h}$ to the left-hand side and including $(\widehat{\boldsymbol{\varepsilon}}_{t+2}, \dots, \widehat{\boldsymbol{\varepsilon}}_{t+h-1})$ as additional regressors. However, re-estimating the coefficients on these projection errors is unnecessary since they have already been estimated in previous LP horizons. To maintain alignment with the structure of the other LP GLS estimators, we implement their estimator by moving these residuals to the left-hand side without re-estimation.

iteratively by setting $\Psi_{t,h} = \sum_{j=1}^h \widehat{\mathbf{B}}_{h-j,(-a)}^\varepsilon \widehat{\boldsymbol{\varepsilon}}_{t+j,(-a)}$.

Remark 3. Although designed for a more general time series framework, the estimators proposed by Perron and González-Coya (2024) and Baillie et al. (2024)—when applied to an LP, one of their key examples—can be viewed as approximations to the approach in Lusompa (2023). While the latter directly implements a feasible GLS transformation based on the MA structure of the LP errors, Perron and González-Coya (2024) and Baillie et al. (2024) approximate the same transformation using an $\text{AR}(\infty)$ representation of the MA error process. This approximation is made feasible by truncating the AR expansion, resulting in estimators that are only approximately correct rather than an exact solution. Consequently, we do not explicitly consider these estimators.

2.4.2 LP GLS Based on LP Residuals

When the VAR is misspecified, the VMA expression in eq.(7) is no longer valid and must be replaced by the extended VMA expression provided in Theorem 1. Using this extended expression, the LP in eq.(4) can be written as:

$$\mathbf{y}_{t+h} = \mathbf{B}_h \mathbf{y}_t + \sum_{j=1}^{h-1} \mathbf{B}_{h-j} \mathbf{v}_{t+j,j} + \mathbf{v}_{t+h,h}.$$

By replacing the population coefficients \mathbf{B}_{h-j} and errors $\mathbf{v}_{t+j,j}$ for $j = 1, \dots, h-1$ with estimates from the previous horizons, a feasible GLS estimator, $\widehat{\mathbf{B}}_{h,(-a)}^v$, can naturally be constructed iteratively by setting $\Psi_{t,h} = \sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j,(-a)}^v \widehat{\mathbf{v}}_{t+j,j,(-a)}$.

3 GLS Estimation of LPs: Efficiency–Robustness Trade-Off

This section analyzes the trade-off between efficiency and robustness that underlies the LP GLS estimators introduced in Subsection 2.4, by examining whether they align with the efficient VAR estimator or retain the robustness of LP OLS. All results are derived under model misspecification (Assumption 1), but continue to hold under correct specification as a special case. We distinguish two strands of LP GLS estimators—those based on VAR residuals and those based on LP residuals—reflecting fundamentally different properties.

3.1 LP GLS Based on VAR Residuals

GLS estimation of LPs using VAR residuals achieves efficiency gains by incorporating aspects of the VAR dynamics into the LP framework. Notably, the LP GLS estimator $\widehat{\mathbf{B}}_h^\varepsilon$, which fully utilizes the VAR residuals, even numerically replicates the VAR IRs $\widehat{\mathbf{A}}^h$. This result is formalized in the following proposition.

Proposition 1. Under Assumption 1, the LP GLS estimator $\widehat{\mathbf{B}}_h^\varepsilon$, which fully incorporates VAR residuals in the GLS transformation, is numerically identical to the VAR IR estimator $\widehat{\mathbf{A}}^h$: $\widehat{\mathbf{B}}_h^\varepsilon = \widehat{\mathbf{A}}^h$ for all $h = 1, \dots, H$.

Proof. Consider that the estimated h -step-ahead forward iterated VAR is given by

$$\mathbf{y}_{t+h} = \widehat{\mathbf{A}}^h \mathbf{y}_t + \sum_{j=1}^h \widehat{\mathbf{A}}^{h-j} \widehat{\boldsymbol{\varepsilon}}_{t+j}. \quad (13)$$

Substituting this VAR expansion into the implemented GLS transformation $\mathbf{y}_{t+h}^\varepsilon = \mathbf{y}_{t+h} - \boldsymbol{\Psi}_{t,h}$ of eq.(12), with $\boldsymbol{\Psi}_{t,h} = \sum_{j=1}^h \widehat{\mathbf{B}}_{h-j}^\varepsilon \widehat{\boldsymbol{\varepsilon}}_{t+j}$, gives in turn:

$$\mathbf{y}_{t+h}^\varepsilon = \widehat{\mathbf{A}}^h \mathbf{y}_t + \sum_{j=1}^h \left(\widehat{\mathbf{A}}^{h-j} - \widehat{\mathbf{B}}_{h-j}^\varepsilon \right) \widehat{\boldsymbol{\varepsilon}}_{t+j}. \quad (14)$$

We can then proceed by mathematical induction to show that $\widehat{\mathbf{B}}_h^\varepsilon = \widehat{\mathbf{A}}^h$ for all $h = 1, \dots, H$. Note that from the initialization in eq.(12), it follows directly for $h = 1$ that $\widehat{\mathbf{B}}_1^\varepsilon = \widehat{\mathbf{B}}_1 = \widehat{\mathbf{A}}$, and eq.(14) implies that further equivalence at any h follows directly from that obtained at previous horizons. That is, assume as the (strong) inductive hypothesis for $h - 1$ that $\widehat{\mathbf{B}}_{h-j}^\varepsilon = \widehat{\mathbf{A}}^{h-j}$ for all $j = 1, \dots, h$ and any $h \geq 2$, and note that in this case the summation term in eq.(14) cancels, leaving $\mathbf{y}_{t+h}^\varepsilon = \widehat{\mathbf{A}}^h \mathbf{y}_t$. Accordingly, substituting this into the expression for the LP GLS estimator in eq.(12) gives as the result for h :

$$\widehat{\mathbf{B}}_h^\varepsilon = \left(\sum_{t=1}^{T-h} \mathbf{y}_{t+h}^\varepsilon \mathbf{y}_t' \right) \left(\sum_{t=1}^{T-h} \mathbf{y}_t \mathbf{y}_t' \right)^{-1} = \left(\sum_{t=1}^{T-h} \widehat{\mathbf{A}}^h \mathbf{y}_t \mathbf{y}_t' \right) \left(\sum_{t=1}^{T-h} \mathbf{y}_t \mathbf{y}_t' \right)^{-1} = \widehat{\mathbf{A}}^h.$$

We can therefore conclude that $\widehat{\mathbf{B}}_h^\varepsilon = \widehat{\mathbf{A}}^h$ for all $h = 1, \dots, H$. □

Intuitively, the forward-iterated estimated VAR in eq.(13) demonstrates that by conditioning on $\widehat{\boldsymbol{\varepsilon}}_{t+1}, \dots, \widehat{\boldsymbol{\varepsilon}}_{t+h}$, the LP GLS transformation, $\mathbf{y}_{t+h} - \sum_{j=1}^h \widehat{\mathbf{B}}_{h-j}^\varepsilon \widehat{\boldsymbol{\varepsilon}}_{t+j}$, eliminates any error term on the right-hand side of the LP equation, which iteratively ensures that $\widehat{\mathbf{B}}_h^\varepsilon = \widehat{\mathbf{A}}^h$ by construction. This numerical equivalence holds regardless of the sample size or whether the VAR model is correctly specified. As a result, implementing the LP GLS estimator $\widehat{\mathbf{B}}_h^\varepsilon$ offers no additional value, as it merely replicates the VAR IRs $\widehat{\mathbf{A}}^h$.

The LP GLS estimator $\widehat{\mathbf{B}}_h^{\text{Lu}}$ proposed by Lusompa (2023)—hereafter LP GLS-Lu—is a variant of the $\widehat{\mathbf{B}}_h^\varepsilon$ estimator that conditions on all available VAR residuals except the current-horizon residual $\widehat{\boldsymbol{\varepsilon}}_{t+h}$. The following corollary outlines its properties relative to the LP OLS estimator, $\widehat{\mathbf{B}}_h$, and the VAR IR estimator, $\widehat{\mathbf{A}}^h$.

Corollary 1. Under Assumption 1, as $T \rightarrow \infty$, the LP GLS estimator $\widehat{\mathbf{B}}_h^{\text{Lu}}$ of Lusompa (2023) deviates

from the LP OLS and VAR IR estimators $\widehat{\mathbf{B}}_h$ and $\widehat{\mathbf{A}}^h$ as follows for $h = 2, \dots, H$:

$$\widehat{\mathbf{B}}_h^{\text{Lu}} = \widehat{\mathbf{B}}_h + \boldsymbol{\psi}_h^{\text{B}} + O_p(T^{-1/2}), \quad (15)$$

$$\widehat{\mathbf{B}}_h^{\text{Lu}} = \widehat{\mathbf{A}}^h + \boldsymbol{\psi}_h^{\text{A}} + O_p(T^{-1/2}), \quad (16)$$

where $\boldsymbol{\psi}_h^{\text{B}} = -\sum_{j=1}^{h-1} (\mathbf{B}_{h-j} + \boldsymbol{\psi}_{h-j}^{\text{B}}) \mathbf{C}_j$ and $\boldsymbol{\psi}_h^{\text{A}} = \mathbf{C}_h - \sum_{j=1}^{h-1} \boldsymbol{\psi}_{h-j}^{\text{A}} \mathbf{C}_j$, with $\boldsymbol{\psi}_1^{\text{B}} = \boldsymbol{\psi}_1^{\text{A}} = \mathbf{0}$, and \mathbf{C}_j as defined in Lemma 1.

Proof. See Appendix A.

Corollary 1 shows that when the VAR is misspecified—reflected in nonzero misspecification terms \mathbf{C}_j for some $j < h$ —the LP GLS-Lu estimator $\widehat{\mathbf{B}}_h^{\text{Lu}}$ asymptotically deviates from both the LP OLS estimator $\widehat{\mathbf{B}}_h$ and the VAR estimator $\widehat{\mathbf{A}}^h$. These deviations are captured by the terms $\boldsymbol{\psi}_h^{\text{B}}$ and $\boldsymbol{\psi}_h^{\text{A}}$, respectively. In contrast, under correct specification—when the VAR includes a sufficient number of lags such that the forward dependence terms $\boldsymbol{\phi}_j = \mathbb{E}[\boldsymbol{\varepsilon}_{t+j} \mathbf{y}'_t]$ vanish for all $j \geq 1$ —the misspecification terms \mathbf{C}_j are zero, and all three estimators consistently estimate the true IRs.

The deviation from LP OLS arises because $\widehat{\mathbf{B}}_h^{\text{Lu}}$ conditions on the intermediate-horizon VAR projection errors $(\boldsymbol{\varepsilon}_{t+1}, \dots, \boldsymbol{\varepsilon}_{t+h-1})$ in its GLS transformation, thereby partially imposing the VAR's dynamic structure. If these projection errors are correlated with the regressors \mathbf{y}_t , this induces a deviation $\boldsymbol{\psi}_h^{\text{B}}$ from the LP OLS estimator. The deviation from the VAR estimator arises because $\widehat{\mathbf{B}}_h^{\text{Lu}}$ does not condition on $\boldsymbol{\varepsilon}_{t+h}$, which prevents it from fully replicating the VAR dynamics. If $\boldsymbol{\varepsilon}_{t+h}$ is correlated with \mathbf{y}_t , this omission results in a nonzero deviation $\boldsymbol{\psi}_h^{\text{A}}$.

Corollary 1 establishes limit expressions of the LP GLS-Lu estimator relative to the LP OLS and VAR estimates, identifying the values to which it converges under general conditions. However, it does not characterize the full asymptotic distribution. Importantly, unlike the other LP GLS estimators considered in this paper—which are numerically or asymptotically equivalent to either VAR or LP OLS—the LP GLS-Lu estimator remains asymptotically distinct. Even when all three estimators are consistent for the same IRs, their asymptotic variances typically differ, reflecting persistent efficiency differences that remain relevant for inference. Moreover, asymptotic bias may still arise when convergence is only local or when the DGP induces slowly vanishing bias terms.

Given the generality of Assumption 1, any expression for the asymptotic distribution would necessarily involve unspecified projection error covariances, precluding a general characterization of the estimators' relative bias–variance properties. Nonetheless, based on the structure of the GLS transformation, we conjecture that the LP GLS-Lu estimator typically lies between LP OLS and VAR in terms of bias and efficiency: it is likely (i) less biased than VAR but more biased than LP OLS, and (ii) more efficient than LP OLS but less so than VAR. These patterns should be understood as general tendencies rather than universal results, since bias–variance trade-offs ultimately depend on the underlying DGP and projection horizon. While Lusompa (2023) does not formally prove that $\widehat{\mathbf{B}}_h^{\text{Lu}}$ is uniformly more efficient than LP OLS, the paper does illustrate

potential efficiency gains for an AR(1) model. To make these trade-offs more concrete, Section 4 turns to specific, empirically relevant DGPs that allow us to analytically derive or simulate the bias and variance of VAR, LP OLS, and LP GLS estimators.

Remark 4. Corollary 1 reflects a key nonparametric result from Plagborg-Møller and Wolf (2021), who show that when a VAR(p) is estimated and p lags are included as controls in the LP, the IR estimands of the VAR and LP methods coincide for projection horizons $h \leq p$, even if the VAR is misspecified. In line with this result, Corollary 1 shows that for $h \leq p$, the deviation terms ψ_h^A and ψ_h^B are zero, implying that the LP GLS-Lu estimator also coincides with the VAR and LP OLS estimators in this case. This equivalence arises because the VAR projection errors ε_{t+j} are, by construction, orthogonal to \mathbf{y}_t for all $j \leq p$, which implies that the misspecification terms C_j vanish for $j < h$. As noted in Remark 2, although our setup is expressed in terms of a VAR(1), it naturally accommodates higher-order VAR(p) models through their standard VAR(1) companion form.

Turning to the LP GLS estimator $\hat{\mathbf{B}}_h^{\text{BB}}$ proposed by Breitung and Brüggemann (2023), this is likewise a restricted version of $\hat{\mathbf{B}}_h^e$ as it conditions on all VAR residuals except the horizon-1 residual, $\hat{\varepsilon}_{t+1}$. The following corollary establishes its equivalence to the VAR IR estimator, $\hat{\mathbf{A}}^h$.

Corollary 2. *Under Assumption 1, the LP GLS estimator $\hat{\mathbf{B}}_{h,(-a)}^{\text{BB}}$ proposed by Breitung and Brüggemann (2023) exhibits the following properties relative to the VAR IR estimate $\hat{\mathbf{A}}_{(-a)}^h$, depending on the sample:*

(i) $a = H$: $\hat{\mathbf{B}}_{h,(-H)}^{\text{BB}} = \hat{\mathbf{A}}_{(-H)}^h$ for all $h = 1, \dots, H$.

(ii) $a = h$: $\hat{\mathbf{B}}_h^{\text{BB}} = \hat{\mathbf{A}}^h + O_p(T^{-1})$ for all $h = 2, \dots, H$.

Proof. See Appendix A.

Corollary 2 indicates that the LP GLS estimator proposed by Breitung and Brüggemann (2023) is numerically equivalent to the VAR IR estimator $\hat{\mathbf{A}}^h$ when both are estimated using a reduced sample of $T - H$ observations. This equivalence arises because, at each horizon h , the transformed LP error term equals $\hat{\varepsilon}_{t+1,(-H)}$, which is orthogonal to \mathbf{y}_t due to the properties of OLS estimation. When both estimators are computed using the longest available sample, the covariance between \mathbf{y}_t and the transformed LP error term $\hat{\varepsilon}_{t+1}$ is $O_p(T^{-1})$. In this case, $\hat{\mathbf{B}}_h^{\text{BB}}$ and $\hat{\mathbf{A}}^h$ are asymptotically equivalent as $T \rightarrow \infty$, as already established by Breitung and Brüggemann (2023). This result holds irrespective of whether the model is correctly specified. The rate of convergence is sufficiently fast to ensure that $\hat{\mathbf{B}}_h^{\text{BB}}$ shares the same asymptotic distribution as $\hat{\mathbf{A}}^h$. Consequently, $\hat{\mathbf{B}}_h^{\text{BB}}$ exhibits a lower variance than the LP OLS estimator $\hat{\mathbf{B}}_h$, but its asymptotic equivalence to $\hat{\mathbf{A}}^h$ highlights that this variance reduction is achieved by fully imposing the VAR specification across the entire forecast horizon.

3.2 LP GLS Based on LP Residuals

The LP GLS estimator using LP residuals adjusts for residual serial correlation by relying on the LP framework itself. The following proposition shows its equivalence to the LP OLS estimator.

Proposition 2. Under Assumption 1, the LP GLS estimator $\widehat{\mathbf{B}}_{h,(-a)}^v$, which fully incorporates LP residuals in the GLS transformation, satisfies the following properties relative to the LP OLS estimator $\widehat{\mathbf{B}}_{h,(-a)}$, depending on the employed sample:

- (i) $a = H$: $\widehat{\mathbf{B}}_{h,(-H)}^v = \widehat{\mathbf{B}}_{h,(-H)}$ for all $h = 1, \dots, H$.
- (ii) $a = h$: $\widehat{\mathbf{B}}_h^v = \widehat{\mathbf{B}}_h + O_p(T^{-1})$ for all $h = 2, \dots, H$.

Proof. The difference between the GLS estimator, $\widehat{\mathbf{B}}_{h,(-a)}^v$, and the OLS estimator, $\widehat{\mathbf{B}}_{h,(-a)}$, is given by:

$$\widehat{\mathbf{B}}_{h,(-a)}^v - \widehat{\mathbf{B}}_{h,(-a)} = - \sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j,(-a)}^v \left(\frac{1}{T} \sum_{t=1}^{T-a} \widehat{\mathbf{v}}_{t+j,j,(-a)} \mathbf{y}'_t \right) \left(\frac{1}{T} \sum_{t=1}^{T-a} \mathbf{y}_t \mathbf{y}'_t \right)^{-1}. \quad (17)$$

Case (i): $a = H$. When a reduced sample of $T - H$ observations is used, $\widehat{\mathbf{v}}_{t+j,j,(-H)}$ and \mathbf{y}'_t are orthogonal by construction. This orthogonality is a numerical property of OLS estimation in the transformed LPs, where \mathbf{y}_t serves as the explanatory variable and $\mathbf{v}_{t+j,j,(-H)}$ is the error term for $j = 1, \dots, h - 1$. Making use of this orthogonality in eq.(17) shows that $\widehat{\mathbf{B}}_{h,(-H)}^v = \widehat{\mathbf{B}}_{h,(-H)}$, proving part (i).

Case (ii): $a = h$. When the longest available sample of $T - h$ observations is used at each horizon h , $\widehat{\mathbf{v}}_{t+j,j}$ and \mathbf{y}_t are orthogonal by construction over the sample period $t = 1, \dots, T - j$, but the summation in the numerator of eq.(17) runs over $t = 1, \dots, T - h$ with $h > j$. Therefore, we can decompose the summation as follows:

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T-h} \widehat{\mathbf{v}}_{t+j,j} \mathbf{y}'_t &= \frac{1}{T} \left[\sum_{t=1}^{T-j} \widehat{\mathbf{v}}_{t+j,j} \mathbf{y}'_t - \sum_{\ell=1}^{h-j} \widehat{\mathbf{v}}_{T-h+\ell+j}^j \mathbf{y}'_{T-h+\ell} \right] \\ &= -\frac{1}{T} \sum_{\ell=1}^{h-j} \widehat{\mathbf{v}}_{T-h+\ell+j}^j \mathbf{y}'_{T-h+\ell}, \end{aligned}$$

where the first term is zero due to orthogonality over $t = 1, \dots, T - j$. Noting that $\widehat{\mathbf{v}}_{T-h+\ell+j}^j \mathbf{y}'_{T-h+\ell} = O_p(1)$ for each ℓ , and that $h - j$ is a fixed, finite integer, it follows that:

$$\frac{1}{T} \sum_{t=1}^{T-h} \widehat{\mathbf{v}}_{t+j,j} \mathbf{y}'_t = O_p(T^{-1}). \quad (18)$$

From Lemma 2 in Appendix A, we know that $\left(\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{y}_t \mathbf{y}'_t \right)^{-1} = \mathbf{\Gamma}^{-1} + O_p(T^{-1/2})$, with $\mathbf{\Gamma} = \mathbb{E}(\mathbf{y}_t \mathbf{y}'_t) = O(1)$. Substituting this into eq.(17) and using the initial conditions $\widehat{\mathbf{B}}_0^v = \mathbf{I}_k$ and $\widehat{\mathbf{B}}_1^v = \widehat{\mathbf{B}}_1$, we iteratively find from using eq.(18) that:

$$\widehat{\mathbf{B}}_h^v - \widehat{\mathbf{B}}_h = O_p(T^{-1}),$$

for each $h = 2, \dots, H$. This proves part (ii). \square

The key insight of Proposition 2 is that the LP GLS estimator $\widehat{\mathbf{B}}_h^v$, which uses LP residuals, repli-

cates the LP OLS estimator $\widehat{\mathbf{B}}_h$. This result holds irrespective of whether the model is correctly specified. When the same reduced sample size of $T - H$ observations is used for each projection horizon h , the orthogonality property of OLS ensures numerical equivalence between the two estimators. When the longest available sample of $T - h$ observations is used for each projection horizon h , the orthogonality between the explanatory variable \mathbf{y}_t and the LP GLS error terms only holds up to an $O_p(T^{-1})$ term. This occurs because the LP is estimated on slightly different samples across horizons. Consequently, in this case, $\widehat{\mathbf{B}}_h^v$ and $\widehat{\mathbf{B}}_h$ are asymptotically equivalent as $T \rightarrow \infty$, with a rate of convergence that is sufficiently fast to ensure that $\widehat{\mathbf{B}}_h^v$ shares the same asymptotic distribution as $\widehat{\mathbf{B}}_h$. As a result, GLS estimation using LP residuals offers no practical advantages over standard LP OLS.

4 Illustrative Examples

To clarify the asymptotic properties of the LP GLS-Lu estimator—the only LP GLS variant that is asymptotically distinct from both LP OLS and VAR—we turn to specific DGPs that allow for analytical and simulation-based comparisons. We begin with a stylized example under shrinking local misspecification, where closed-form bias and variance expressions can be derived. We then extend this setting to a simulation design that retains the same DGP but replaces shrinking misspecification with lag selection via standard empirical criteria. Next, we examine whether the observed patterns persist under richer dynamics in simulations based on the DSGE model of Smets and Wouters (2005). Finally, we assess empirical relevance in an application to U.S. monetary policy using external instruments à la Gertler and Karadi (2015).

4.1 A Stylized Example of Local Misspecification

To complement the general framework of Section 3, we now consider a more specific setting that permits explicit analytical results. We adopt the local misspecification framework of Schorfheide (2005), Li et al. (2024), and Olea et al. (2024), in which the degree of misspecification vanishes at rate $T^{-1/2}$, allowing for a tractable fixed-lag asymptotic analysis. We also explore a closely related simulation design based on a similar DGP but using lag selection rules to determine model complexity. This more conventional empirical setup allows us to evaluate whether the results from the stylized example extend to more realistic conditions.

4.1.1 Local Misspecification via Vanishing MA Distortion

We consider the following autoregressive moving average (ARMA) process with a shrinking MA component:

$$w_{t+1} = \rho w_t + \beta \mu_{1,t} + \mu_{2,t+1} + \frac{\alpha}{\sqrt{T}} \mu_{2,t}, \quad (19)$$

where $|\rho| < 1$, $\boldsymbol{\mu}_t = (\mu_{1,t}, \mu_{2,t})'$ is a zero-mean i.i.d. white noise process with $\text{Var}(\boldsymbol{\mu}_t) = \text{diag}(\sigma_1^2, \sigma_2^2)$ and finite fourth moments. We assume that $\mathbf{y}_t = (\mu_{1,t}, w_t)'$ is observed, while $\mu_{2,t}$ remains latent.

As $T \rightarrow \infty$, the term α/\sqrt{T} shrinks to zero, causing the MA component $\mu_{2,t}$ to vanish asymptotically. In the limit, the DGP in eq.(19) converges to a stationary AR(1) process driven by the exogenous regressor $\mu_{1,t}$ and the innovation $\mu_{2,t+1}$. The observed process \mathbf{y}_t is then well approximated by a correctly specified VAR(1) model of the form in eq.(2), with

$$\mathbf{A} = \begin{pmatrix} 0 & 0 \\ \beta & \rho \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\varepsilon}_{t+1} = \begin{pmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} \end{pmatrix}.$$

This local misspecification setup captures the idea that finite-order VARs provide useful but imperfect representations of the true DGP in finite samples. By introducing a vanishing deviation from the VAR(1) benchmark, the framework delivers a tractable approximation that allows us to derive closed-form asymptotic distributions for estimators while retaining the essential features of the bias-variance trade-off caused by misspecification.

Our objective is to estimate the response of w_{t+h} , for $h \geq 1$, to a one-unit innovation in $\mu_{1,t}$. The true IR function is given by $\theta_h = \mathbf{e}'_2 \mathbf{A}^h \mathbf{e}_1 = \rho^{h-1} \beta$, for $h \geq 1$, where \mathbf{e}_j denotes the 2×1 unit vector with a one in position j and a zero in the other entry, for $j = 1, 2$.

Note that the shock $\mu_{1,t}$ enters eq.(19) with a one-period lag, such that it affects w_{t+1} rather than w_t . This ensures that the reduced-form IRs θ_h coincide with the structural IRs. The timing convention is without loss of generality: $\mu_{1,t}$ can always be interpreted—or recorded in the dataset—as a one-period lead of a structural shock, such that it contemporaneously affects the system while remaining exogenous. This allows for a structural interpretation of the reduced-form IRs without imposing additional identifying restrictions.

The estimators considered in Section 2 follow from specifying $\mathbf{y}_t = (\mu_{1,t}, w_t)'$. The VAR IR estimator for θ_h is given by $\hat{\theta}_h^{\text{VAR}} = \mathbf{e}'_2 \hat{\mathbf{A}}^h \mathbf{e}_1 = \hat{\rho}^{h-1} \hat{\beta}$, the LP OLS estimator by $\hat{\theta}_h^{\text{LP}} = \mathbf{e}'_2 \hat{\mathbf{B}}_h \mathbf{e}_1$ and the LP GLS-Lu estimator by $\hat{\theta}_h^{\text{Lu}} = \mathbf{e}'_2 \hat{\mathbf{B}}_h^{\text{Lu}} \mathbf{e}_1$, with $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}_h$ and $\hat{\mathbf{B}}_h^{\text{Lu}}$ defined in Section 2. We then obtain the following result:

Proposition 3. *Consider the DGP in eq.(19), with $|\rho| < 1$ and $\alpha \in \mathbb{R}$. Assume $\sigma_j^2 > 0$ and $\mathbb{E}(\mu_{j,t}^4) < \infty$*

for $j = 1, 2$. Define $\sigma_w^2 = (\beta^2\sigma_1^2 + \sigma_2^2)/(1 - \rho^2)$. Then, as $T \rightarrow \infty$

$$\sqrt{T} \left(\hat{\theta}_h^{\text{est}} - \theta_h \right) \xrightarrow{d} \mathcal{N}(b_h^{\text{est}}, V_h^{\text{est}}),$$

for $\text{est} \in \{\text{VAR}, \text{LP}, \text{Lu}\}$, where the asymptotic bias and variance terms b_h^{est} and V_h^{est} are given as follows.

For all $h \geq 1$,

$$\begin{aligned} b_h^{\text{VAR}} &= (h-1) \rho^{h-2} \beta \frac{\alpha \sigma_2^2}{\sigma_w^2}, & V_h^{\text{VAR}} &= \rho^{2(h-1)} \frac{\sigma_2^2}{\sigma_1^2} + (h-1)^2 \frac{\sigma_2^2}{\sigma_w^2} \rho^{2(h-2)} \beta^2, \\ b_h^{\text{LP}} &= 0, & V_h^{\text{LP}} &= \left(1 - \rho^{2h}\right) \frac{\sigma_w^2}{\sigma_1^2} - \rho^{2(h-1)} \beta^2. \end{aligned}$$

For all $h \geq 2$,

$$\begin{aligned} b_h^{\text{Lu}} &= (h-2) \rho^{h-2} \beta \frac{\alpha \sigma_2^2}{\sigma_w^2}, \\ V_h^{\text{Lu}} &= \left(1 + \rho^{2(h-1)}\right) \frac{\sigma_2^2}{\sigma_1^2} + \left(1 + h(h-2) \frac{\sigma_2^2}{\sigma_w^2}\right) \rho^{2(h-2)} \beta^2 + (h-2)^2 \rho^{2(h-3)} \beta^4 \frac{\sigma_1^2}{\sigma_w^2}. \end{aligned}$$

For $h = 1, 2$, we have $\hat{\theta}_h^{\text{Lu}} = \hat{\theta}_h^{\text{LP}}$.

Proof. See Appendix B (Online Supplementary Material).

When $\rho \neq 0$ and $\beta \neq 0$, Proposition 3 establishes the following ranking in terms of bias magnitude and variance:

$$|b_h^{\text{VAR}}| > |b_h^{\text{Lu}}| > |b_h^{\text{LP}}| = 0, \quad V_h^{\text{VAR}} < V_h^{\text{Lu}} < V_h^{\text{LP}}, \quad \text{for } h > 2.$$

The bias rankings reflect that the VAR estimator fully imposes the misspecified dynamic structure, thereby inducing the largest bias, while LP OLS remains unbiased in this setting because the misspecification term is not correlated with $\mu_{1,t}$. The LP GLS-Lu estimator partially imposes the VAR structure, resulting in an intermediate bias. Similarly, the efficiency ranking follows from the degree of structure that is imposed: the VAR estimator achieves the highest efficiency by fully exploiting model structure, LP GLS-Lu partially reduces noise relative to LP OLS, and LP OLS remains the most variable due to minimal restrictions. Figure 1 plots the asymptotic bias and standard deviation across horizons under low and high persistence ($\rho = 0.6$ and $\rho = 0.9$), confirming these trade-offs.

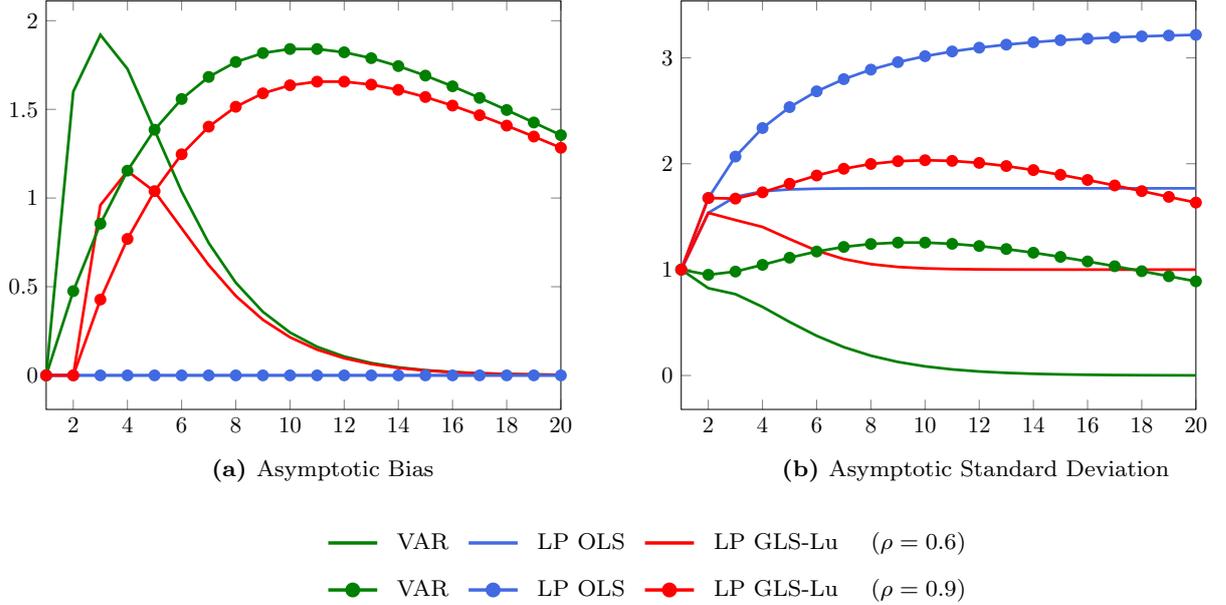
To assess overall performance, we compute a weighted RMSE for each estimator $\text{est} \in \{\text{VAR}, \text{LP}, \text{Lu}\}$:

$$\text{RMSE}_{\lambda, h}^{\text{est}} = \sqrt{\lambda (b_h^{\text{est}})^2 + (1 - \lambda) V_h^{\text{est}}}, \quad (20)$$

where $\lambda \in [0, 1]$ determines the weight placed on squared bias relative to variance.

Figure 2 displays the estimator that achieves the lowest weighted RMSE across projection hori-

Figure 1: Asymptotic Bias and Standard Deviation—Shrinking Local Misspecification



Notes: Reported are the asymptotic bias and standard deviation of the VAR, LP OLS, and LP GLS-Lu IRs estimators for the DGP in eq.(19), computed using the expressions in Proposition 3 under parameter values $\beta = \sigma_1^2 = \sigma_2^2 = 1$, $\rho \in \{0.6, 0.9\}$ and a misspecification term of $\alpha = 5$. The horizontal axis denotes the projection horizon $h = 1, \dots, 20$.

zons $h = 1, \dots, 20$ and bias weights $\lambda \in [0, 1]$. Color intensity reflects the strength of dominance, measured by the percentage RMSE reduction relative to the second-best estimator: darker shades indicate stronger dominance, while lighter shades reflect smaller gains. Black dots mark regions where specifically the LP GLS-Lu estimator ranks second-best. The results show that the preferred estimator depends on the weight assigned to bias: VAR dominates when bias is not weighted too heavily and at longer horizons, while LP OLS is favored when bias receives a high weight, particularly at shorter horizons. The LP GLS-Lu estimator typically ranks second-best and only occasionally emerges as the top performer, with rather minor RMSE improvements in those cases. Overall, under this stylized local misspecification, the LP GLS-Lu estimator offers no improvement over the benchmark VAR and LP OLS estimators.

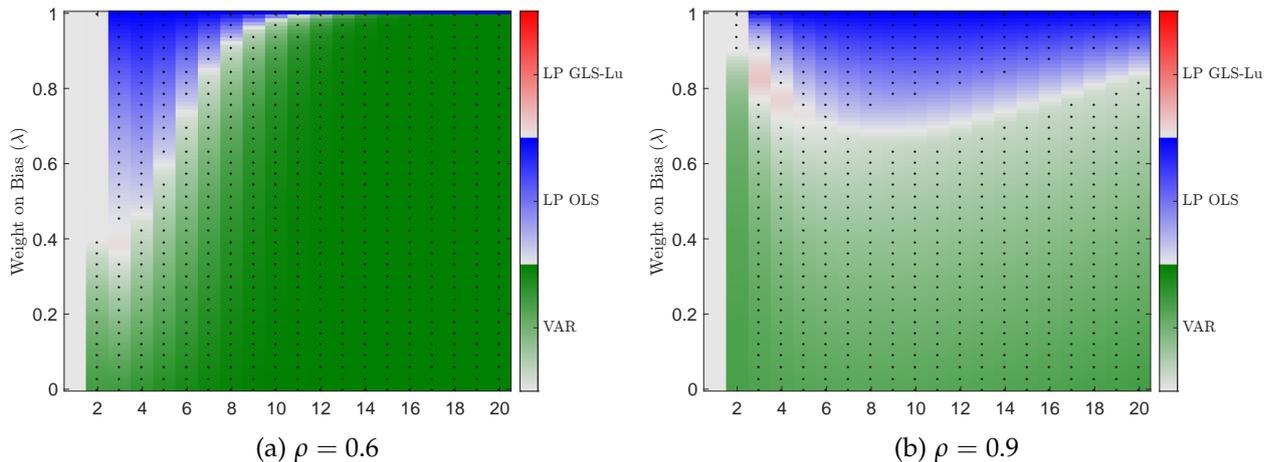
4.1.2 Local Misspecification with Data-Driven Lag Selection

To bridge the gap between the stylized setup and standard practice, we now consider a finite-sample simulation based on the following DGP:

$$w_{t+1} = \rho w_t + \beta \mu_{1,t} + \mu_{2,t+1} + \alpha \mu_{2,t-4}, \quad (21)$$

which replaces the vanishing misspecification term α / \sqrt{T} from eq.(19) with a fixed α . As a result, the DGP no longer converges to a finite-order VAR, reflecting the more realistic case where model misspecification persists in large samples. In practice, such complexity is typically addressed by selecting the lag length using data-driven rules, such as the Akaike Information Criterion (AIC), or by increasing lag length with sample size according to rule-of-thumb formulas like $p = \lfloor T^{1/4} \rfloor$.

Figure 2: Estimator Dominance by Weighted RMSE—Shrinking Local Misspecification



Notes: The heatmaps visualize estimator dominance across forecast horizons ($h = 1, \dots, 20$), plotted on the x-axis, and squared-bias weights ($\lambda \in [0, 1]$), plotted on the y-axis. Each cell color corresponds to the estimator—VAR IR, LP OLS, or LP GLS-Lu—minimizing the weighted RMSE defined in eq.(20), computed from data simulated from eq.(19) with parameters $\beta = \sigma_1^2 = \sigma_2^2 = 1$, $\rho \in \{0.6, 0.9\}$ and a misspecification term $\alpha = 5$. Color intensity reflects the relative dominance strength, measured as the percentage RMSE reduction compared to the second-best estimator: darker shades indicate stronger dominance, and lighter shades weaker dominance. Black dots highlight regions where LP GLS-Lu ranks second-best. For visual clarity, they are shown only every third weight step.

Simulation results for $T = 250$ are reported in Appendix B. Results are shown for both a low-persistence setting ($\rho = 0.6$) and a high-persistence setting ($\rho = 0.9$), with the misspecification parameter fixed at $\alpha = 0.5$. We present the \sqrt{T} -scaled bias and standard deviation of each estimator, along with heatmaps identifying the method that minimizes the weighted RMSE across projection horizons $h = 1, \dots, 20$ for a range of bias weights $\lambda \in [0, 1]$. Under AIC selection, the median lag is 3 for $\rho = 0.6$ and 1 for $\rho = 0.9$. To evaluate robustness, we also consider the rule-of-thumb lag length $p = \lfloor T^{1/4} \rfloor = 4$ and a larger fixed lag length $p = 8$. We apply the selected lag length uniformly across all three estimators.

The results confirm that the core features of the bias-variance trade-off persist: LP OLS remains less biased but more variable, while VAR is more precise but exhibits greater bias. The LP GLS-Lu estimator continues to interpolate between the two but tends to lie closer to VAR, with slightly reduced bias and slightly increased variance. It seldom outperforms either benchmark, and when it does, the RMSE gains are modest. The choice between data-driven and fixed lag length does not materially alter the qualitative ranking among estimators. Consistent with the findings of Plagborg-Møller and Wolf (2021) and the discussion in Remark 4, the estimated IRs align closely up to horizon $h = p$, but begin to diverge at longer horizons.

4.2 Simulations Based on the Smets and Wouters (2005) DSGE Model

To assess the finite-sample properties of the estimators in a realistic macroeconomic setting, we simulate data from the DSGE model developed by Smets and Wouters (2005). This model is widely recognized for its ability to capture key nominal and real rigidities underlying U.S. busi-

ness cycle fluctuations. We use Dynare (Adjemian et al., 2024) to solve the model at its estimated posterior mode and obtain its state-space representation, which includes seven structural shocks that propagate through twenty state variables, jointly driving the dynamics of seven observed macroeconomic indicators.

Following Olea et al. (2024), we focus on a subset of four variables from the simulated data— inflation, wages, hours worked, and the wage cost-push shock—and examine the dynamic response of inflation to the wage cost-push shock. In Smets and Wouters (2005), this shock follows an ARMA(1,1) process, implying that any finite-order VAR is inherently misspecified. However, because the shock is observed and included in the system, the misspecification is local: it results from approximating a process with VMA dynamics using a finite-lag VAR. As the lag length increases with the sample size, this approximation improves and the misspecification vanishes asymptotically.

To identify structural responses, we place the wage cost-push shock first in a recursive VAR, following standard practice. Reduced-form IRs are estimated using the VAR, LP OLS, and LP GLS-Lu estimators. Structural IRs are then obtained by post-multiplying the reduced-form responses with the Cholesky impact matrix from the VAR.

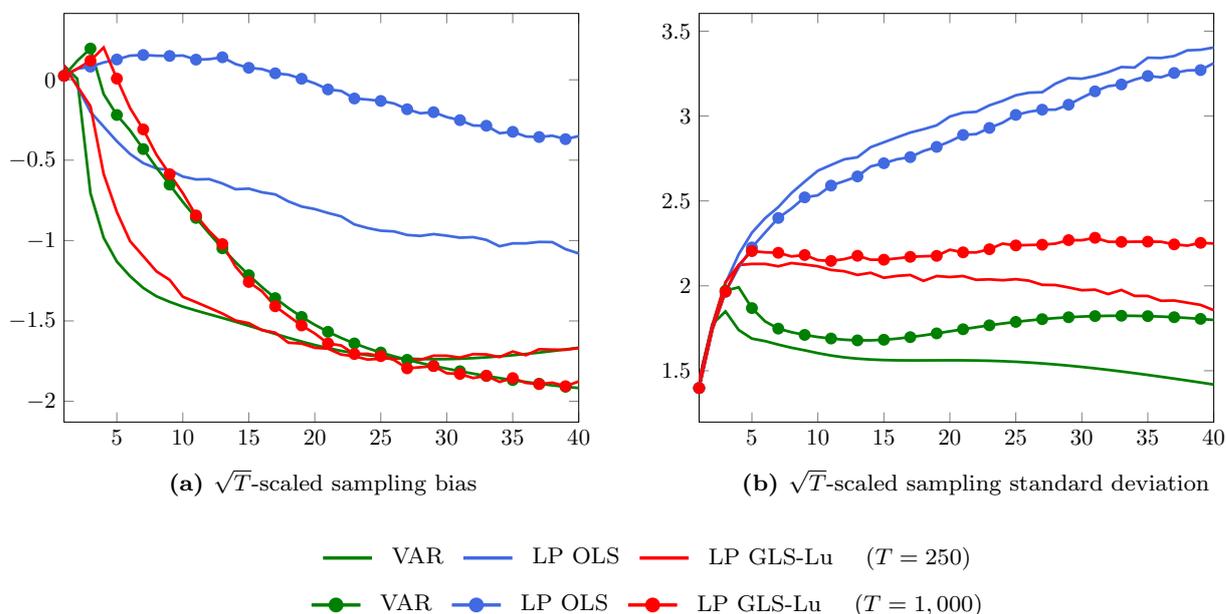
We consider two sample sizes, $T \in \{250, 1000\}$. The smaller sample ($T = 250$) reflects typical macroeconomic applications, while the larger one ($T = 1000$) allows us to assess how estimator performance evolves with increased sample size. The VAR lag order p is selected using the AIC, subject to a maximum of $\lfloor T^{1/4} \rfloor$. The same lag length is then applied to the LP OLS and LP GLS-Lu estimators. For $T = 250$, the median selected p is 2; for $T = 1000$, it increases to 3. Allowing the maximum lag order to grow at a faster rate or using alternative information criteria does not materially affect the results.

Figure 3 summarizes the \sqrt{T} -scaled bias and standard deviation of the VAR, LP OLS, and LP GLS-Lu estimators for the two considered sample sizes. As in the earlier results, all three estimators are highly similar at horizons shorter than or equal to the selected lag length, consistent with the findings of Plagborg-Møller and Wolf (2021) and the discussion in Remark 4. Beyond these horizons, and again in line with the analytical results presented earlier, the LP OLS estimator exhibits lower bias than VAR and LP GLS-Lu. The biases of VAR and LP GLS-Lu remain similar, although LP GLS-Lu shows a slightly lower bias at shorter horizons. In terms of variability, LP OLS consistently exhibits a higher standard deviation compared to both VAR and LP GLS-Lu, while LP GLS-Lu has a higher standard deviation than VAR.

Figure 4 visualizes the estimator achieving the lowest weighted RMSE across projection horizons $h = 1, \dots, 40$ and squared-bias weights $\lambda \in [0, 1]$. The VAR and LP OLS estimators are most frequently preferred: VAR dominates for moderate bias weights, while LP OLS is favored when bias receives a higher weight. The LP GLS-Lu estimator seldomly improves compared to the benchmarks, achieving the lowest weighted RMSE only in a few isolated cases, and then with minimal dominance. The dot-markers indicate that LP GLS-Lu tends to align more closely with

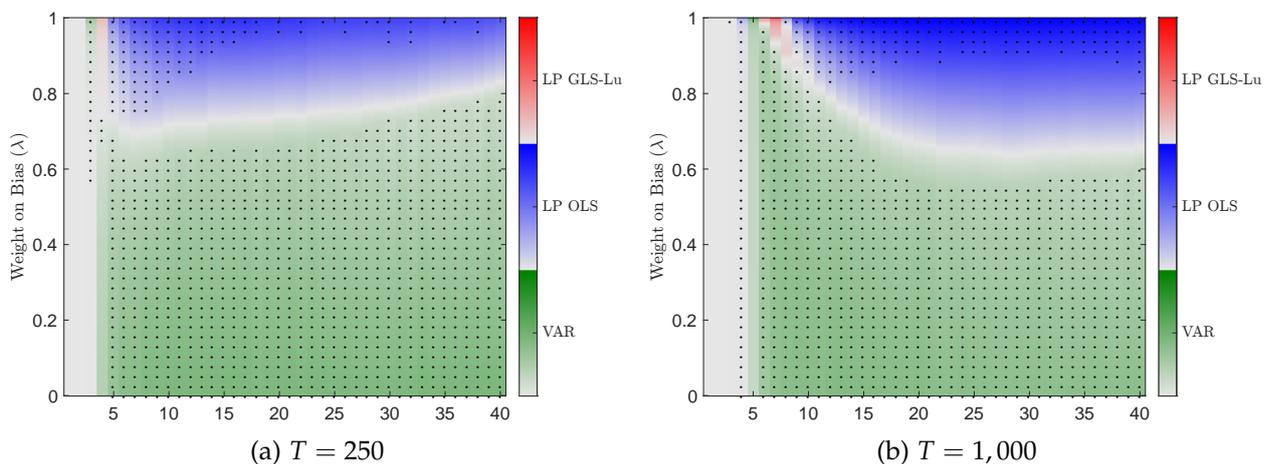
VAR, typically ranking second-best when VAR dominates. Conversely, when LP OLS is preferred, VAR is generally the runner-up. This pattern reflects the fact that LP GLS-Lu has a bias similar to VAR but generally exhibits higher variance. An exception occurs at shorter horizons when $T = 250$, where LP GLS-Lu has slightly lower bias than VAR and thus ranks second to LP OLS when bias is heavily weighted.

Figure 3: Scaled Bias and Standard Deviation—Smets-Wouters DSGE Model



Notes: Displayed are the \sqrt{T} -scaled bias and standard deviation, computed from 10,000 Monte Carlo replications based on data simulated from the Smets–Wouters DSGE model for $T \in \{250, 1,000\}$. The VAR lag length is selected using the AIC and applied uniformly across the VAR, LP OLS, and LP GLS-Lu estimators. The horizontal axis indicates the projection horizon $h = 1, \dots, 40$.

Figure 4: Estimator Dominance by Weighted RMSE—Smets-Wouters DSGE Model



Notes: Displayed are heatmaps of the estimator minimizing the weighted RMSE defined in eq.(20), computed from 10,000 Monte Carlo replications based on data simulated from the Smets–Wouters DSGE model for $T \in \{250, 1,000\}$. The VAR lag length is selected using the AIC and applied uniformly across the VAR, LP OLS, and LP GLS-Lu estimators. The horizontal axis indicates the forecast horizon $h = 1, \dots, 40$; the vertical axis varies the squared-bias weight $\lambda \in [0, 1]$. For interpretation of color shading and dots, see notes to Figure 2.

4.3 Empirical Application: Monetary Policy Transmission

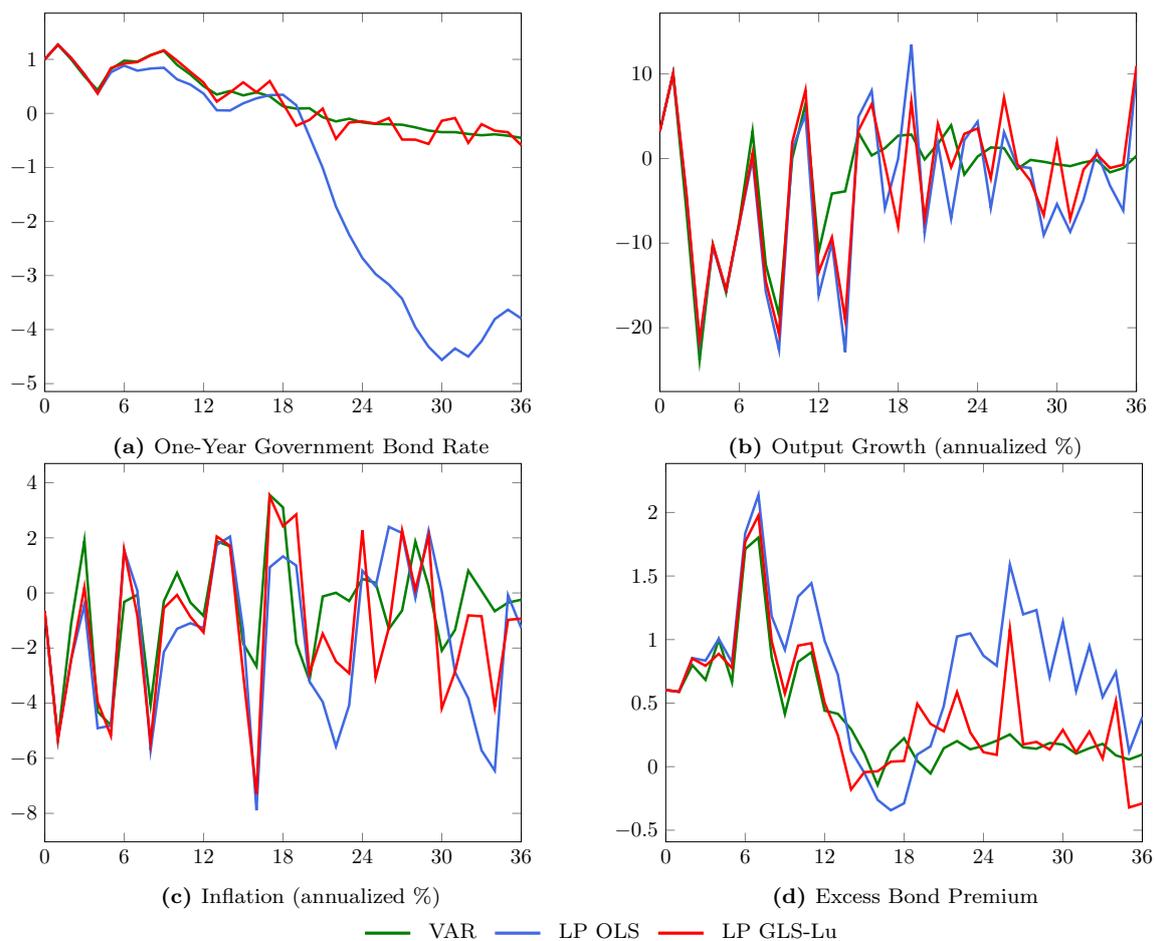
We conclude by illustrating our theoretical results with an application to the transmission of monetary policy shocks. We adopt the framework of Gertler and Karadi (2015), who use high-frequency financial data to construct an external instrument for monetary policy shocks. The vector of macroeconomic variables includes the one-year government bond rate, output growth (log growth rate of industrial production), inflation (log growth rate of the CPI), and the excess bond premium. The dataset is monthly and covers the period from January 1990 to June 2012.²

Following the empirical setup in Plagborg-Møller and Wolf (2021), we include the instrument as the first variable in a recursive VAR à la Ramey (2011), thereby treating it as an “internal instrument” rather than using a conventional SVAR-IV approach. This identification strategy exploits the instrument’s exogeneity and ensures its contemporaneous effect on the remaining variables. As in Section 4.2, we estimate reduced-form IRs using the VAR, LP OLS, and LP GLS-Lu estimators, and obtain structural IRs by post-multiplying the reduced-form responses with the Cholesky impact matrix from the VAR.

Estimated IRs using a fixed lag length of $p = 12$ are shown in Figure 5. Qualitatively similar results are obtained when setting $p = 3$, as selected by the AIC. Consistent with earlier findings, the LP GLS-Lu estimates appear less erratic than LP OLS but exhibit greater variability compared to the VAR IRs. It can be interpreted as a VAR-like estimator with additional variance arising from the fact that it does not condition on ε_{t+h} in the horizon- h projection. This distinction becomes particularly pronounced at longer horizons, where the LP OLS estimates tend to diverge, while the LP GLS-Lu estimates remain anchored to the VAR.

²For details on the construction and motivation of the instrument, see Gertler and Karadi (2015). We use their data, available at: <https://www.aeaweb.org/articles?id=10.1257/mac.20130329>.

Figure 5: Impulse Responses to a Monetary Policy Shock—Gertler-Karadi Application



Notes: Impulse responses to a monetary policy shock, identified using the high-frequency instrument from Gertler and Karadi (2015), and estimated on monthly U.S. data from January 1990 to June 2012. The shock is normalized to raise the one-year government bond rate by 100 basis points on impact. A lag length of 12 is used for all three estimators. The horizontal axis indicates the projection horizon $h = 1, \dots, 36$.

5 Conclusion

This paper critically evaluates the use of GLS transformations in the estimation of IRs via LPs. While GLS is often motivated by the goal of improving finite-sample efficiency, we show that its application entails a fundamental trade-off between efficiency and robustness—one that depends on the residuals used in the transformation. We distinguish two broad strands of LP GLS estimators: the first relies on residuals from an auxiliary VAR and imposes the dynamic structure of the VAR onto the LP framework; the second uses previous-horizon LP residuals to construct the transformation, aiming to retain the flexibility of the LP approach.

The central insight of our analysis is that these LP GLS estimators typically do not yield distinct asymptotic behavior, but instead replicate one of the benchmark estimators. Conditioning on VAR residuals leads to convergence to the VAR estimator, gaining precision under correct specification but losing robustness. Conditioning on LP residuals, by contrast, preserves robustness but offers no efficiency gains. These results hold under general nonparametric conditions and do not require assumptions about whether the auxiliary VAR is correctly specified or misspecified.

The only exception is the LP GLS estimator proposed by Lusompa (2023), which conditions on a subset of VAR residuals. This hybrid structure yields an estimator that is generally asymptotically distinct from both benchmarks. In a stylized local misspecification framework, we show that it strikes a balance between the bias of the VAR and the variance of LP OLS but rarely dominates either benchmark in terms of weighted mean squared error. These patterns persist in simulations based on the Smets and Wouters (2005) DSGE model and are echoed in an empirical application.

It is worth noting that the efficiency–robustness trade-off may still be mitigated by augmenting LPs with observed structural shocks instead of estimated residuals, as proposed by Faust and Wright (2013) and Teulings and Zubanov (2014). When these shocks are exogenous, such augmentations can preserve robustness while partially improving efficiency. Yet, since the observed shocks rarely span the full LP error term, a full GLS correction remains infeasible—unless residuals are substituted for unobserved components, which reintroduces the trade-off. Note also that Teulings and Zubanov (2014) primarily use this augmentation approach to address incidental parameter bias in panel LPs with fixed effects. Such bias-corrections are, however, beyond the scope of this paper and left for future work.

In conclusion, researchers should exercise caution when applying GLS to LPs. While GLS transformations may appear promising at first glance, most implementations either replicate the VAR or LP OLS benchmark, without resolving the bias–variance trade-off. Rather than providing a third alternative, LP GLS estimators merely shift the balance between robustness and efficiency depending on the residuals used.

References

- Adjemian, S., Juillard, M., Karamé, F., Mutschler, W., Pfeifer, J., Ratto, M., Rion, N., and Villemot, S. (2024). Dynare: Reference Manual, Version 6. Dynare Working Papers 80, CEPREMAP.
- Baillie, R. T., Diebold, F. X., Kapetanios, G., Kim, K. H., and Mora, A. (2024). On Robust Inference in Time Series Regression. *The Econometrics Journal*, Article ID: utae019.
- Bhansali, R. J. (1997). Direct Autoregressive Predictors For Multistep Prediction: Order Selection and Performance Relative to the Plug in Predictors. *Statistica Sinica*, 7(2):425–449.
- Braun, P. A. and Mittnik, S. (1993). Misspecifications in Vector Autoregressions and Their Effects on Impulse Responses and Variance Decompositions. *Journal of Econometrics*, 59(3):319–341.
- Breitung, J. and Brüggemann, R. (2023). Projection Estimators for Structural Impulse Responses. *Oxford Bulletin of Economics and Statistics*, 85(6):1320–1340.
- Clark, T. E., Huber, F., Koop, G., Marcellino, M., and Pfarrhofer, M. (2024). Investigating Growth-at-Risk Using a Multicountry Nonparametric Quantile Factor Model. *Journal of Business & Economic Statistics*, 42(4):1302–1317.
- Faust, J. and Wright, J. H. (2013). Efficient prediction of excess returns. *The Review of Economics and Statistics*, 95(4):1223–1233.
- Forni, M., Gambetti, L., and Sala, L. (2019). Structural VARs and Noninvertible Macroeconomic Models. *Journal of Applied Econometrics*, 34(2):221–246.
- Galvao, A. F. and Kato, K. (2014). Estimation and Inference for Linear Panel Data Models Under Misspecification When Both n and T are Large. *Journal of Business & Economic Statistics*, 32(2):285–309.
- Gertler, M. and Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.
- Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1):161–182.
- Kilian, L. and Kim, Y. (2011). How reliable are local projection estimators of impulse responses? *The Review of Economics and Statistics*, 93(4):1460–1466.
- Kolesár, M. and Plagborg-Møller, M. (2024). Dynamic Causal Effects in a Nonlinear World: The Good, the Bad, and the Ugly. Manuscript, arXiv:2411.10415v2.
- Leeper, E. M., Walker, T. B., and Yang, S.-C. S. (2013). Fiscal Foresight and Information Flows. *Econometrica*, 81(3):1115–1145.
- 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.

- Lusompa, A. (2023). Local Projections, Autocorrelation, and Efficiency. *Quantitative Economics*, 14(4):1199–1220.
- Olea, J. L. M., Plagborg-Møller, M., Qian, E., and Wolf, C. K. (2024). Double Robustness of Local Projections and Some Unpleasant VARithmetic. Manuscript, arXiv:2405.09509v2.
- Perron, P. and González-Coya, E. (2024). Feasible GLS for Time Series Regression. Manuscript, Department of Economics, Boston University.
- Plagborg-Møller, M. and Wolf, C. (2021). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*, 89(2):955–980.
- Ramey, V. A. (2011). Identifying Government Spending Shocks: It’s All in the Timing. *Quarterly Journal of Economics*, 126(1):1–50.
- Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, chapter 2, pages 71–162. Elsevier.
- Schorfheide, F. (2005). VAR Forecasting under Misspecification. *Journal of Econometrics*, 128(1):99–136.
- Sims, E. R. (2012). News, Non-Invertibility, and Structural VARs. In Balke, N., Canova, F., Milani, F., and Wynne, M., editors, *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments*, volume 28 of *Advances in Econometrics*, pages 81–135. Emerald Group Publishing Limited, Bingley, UK.
- Smets, F. and Wouters, R. (2005). Comparing Shocks and Frictions in US and Euro Area Business Cycles: a Bayesian DSGE Approach. *Journal of Applied Econometrics*, 20(2):161–183.
- Stock, J. H. and Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *The Economic Journal*, 128(610):917–948.
- Teulings, C. and Zubanov, N. (2014). Is Economic Recovery a Myth? Robust Estimation of Impulse Responses. *Journal of Applied Econometrics*, 29(3):497–514.

Appendix A Proofs and Supporting Results for Sections 2–3

A.1 Preliminaries: Notation and Useful Results

Throughout the proofs, all $O_p(\cdot)$ terms involving matrices or vectors are understood to be with respect to the spectral norm. For simplicity and clarity, explicit norm notation is omitted. Additionally, we assume that H is finite, ensuring that $h/T \rightarrow 0$ as $T \rightarrow \infty$ for all $h = 1, \dots, H$.

Starting from the VAR(1) model in eq.(2), estimated via OLS on the reduced sample $t = 1, \dots, T - a$, we derive the following backward iterated representations, which are useful for the proofs that follow:

$$\mathbf{y}_{t+h} = \widehat{\mathbf{A}}_{(-a)} \mathbf{y}_{t+h-1} + \widehat{\boldsymbol{\varepsilon}}_{t+h,(-a)}, \quad (\text{A-1})$$

$$= \widehat{\mathbf{A}}_{(-a)}^{h-1} \mathbf{y}_{t+1} + \sum_{j=2}^h \widehat{\mathbf{A}}_{(-a)}^{h-j} \widehat{\boldsymbol{\varepsilon}}_{t+j,(-a)}, \quad (\text{A-2})$$

$$= \widehat{\mathbf{A}}_{(-a)}^h \mathbf{y}_t + \sum_{j=1}^h \widehat{\mathbf{A}}_{(-a)}^{h-j} \widehat{\boldsymbol{\varepsilon}}_{t+j,(-a)}, \quad (\text{A-3})$$

where $\widehat{\mathbf{A}}_{(-a)}^0 = \mathbf{I}_k$. For notational simplicity, we adopt the convention of omitting the subscript $(-a)$ when using the full sample, corresponding to $a = 1$ when estimating the VAR(1) in eq.(2).

A.2 Additional Lemmas

Under Assumption 1, the following results hold as $T \rightarrow \infty$:

Lemma 2. $\widehat{\boldsymbol{\Gamma}}_{T-h} = \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{y}_t \mathbf{y}_t' = \boldsymbol{\Gamma} + O_p(T^{-1/2})$ and $\widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} = \boldsymbol{\Gamma}^{-1} + O_p(T^{-1/2})$, with $\boldsymbol{\Gamma} = \mathbb{E}(\mathbf{y}_t \mathbf{y}_t') = O(1)$.

Lemma 3. $\widehat{\mathbf{A}} = \mathbf{A} + O_p(T^{-1/2})$ and $\widehat{\mathbf{B}}_h = \mathbf{B}_h + O_p(T^{-1/2})$ for $h \geq 1$.

Lemma 4. $\widehat{\mathbf{A}}_{(-a)} = \widehat{\mathbf{A}} + O_p(T^{-1})$ for $2 \leq a \leq H$.

Lemma 5. Let $\widehat{\boldsymbol{\phi}}_j = \frac{1}{T} \sum_{t=1}^{T-h} \widehat{\boldsymbol{\varepsilon}}_{t+j} \mathbf{y}_t'$, where $\widehat{\boldsymbol{\varepsilon}}_{t+j}$ are the residuals from the VAR(1) model in eq.(2). Then:

$$\widehat{\boldsymbol{\phi}}_j = \begin{cases} \mathbf{0} & \text{for } j = 1, \\ \boldsymbol{\phi}_j + O_p(T^{-1/2}) & \text{for } j \geq 2, \end{cases}$$

where $\boldsymbol{\phi}_j = \mathbb{E}[\boldsymbol{\varepsilon}_{t+j} \mathbf{y}_t']$, with $\boldsymbol{\phi}_1 = \mathbf{0}$ and $\boldsymbol{\phi}_j = O(1)$ for $j \geq 2$.

A.3 Proofs of Lemmas

The proofs are organized for logical coherence. While Lemma 1 is stated first in the main text, its proof is placed last as it refers to notation established in Lemmas 2-5.

A.3.1 Proof of Lemma 2

Under Assumption 1, \mathbf{y}_t is a zero-mean, weakly dependent stationary series with finite fourth moments. The population second moment (covariance matrix) is given by:

$$\mathbf{\Gamma} = \mathbb{E} [\mathbf{y}_t \mathbf{y}_t'] = \text{Var}(\mathbf{y}_t) = O(1),$$

where $\mathbf{\Gamma} > 0$ (positive definite) by assumption.

From standard results for stationary time series with finite fourth moments, the sample covariance matrix satisfies

$$\hat{\mathbf{\Gamma}}_{T-h} = \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{y}_t \mathbf{y}_t' = \mathbf{\Gamma} + O_p(T^{-1/2}).$$

Because $\mathbf{\Gamma}$ is positive definite and bounded, its inverse $\mathbf{\Gamma}^{-1}$ is also $O(1)$. By the continuous mapping theorem for matrix inverses, it follows that:

$$\hat{\mathbf{\Gamma}}_{T-h}^{-1} = \mathbf{\Gamma}^{-1} + O_p(T^{-1/2}).$$

□

A.3.2 Proof of Lemma 3

The OLS estimator for \mathbf{B}_h in eq.(4) can be written as

$$\hat{\mathbf{B}}_h = \left(\sum_{t=1}^{T-h} \mathbf{y}_{t+h} \mathbf{y}_t' \right) \left(\sum_{t=1}^{T-h} \mathbf{y}_t \mathbf{y}_t' \right)^{-1} = \mathbf{B}_h + \left(\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{e}_{t+h,h} \mathbf{y}_t' \right) \hat{\mathbf{\Gamma}}_{T-h}^{-1}.$$

From the definition of $\mathbf{e}_{t+h,h} \equiv \mathbf{y}_{t+h} - \mathbb{E}(\mathbf{y}_{t+h} | \mathbf{y}_t)$ as a projection error, and under Assumption 1, which states that \mathbf{y}_t is a vector of stationary variables with finite fourth moments, we have

$$\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{e}_{t+h,h} \mathbf{y}_t' = O_p(T^{-1/2}).$$

From Lemma 2, we know that $\widehat{\Gamma}_{T-h}^{-1} = O_p(1)$. Substituting this into the expression for $\widehat{\mathbf{B}}_h$, we obtain:

$$\widehat{\mathbf{B}}_h - \mathbf{B}_h = \left(\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{e}_{t+h,h} \mathbf{y}'_t \right) \widehat{\Gamma}_{T-h}^{-1} = O_p(T^{-1/2}).$$

Finally, setting $h = 1$, we have $\widehat{\mathbf{B}}_1 - \mathbf{B}_1 = \widehat{\mathbf{A}} - \mathbf{A} = O_p(T^{-1/2})$. \square

A.3.3 Proof of Lemma 4

Consider the OLS estimator defined in eq.(3), estimated with $T - a$ observations:

$$\widehat{\mathbf{A}}_{(-a)} = \left(\frac{1}{T} \sum_{t=1}^{T-a} \mathbf{y}_{t+1} \mathbf{y}'_t \right) \left(\frac{1}{T} \sum_{t=1}^{T-a} \mathbf{y}_t \mathbf{y}'_t \right)^{-1} = \widehat{\gamma}_{T-a} \widehat{\Gamma}_{T-a}^{-1}, \quad \text{for } 1 \leq a \leq T-1.$$

For any fixed and finite $2 \leq a \leq H$, we can write:

$$\begin{aligned} \widehat{\Gamma}_{T-a} &= \frac{1}{T} \left[\sum_{t=1}^{T-1} \mathbf{y}_t \mathbf{y}'_t - \sum_{\ell=0}^{a-1} \mathbf{y}_{T-1-\ell} \mathbf{y}'_{T-1-\ell} \right] = \widehat{\Gamma}_{T-1} - \frac{1}{T} \sum_{\ell=0}^{a-1} \mathbf{y}_{T-1-\ell} \mathbf{y}'_{T-1-\ell} = \widehat{\Gamma}_{T-1} + O_p(T^{-1}), \\ \widehat{\gamma}_{T-a} &= \frac{1}{T} \left[\sum_{t=1}^{T-1} \mathbf{y}_{t+1} \mathbf{y}'_t - \sum_{\ell=0}^{a-1} \mathbf{y}_{T-\ell} \mathbf{y}'_{T-1-\ell} \right] = \widehat{\gamma}_{T-1} - \frac{1}{T} \sum_{\ell=0}^{a-1} \mathbf{y}_{T-\ell} \mathbf{y}'_{T-1-\ell} = \widehat{\gamma}_{T-1} + O_p(T^{-1}). \end{aligned}$$

Under Assumption 1, the terms $\sum_{\ell=0}^{a-1} \mathbf{y}_{T-1-\ell} \mathbf{y}'_{T-1-\ell}$ and $\sum_{\ell=0}^{a-1} \mathbf{y}_{T-\ell} \mathbf{y}'_{T-1-\ell}$ are $O_p(1)$ since a is fixed and finite. Substituting these results back into the expression for $\widehat{\mathbf{A}}_{(-a)}$, we have:

$$\widehat{\mathbf{A}}_{(-a)} = \widehat{\gamma}_{T-a} \widehat{\Gamma}_{T-a}^{-1} = \widehat{\gamma}_{T-1} \widehat{\Gamma}_{T-1}^{-1} + O_p(T^{-1}) = \widehat{\mathbf{A}} + O_p(T^{-1}).$$

\square

A.3.4 Proof of Lemma 5

Let $\widehat{\boldsymbol{\phi}}_j = \frac{1}{T} \sum_{t=1}^{T-h} \widehat{\boldsymbol{\varepsilon}}_{t+j} \mathbf{y}'_t$, where $\widehat{\boldsymbol{\varepsilon}}_{t+j}$ are the residuals from the VAR(1) model in eq.(2).

Consider first $j = 1$. By definition, $\boldsymbol{\varepsilon}_{t+1}$ is the projection residual of \mathbf{y}_{t+1} onto \mathbf{y}_t , which makes it orthogonal to \mathbf{y}_t and implies $\boldsymbol{\phi}_1 = \mathbb{E}[\boldsymbol{\varepsilon}_{t+1} \mathbf{y}'_t] = \mathbf{0}$. By the first order conditions of OLS, the orthogonality property similarly holds for the estimated $\widehat{\boldsymbol{\varepsilon}}_{t+1}$, such that the sample counterpart $\widehat{\boldsymbol{\phi}}_1$ also satisfies $\widehat{\boldsymbol{\phi}}_1 = \frac{1}{T} \sum_{t=1}^{T-h} \widehat{\boldsymbol{\varepsilon}}_{t+1} \mathbf{y}'_t = \mathbf{0}$.

For $j \geq 2$, substituting the definition of $\widehat{\boldsymbol{\varepsilon}}_{t+j}$ into $\widehat{\boldsymbol{\phi}}_j$ yields:

$$\widehat{\boldsymbol{\phi}}_j = \frac{1}{T} \sum_{t=1}^{T-h} \widehat{\boldsymbol{\varepsilon}}_{t+j} \mathbf{y}'_t = \frac{1}{T} \sum_{t=1}^{T-h} \left(\boldsymbol{\varepsilon}_{t+j} - (\widehat{\mathbf{A}} - \mathbf{A}) \mathbf{y}_{t+j-1} \right) \mathbf{y}'_t$$

$$\begin{aligned}
&= \frac{1}{T} \sum_{t=1}^{T-h} \varepsilon_{t+j} \mathbf{y}'_t - (\widehat{\mathbf{A}} - \mathbf{A}) \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{y}_{t+j-1} \mathbf{y}'_t \\
&= \frac{1}{T} \sum_{t=1}^{T-h} \varepsilon_{t+j} \mathbf{y}'_t + O_p(T^{-1/2})
\end{aligned}$$

where the last line makes use of $\widehat{\mathbf{A}} - \mathbf{A} = O_p(T^{-1/2})$ from Lemma 3 and $\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{y}_{t+j-1} \mathbf{y}'_t = O_p(1)$ under Assumption 1. Recalling that $\boldsymbol{\phi}_j = \mathbb{E} [\varepsilon_{t+j} \mathbf{y}'_t]$, and noting that ε_{t+j} and \mathbf{y}_t are covariance stationary with finite fourth moments under Assumption 1, it follows that:

$$\frac{1}{T} \sum_{t=1}^{T-h} \varepsilon_{t+j} \mathbf{y}'_t - \boldsymbol{\phi}_j = O_p(T^{-1/2}).$$

Combining these results gives:

$$\widehat{\boldsymbol{\phi}}_j = \boldsymbol{\phi}_j + O_p(T^{-1/2}) \quad \text{for } j \geq 2.$$

□

A.3.5 Proof of Lemma 1 (from the main text)

First note that from the definition of \mathbf{B}_h as the coefficients of the best linear projection of \mathbf{y}_{t+h} onto \mathbf{y}_t , we have:

$$\mathbb{E} [\mathbf{y}_{t+h} \mid \mathbf{y}_t] = \mathbf{B}_h \mathbf{y}_t. \tag{A-4}$$

Taking the conditional expectation given \mathbf{y}_t of the forward-iterated VAR representation in eq.(6), we have:

$$\mathbb{E} [\mathbf{y}_{t+h} \mid \mathbf{y}_t] = \mathbf{A}^h \mathbf{y}_t + \sum_{j=1}^h \mathbf{A}^{h-j} \mathbb{E} [\varepsilon_{t+j} \mid \mathbf{y}_t] = \left(\mathbf{A}^h + \sum_{j=1}^h \mathbf{A}^{h-j} \mathbf{C}_j \right) \mathbf{y}_t. \tag{A-5}$$

where we used that the conditional expectation of ε_{t+j} given \mathbf{y}_t is the linear projection of ε_{t+j} onto \mathbf{y}_t , given by $\mathbb{E} [\varepsilon_{t+j} \mid \mathbf{y}_t] = \mathbf{C}_j \mathbf{y}_t$, with $\mathbf{C}_j = \boldsymbol{\phi}_j \boldsymbol{\Gamma}^{-1}$, $\boldsymbol{\phi}_j = \mathbb{E} [\varepsilon_{t+j} \mathbf{y}'_t]$ and $\boldsymbol{\Gamma} = \mathbb{E} [\mathbf{y}_t \mathbf{y}'_t]$.

Hence, equating (A-5) to the projection definition in eq.(A-4) reveals that:

$$\mathbf{B}_h = \mathbf{A}^h + \sum_{j=1}^h \mathbf{A}^{h-j} \mathbf{C}_j.$$

□

A.4 Proofs for the Results in the Corollaries

A.4.1 Proof of eq.(15) in Corollary 1

By setting $\Psi_{t,h} = \sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \widehat{\boldsymbol{\varepsilon}}_{t+j}$ in eq.(12), and defining the shorthand $\widehat{\boldsymbol{\Gamma}}_{T-h} = \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{y}_t \mathbf{y}'_t$, the $\widehat{\mathbf{B}}_h^{\text{Lu}}$ estimator can be written as:

$$\begin{aligned} \widehat{\mathbf{B}}_h^{\text{Lu}} &= \left(\frac{1}{T} \sum_{t=1}^{T-h} \left(\mathbf{y}_{t+h} - \sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \widehat{\boldsymbol{\varepsilon}}_{t+j} \right) \mathbf{y}'_t \right) \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} \\ &= \widehat{\mathbf{B}}_h - \left(\sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \widehat{\boldsymbol{\phi}}_j \right) \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1}, \end{aligned} \quad (\text{A-6})$$

where we defined $\widehat{\boldsymbol{\phi}}_j = \frac{1}{T} \sum_{t=1}^{T-h} \widehat{\boldsymbol{\varepsilon}}_{t+j} \mathbf{y}'_t$. For $h = 1$, we have by definition that $\widehat{\mathbf{B}}_1^{\text{Lu}} = \widehat{\mathbf{B}}_1$, such that it follows from Lemma 3 that

$$\widehat{\mathbf{B}}_1^{\text{Lu}} = \mathbf{B}_1 + O_p(T^{-1/2}). \quad (\text{A-7})$$

Next, setting $h = 2$ in eq.(A-6) gives

$$\widehat{\mathbf{B}}_2^{\text{Lu}} = \widehat{\mathbf{B}}_2 - \widehat{\mathbf{B}}_1^{\text{Lu}} \widehat{\boldsymbol{\phi}}_1 \widehat{\boldsymbol{\Gamma}}_{T-2}^{-1}.$$

Making use of eq.(A-7), $\widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} = \boldsymbol{\Gamma}^{-1} + O_p(T^{-1/2})$ by Lemma 2 and $\widehat{\boldsymbol{\phi}}_j = \boldsymbol{\phi}_j + O_p(T^{-1/2})$ by Lemma 5, with $\boldsymbol{\phi}_j = O_p(1)$, then leads to

$$\widehat{\mathbf{B}}_2^{\text{Lu}} = \widehat{\mathbf{B}}_2 - \mathbf{B}_1 \boldsymbol{\phi}_1 \boldsymbol{\Gamma}^{-1} + O_p(T^{-1/2}).$$

Thus, using Lemma 3, we have $\widehat{\mathbf{B}}_2^{\text{Lu}} = \mathbf{B}_2 - \mathbf{B}_1 \boldsymbol{\phi}_1 \boldsymbol{\Gamma}^{-1} + O_p(T^{-1/2})$, which implies in turn that $\widehat{\mathbf{B}}_2^{\text{Lu}}$ remains $O_p(1)$.

Setting next $h = 3$ in eq.(A-6) gives

$$\widehat{\mathbf{B}}_3^{\text{Lu}} = \widehat{\mathbf{B}}_3 - \widehat{\mathbf{B}}_2^{\text{Lu}} \widehat{\boldsymbol{\phi}}_1 \widehat{\boldsymbol{\Gamma}}_{T-3}^{-1} - \widehat{\mathbf{B}}_1^{\text{Lu}} \widehat{\boldsymbol{\phi}}_2 \widehat{\boldsymbol{\Gamma}}_{T-3}^{-1}.$$

where substituting in the result for $\widehat{\mathbf{B}}_2^{\text{Lu}}$ and using the same lemmas as above results in

$$\widehat{\mathbf{B}}_3^{\text{Lu}} = \widehat{\mathbf{B}}_3 - \left(\mathbf{B}_2 - \mathbf{B}_1 \boldsymbol{\phi}_1 \boldsymbol{\Gamma}^{-1} \right) \boldsymbol{\phi}_1 \boldsymbol{\Gamma}^{-1} - \mathbf{B}_1 \boldsymbol{\phi}_2 \boldsymbol{\Gamma}^{-1} + O_p(T^{-1/2}).$$

which too is a bounded quantity.

Accordingly, by iterating the steps above for general $h > 1$, the recursive structure of eq.(A-6) implies that

$$\widehat{\mathbf{B}}_h^{\text{Lu}} = \widehat{\mathbf{B}}_h + \boldsymbol{\psi}_h^{\text{B}} + O_p(T^{-1/2}),$$

where the deviation term $\boldsymbol{\psi}_h^B$ is defined recursively as:

$$\boldsymbol{\psi}_h^B = - \sum_{j=1}^{h-1} \left(\mathbf{B}_{h-j} + \boldsymbol{\psi}_{h-j}^B \right) \boldsymbol{\phi}_j \boldsymbol{\Gamma}^{-1}, \quad \text{with } \boldsymbol{\psi}_1^B = 0.$$

□

A.4.2 Proof of eq.(16) in Corollary 1

Substituting eq.(A-3) into the expression for $\widehat{\mathbf{B}}_h^{\text{Lu}}$, we have for $h \geq 2$:

$$\begin{aligned} \widehat{\mathbf{B}}_h^{\text{Lu}} &= \left(\frac{1}{T} \sum_{t=1}^{T-h} \left(\mathbf{y}_{t+h} - \sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \widehat{\boldsymbol{\varepsilon}}_{t+j} \right) \mathbf{y}'_t \right) \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} \\ &= \left(\frac{1}{T} \sum_{t=1}^{T-h} \left(\widehat{\mathbf{A}}^h \mathbf{y}_t + \sum_{j=1}^{h-1} \left(\widehat{\mathbf{A}}^{h-j} - \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \right) \widehat{\boldsymbol{\varepsilon}}_{t+j} + \widehat{\boldsymbol{\varepsilon}}_{t+h} \right) \mathbf{y}'_t \right) \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} \\ &= \widehat{\mathbf{A}}^h + \left(\frac{1}{T} \sum_{t=1}^{T-h} \left(\sum_{j=1}^{h-1} \left(\widehat{\mathbf{A}}^{h-j} - \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \right) \widehat{\boldsymbol{\varepsilon}}_{t+j} + \widehat{\boldsymbol{\varepsilon}}_{t+h} \right) \mathbf{y}'_t \right) \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} \\ &= \widehat{\mathbf{A}}^h + \left(\widehat{\boldsymbol{\phi}}_h - \sum_{j=1}^{h-1} \left(\widehat{\mathbf{B}}_{h-j}^{\text{Lu}} - \widehat{\mathbf{A}}^{h-j} \right) \widehat{\boldsymbol{\phi}}_j \right) \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} \end{aligned} \quad (\text{A-8})$$

where $\widehat{\boldsymbol{\phi}}_j = \frac{1}{T} \sum_{t=1}^{T-h} \widehat{\boldsymbol{\varepsilon}}_{t+j} \mathbf{y}'_t$ and $\widehat{\boldsymbol{\Gamma}}_{T-h} = \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{y}_t \mathbf{y}'_t$.

We start with $h = 2$. Since $(\widehat{\mathbf{B}}_1^{\text{Lu}} - \widehat{\mathbf{A}}) = \mathbf{0}$ by definition, Lemmas 2 and 5 imply:

$$\widehat{\mathbf{B}}_2^{\text{Lu}} - \widehat{\mathbf{A}}^2 = \widehat{\boldsymbol{\phi}}_2 \widehat{\boldsymbol{\Gamma}}_{T-2}^{-1} = \boldsymbol{\psi}_2^A + O_p(T^{-1/2}),$$

where $\boldsymbol{\psi}_2^A = \boldsymbol{\phi}_2 \boldsymbol{\Gamma}^{-1}$.

For $h = 3$, we can substitute the result for $h = 2$ into the recursion, and find using the same lemmas as above:

$$\widehat{\mathbf{B}}_3^{\text{Lu}} - \widehat{\mathbf{A}}^3 = (\boldsymbol{\phi}_3 - \boldsymbol{\psi}_2^A \boldsymbol{\phi}_1) \boldsymbol{\Gamma}^{-1} + O_p(T^{-1/2}) = \boldsymbol{\psi}_3^A + O_p(T^{-1/2}).$$

Hence, by recursively substituting back into the expression (A-8) and making use of Lemmas 2 and 5, we obtain the following result for general $h \geq 2$:

$$\widehat{\mathbf{B}}_h^{\text{Lu}} - \widehat{\mathbf{A}}^h = \boldsymbol{\psi}_h^A + O_p(T^{-1/2}),$$

where $\boldsymbol{\psi}_h^A$ is defined recursively as:

$$\boldsymbol{\psi}_h^A = \left(\boldsymbol{\phi}_h - \sum_{j=1}^{h-2} \boldsymbol{\psi}_{h-j}^A \boldsymbol{\phi}_j \right) \boldsymbol{\Gamma}^{-1}.$$

□

A.5 Proof of Corollary 2

Substituting eq.(A-2) into the expression for $\widehat{\mathbf{B}}_{h,(-a)}^{\text{BB}}$, we have:

$$\widehat{\mathbf{B}}_{h,(-a)}^{\text{BB}} = \left(\sum_{t=1}^{T-a} \left(\widehat{\mathbf{A}}_{(-a)}^{h-1} \mathbf{y}_{t+1} - \sum_{j=2}^{h-2} \left(\widehat{\mathbf{B}}_{h-j,(-a)}^{\text{BB}} - \widehat{\mathbf{A}}_{(-a)}^{h-j} \right) \widehat{\boldsymbol{\varepsilon}}_{t+j,(-a)} \right) \mathbf{y}'_t \right) \left(\sum_{t=1}^{T-a} \mathbf{y}_t \mathbf{y}'_t \right)^{-1}, \quad (\text{A-9})$$

where we used that by definition $\widehat{\mathbf{B}}_{0,(-a)}^{\text{BB}} = \widehat{\mathbf{A}}_{(-a)}^0 = \mathbf{I}_k$ and $\widehat{\mathbf{B}}_{1,(-a)}^{\text{BB}} = \widehat{\mathbf{A}}_{(-a)}$.

Case 1 ($a = H$): The LP is estimated over the fixed sample $t = 1, \dots, T - H$.

We shall proceed by induction. For $h = 1$, we have by construction that $\widehat{\mathbf{B}}_{1,(-H)}^{\text{BB}} = \widehat{\mathbf{A}}_{(-H)}$. Assume then as the strong inductive hypothesis for $h - 1$ that $\widehat{\mathbf{B}}_{h-j,(-H)}^{\text{BB}} = \widehat{\mathbf{A}}_{(-H)}^{h-j}$ holds for all $j = 2, \dots, h - 2$ and $h > 1$. Substituting this hypothesis into eq.(A-9) gives for h that:

$$\widehat{\mathbf{B}}_{h,(-H)}^{\text{BB}} = \left(\sum_{t=1}^{T-H} \widehat{\mathbf{A}}_{(-H)}^{h-1} \mathbf{y}_{t+1} \mathbf{y}'_t \right) \left(\sum_{t=1}^{T-H} \mathbf{y}_t \mathbf{y}'_t \right)^{-1} = \widehat{\mathbf{A}}_{(-H)}^{h-1} \widehat{\mathbf{A}}_{(-H)} = \widehat{\mathbf{A}}_{(-H)}^h.$$

Hence, by mathematical induction, the result $\widehat{\mathbf{B}}_{h,(-H)}^{\text{BB}} = \widehat{\mathbf{A}}_{(-H)}^h$ holds for all $h = 1, \dots, H$.

Case 2 ($a = h$): The LP is estimated over the longest possible sample $t = 1, \dots, T - h$.

In this case, \mathbf{A} is estimated using the full sample $t = 1, \dots, T - 1$, so that $\widehat{\mathbf{A}}_{(-a)}^j = \widehat{\mathbf{A}}_{(-1)}^j$. By our adopted convention, $\widehat{\mathbf{A}}_{(-1)}^j = \widehat{\mathbf{A}}^j$, the $(-a)$ subscripts can be omitted from eq.(A-9). For $h = 1$, we have by construction that $\widehat{\mathbf{B}}_1^{\text{BB}} = \widehat{\mathbf{A}}$. For $h = 2$, substituting into eq.(A-9), we obtain:

$$\widehat{\mathbf{B}}_2^{\text{BB}} = \widehat{\mathbf{A}} \left(\sum_{t=1}^{T-2} \mathbf{y}_{t+1} \mathbf{y}'_t \right) \left(\sum_{t=1}^{T-2} \mathbf{y}_t \mathbf{y}'_t \right)^{-1} = \widehat{\mathbf{A}} \widehat{\mathbf{A}}_{(-2)} = \widehat{\mathbf{A}}^2 + O_p(T^{-1}),$$

with the simplification in the last step following from $\widehat{\mathbf{A}}_{(-2)} = \widehat{\mathbf{A}} + O_p(T^{-1})$ of Lemma 4.

Now assume for $h - 1$ and $h \geq 3$ that $\widehat{\mathbf{B}}_{h-j}^{\text{BB}} = \widehat{\mathbf{A}}^{h-j} + O_p(T^{-1})$ holds for all $j = 2, \dots, h - 2$, and substitute the hypothesis into eq.(A-9). Using Lemma 4, we thus obtain for h that:

$$\widehat{\mathbf{B}}_h^{\text{BB}} = \widehat{\mathbf{A}}^{h-1} \widehat{\mathbf{A}}_{(-h)} + O_p(T^{-1}) = \widehat{\mathbf{A}}^{h-1} \left(\widehat{\mathbf{A}} + O_p(T^{-1}) \right) = \widehat{\mathbf{A}}^h + O_p(T^{-1}).$$

Hence, by mathematical induction, the result $\widehat{\mathbf{B}}_h^{\text{BB}} = \widehat{\mathbf{A}}^h + O_p(T^{-1})$ holds for all $h = 2, \dots, H$. □

Appendix B Supplementary Material

The online Supplementary Material contains all derivations and additional simulation results supporting the illustrations in Section 4.

Supplementary Material for GLS Estimation of Local Projections: Trading Robustness for Efficiency

Ignace De Vos^{a,b} and Gerdie Everaert^{c,*}

^aVU Amsterdam, Department of Econometrics and Data Science

^bTinbergen Institute

^cGhent University, Department of Economics

This supplement contains two sections. Supplement 1 provides the derivations underlying Proposition 3 in Section 4.1.1 of the main text. Supplement 2 reports simulation results for the local misspecification setting with data-driven lag selection, as discussed in Section 4.1.2 of the main text.

Supplement 1 Proof of Proposition 3 from the Main Paper

The proofs in this section rely on the DGP and assumptions introduced in Subsection 4.1.1 of the main paper. The derivation of the asymptotic distributions for the LP OLS and VAR IR estimators follows similar arguments to the proofs provided in the working paper version of Li et al. (2022).

For convenience, we restate the DGP from eq.(19):

$$w_{t+1} = \rho w_t + \beta \mu_{1,t} + \mu_{2,t+1} + \frac{\alpha}{\sqrt{T}} \mu_{2,t}, \quad (\text{S1.1})$$

along with the accompanying assumptions: namely, that $|\rho| < 1$ and that $\boldsymbol{\mu}_t = (\mu_{1,t}, \mu_{2,t})'$ follows an i.i.d. white noise process with variance $\text{Var}(\boldsymbol{\mu}_t) = \text{diag}(\sigma_1^2, \sigma_2^2)$. Furthermore, w_0 is drawn from its stationary distribution. Recall that \mathbf{e}_j denotes a 2×1 vector with a one in the j -th position and zeros elsewhere, for $j = 1, 2$.

*Corresponding author. E-mail: gerdie.everaert@ugent.be

We begin by establishing several preliminary results that will be instrumental in deriving the asymptotic distributions of the VAR, LP OLS and the LP GLS-Lu estimators.

S1.1 Some Preliminary Results

By the Law of Large Numbers for stationary processes, the sample covariance matrix of $\mathbf{y}_t = (\mu_{1,t}, w_t)'$ satisfies

$$\hat{\Gamma} = \frac{1}{T} \sum_{t=1}^{T-1} \mathbf{y}_t \mathbf{y}_t' = \Gamma + o_p(1), \quad (\text{S1.2})$$

where we define the population covariance matrix Γ as the limiting second moment of \mathbf{y}_t as $T \rightarrow \infty$:

$$\Gamma = \lim_{T \rightarrow \infty} \mathbb{E}[\mathbf{y}_t \mathbf{y}_t'] = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_w^2 \end{bmatrix},$$

with $\sigma_w^2 = (\beta^2 \sigma_1^2 + \sigma_2^2) / (1 - \rho^2)$. This result is unaffected by the local misspecification term α / \sqrt{T} , which vanishes asymptotically.

Another useful result follows from the properties of the innovation terms. Since $\mu_{1,t}$ and $\mu_{2,t}$ are i.i.d. white noise with finite fourth moments, and w_t follows the stationary process in eq.(19), independence implies $\mathbb{E}[\mu_{2,t+s} \mu_{1,t}] = 0$ for all $s \geq 0$. By stationarity, $\mathbb{E}[\mu_{2,t} w_t] = \sigma_2^2$. Moreover, since $\mu_{2,t+s}$ is independent of past terms for $s > 0$, it follows that $\mathbb{E}[\mu_{2,t+s} w_t] = 0$. Hence, we obtain:

$$\frac{1}{T} \sum_{t=1}^{T-1} \mu_{2,t+s} \mu_{1,t} = O_p(T^{-1/2}), \quad (\text{S1.3})$$

$$\frac{1}{T} \sum_{t=1}^{T-1} \mu_{2,t+s} w_t = \sigma_2^2 \mathbf{1}_{\{s=0\}} + O_p(T^{-1/2}), \quad (\text{S1.4})$$

for all $s \geq 0$, where $\mathbf{1}_{\{s=0\}}$ is an indicator function that equals one if $s = 0$ and zero otherwise. Accordingly, also $\boldsymbol{\mu}_{t+1} = (\mu_{1,t+1}, \mu_{2,t+2})'$ is independent of $\mathbf{y}_t = (\mu_{1,t}, w_t)'$, so that $\mathbb{E}[\boldsymbol{\mu}_{t+1} \mathbf{y}_t'] = \mathbf{0}$, and therefore

$$\frac{1}{T} \sum_{t=1}^{T-1} \boldsymbol{\mu}_{t+1} \mathbf{y}_t' = O_p(T^{-1/2}). \quad (\text{S1.5})$$

S1.2 Asymptotic Distributions

S1.2.1 VAR IR estimator

Defining the population coefficient matrix $\mathbf{A}_0 = \begin{pmatrix} 0 & 0 \\ \beta & \rho \end{pmatrix}$, the OLS estimator for \mathbf{A} in eq.(2) from the main paper is given by

$$\begin{aligned} \hat{\mathbf{A}} &= \mathbf{A}_0 + \left(\frac{1}{T} \sum_{t=1}^{T-1} \left(\begin{bmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} + \frac{\alpha}{\sqrt{T}} \mu_{2,t} \end{bmatrix} \begin{bmatrix} \mu_{1,t} & w_t \end{bmatrix} \right) \right) \hat{\mathbf{\Gamma}}^{-1}, \\ &= \mathbf{A}_0 + \left(\frac{1}{T} \sum_{t=1}^{T-1} \mu_{t+1} \mathbf{y}'_t + \frac{\alpha \sigma_2^2}{\sqrt{T}} \mathbf{e}_2 \mathbf{e}'_2 + O_p(T^{-1}) \right) \hat{\mathbf{\Gamma}}^{-1}, \\ &= \mathbf{A}_0 + \left(\frac{1}{T} \sum_{t=1}^{T-1} \mu_{t+1} \mathbf{y}'_t + \frac{\alpha \sigma_2^2}{\sqrt{T}} \mathbf{e}_2 \mathbf{e}'_2 \right) \hat{\mathbf{\Gamma}}^{-1} + O_p(T^{-1}), \end{aligned} \quad (\text{S1.6})$$

$$= \mathbf{A}_0 + O_p(T^{-1/2}). \quad (\text{S1.7})$$

by eqs.(S1.3)–(S1.5) together with $\hat{\mathbf{\Gamma}}^{-1} = O_p(1)$. As such, $\hat{\mathbf{A}}$ is a consistent estimator of \mathbf{A}_0 .

However, rewriting eq.(S1.6) and making use of (S1.2) gives as $T \rightarrow \infty$

$$\sqrt{T} (\hat{\mathbf{A}} - \mathbf{A}_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \mu_{t+1} \mathbf{y}'_t \mathbf{\Gamma}^{-1} + \alpha \sigma_2^2 \mathbf{e}_2 \mathbf{e}'_2 \mathbf{\Gamma}^{-1} + o_p(1), \quad (\text{S1.8})$$

which given $\mathbb{E}(\mu_{t+1} \mathbf{y}'_t) = \mathbf{0}$ reveals that there is an asymptotic bias term when $\alpha \neq 0$. To see how this affects the distribution of the VAR IR estimator for $\theta_h = \rho^{h-1} \beta$, we first derive the asymptotic distribution of the VAR estimator for ρ and β .

Using $\mathbf{e}'_2 \mathbf{A}_0 = (\beta, \rho)$ to select the second row of \mathbf{A}_0 , applying a standard martingale central limit theorem to Eq.(S1.8) yields:

$$\sqrt{T} (\hat{\mathbf{A}} - \mathbf{A}_0)' \mathbf{e}_2 \xrightarrow{d} \mathcal{N} \left(\text{aBias}(\hat{\mathbf{A}}' \mathbf{e}_2), \text{aVar}(\hat{\mathbf{A}}' \mathbf{e}_2) \right). \quad (\text{S1.9})$$

where

$$\text{aBias}(\hat{\mathbf{A}}' \mathbf{e}_2) = \mathbf{\Gamma}^{-1} \mathbb{E} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \mu_{t+1} \mathbf{y}'_t + \alpha \sigma_2^2 \mathbf{e}_2 \mathbf{e}'_2 \right)' \mathbf{e}_2 = \frac{\alpha \sigma_2^2}{\sigma_w^2} \mathbf{e}_2,$$

from $\mathbb{E} [\mu_{t+1} \mathbf{y}'_t] = \mathbf{0}$.

The asymptotic variance follows from eq.(S1.8) as:

$$\begin{aligned}
\text{aVar}(\widehat{\mathbf{A}}' \mathbf{e}_2) &= \mathbf{\Gamma}^{-1} \lim_{T \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \boldsymbol{\mu}_{t+1} \mathbf{y}'_t \right)' \mathbf{e}_2 \mathbf{e}'_2 \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \boldsymbol{\mu}_{t+1} \mathbf{y}'_t \right) \right] \mathbf{\Gamma}^{-1}, \\
&= \mathbf{\Gamma}^{-1} \lim_{T \rightarrow \infty} \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^{T-1} \begin{bmatrix} \mu_{1,t}^2 \mu_{2,t+1}^2 & \mu_{1,t} w_t \mu_{2,t+1}^2 \\ \mu_{1,t} w_t \mu_{2,t+1}^2 & \mu_{2,t+1}^2 w_t^2 \end{bmatrix} \right] \mathbf{\Gamma}^{-1}, \\
&= \mathbf{\Gamma}^{-1} \left(\sigma_2^2 \mathbf{\Gamma} \right) \mathbf{\Gamma}^{-1} = \sigma_2^2 \mathbf{\Gamma}^{-1},
\end{aligned}$$

where the expectation follows from independence:

$$\mathbb{E} \begin{bmatrix} \mu_{1,t}^2 \mu_{2,t+1}^2 & \mu_{1,t} w_t \mu_{2,t+1}^2 \\ \mu_{1,t} w_t \mu_{2,t+1}^2 & \mu_{2,t+1}^2 w_t^2 \end{bmatrix} = \sigma_2^2 \mathbf{\Gamma}. \quad (\text{S1.10})$$

Applying the Delta Method to the function $g(\mathbf{A}_0) = \mathbf{e}'_2 \mathbf{A}_0^h \mathbf{e}_1 = \rho^{h-1} \beta = \theta_h$, the asymptotic distribution of the impulse response estimator $\widehat{\theta}_h^{\text{VAR}} = \widehat{\rho}^{h-1} \widehat{\beta}$ follows as

$$\sqrt{T} \left(\widehat{\theta}_h^{\text{VAR}} - \theta_h \right) \xrightarrow{d} \mathcal{N} \left(b_h^{\text{VAR}}, V_h^{\text{VAR}} \right), \quad (\text{S1.11})$$

where

$$\begin{aligned}
b_h^{\text{VAR}} &= \mathbf{J}_0 \frac{\alpha \sigma_2^2}{\sigma_w^2} \mathbf{e}_2 = (h-1) \rho^{h-2} \beta \alpha \frac{\sigma_2^2}{\sigma_w^2}, \\
V_h^{\text{VAR}} &= \mathbf{J}_0 \sigma_2^2 \mathbf{\Gamma}^{-1} \mathbf{J}'_0 = \rho^{2(h-1)} \frac{\sigma_2^2}{\sigma_1^2} + (h-1)^2 \rho^{2(h-2)} \beta^2 \frac{\sigma_2^2}{\sigma_w^2},
\end{aligned}$$

using that the Jacobian, evaluated at $\text{plim } \mathbf{e}'_2 \widehat{\mathbf{A}} = \mathbf{e}'_2 \mathbf{A}_0 = (\beta, \rho)$, is given by

$$\mathbf{J}_0 = \left. \frac{\partial \mathbf{e}'_2 \widehat{\mathbf{A}}^h \mathbf{e}_1}{\partial (\mathbf{e}'_2 \widehat{\mathbf{A}})} \right|_{\mathbf{e}'_2 \widehat{\mathbf{A}} = \mathbf{e}'_2 \mathbf{A}_0} = \left(\rho^{h-1}, (h-1) \rho^{h-2} \beta \right)'.$$

S1.2.2 LP OLS estimator

Define the population coefficient matrix as $\mathbf{B}_{0,h} = \begin{pmatrix} 0 & 0 \\ \rho^{h-1} \beta & \rho^h \end{pmatrix}$ for $h > 0$, and $\mathbf{B}_{0,0} = \mathbf{I}_2$ for $h = 0$. The scaled OLS estimator for \mathbf{B}_h in eq.(4) from the main paper can be written as:

$$\sqrt{T} (\widehat{\mathbf{B}}_h - \mathbf{B}_{0,h}) = \sqrt{T} \left(\frac{1}{T} \sum_{t=1}^{T-h} \sum_{j=0}^{h-1} \left(\mathbf{B}_{0,j} \begin{bmatrix} \mu_{1,t+h-j} \\ \mu_{2,t+h-j} + \frac{\alpha}{\sqrt{T}} \mu_{2,t+h-j-1} \end{bmatrix} \begin{bmatrix} \mu_{1,t} & w_t \end{bmatrix} \right) \right) \widehat{\mathbf{\Gamma}}^{-1},$$

$$\begin{aligned}
&= \sqrt{T} \left(\frac{1}{T} \sum_{t=1}^{T-h} \sum_{j=0}^{h-1} \mathbf{B}_{0,j} \boldsymbol{\mu}_{t+h-j} \mathbf{y}'_t + \mathbf{B}_{0,h-1} \frac{\alpha \sigma_2^2}{\sqrt{T}} \mathbf{e}_2 \mathbf{e}'_2 + O_p(T^{-1}) \right) \widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \sqrt{T} \left(\frac{1}{T} \sum_{t=1}^{T-h} \sum_{j=0}^{h-1} \mathbf{B}_{0,j} \boldsymbol{\mu}_{t+h-j} \mathbf{y}'_t + \rho^{h-1} \frac{\alpha \sigma_2^2}{\sqrt{T}} \mathbf{e}_2 \mathbf{e}'_2 \right) \widehat{\boldsymbol{\Gamma}}^{-1} + O_p(T^{-1/2}), \\
&= \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \sum_{j=0}^{h-1} \mathbf{B}_{0,j} \boldsymbol{\mu}_{t+h-j} \mathbf{y}'_t + \rho^{h-1} \alpha \sigma_2^2 \mathbf{e}_2 \mathbf{e}'_2 \right) \boldsymbol{\Gamma}^{-1} + o_p(1), \tag{S1.12}
\end{aligned}$$

where we use $\mathbf{B}_{0,h-1} \mathbf{e}_2 \mathbf{e}'_2 = \begin{pmatrix} \rho^0 & 0 \\ \rho^{h-2} \beta & \rho^{h-1} \end{pmatrix} = \rho^{h-1} \mathbf{e}_2 \mathbf{e}'_2$ along with eqs.(S1.2), (S1.3) and (S1.4).

Using $\mathbf{e}'_2 \mathbf{B}_{0,h} \mathbf{e}_1 = \rho^{h-1} \beta = \theta_h$ to select the relevant element in $\mathbf{B}_{0,h}$, applying a standard martingale central limit theorem to Eq.(S1.12) yields:

$$\sqrt{T} \left(\widehat{\theta}_h^{\text{LP}} - \theta_h \right) \xrightarrow{d} \mathcal{N} \left(0, V_h^{\text{LP}} \right). \tag{S1.13}$$

The absence of asymptotic bias follows from eq.(S1.12) as:

$$b_h^{\text{LP}} = \lim_{T \rightarrow \infty} \mathbb{E} \left[\sqrt{T} \mathbf{e}'_2 \left(\widehat{\mathbf{B}}_h - \mathbf{B}_{0,h} \right) \mathbf{e}_1 \right] = \lim_{T \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \sum_{j=0}^{h-1} \mathbf{e}'_2 \mathbf{B}_{0,j} \boldsymbol{\mu}_{t+h-j} \boldsymbol{\mu}_{1,t} \right) \sigma_1^{-2} \right] = 0,$$

using $\mathbb{E} \left[\boldsymbol{\mu}_{t+s} \boldsymbol{\mu}_{1,t} \right] = \mathbf{0}$ for all $s > 0$.

The asymptotic variance follows from eq.(S1.12) as

$$\begin{aligned}
V_h^{\text{LP}} &= \lim_{T \rightarrow \infty} \text{Var} \left(\sqrt{T} \mathbf{e}'_2 \left(\widehat{\mathbf{B}}_h - \mathbf{B}_{0,h} \right) \mathbf{e}_1 \right), \\
&= \sigma_1^{-4} \lim_{T \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \sum_{j=0}^{h-1} \mathbf{e}'_2 \mathbf{B}_{0,j} \boldsymbol{\mu}_{t+h-j} \boldsymbol{\mu}_{1,t} \right)^2 \right], \\
&= \sigma_1^{-4} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E} \left[\sum_{j=1}^{h-1} \rho^{j-1} \beta \boldsymbol{\mu}_{1,t+h-j} \boldsymbol{\mu}_{1,t} + \sum_{j=0}^{h-1} \rho^j \boldsymbol{\mu}_{2,t+h-j} \boldsymbol{\mu}_{1,t} \right]^2, \\
&= \sigma_1^{-2} \left(\beta^2 \sigma_1^2 \sum_{j=1}^{h-1} \rho^{2(j-1)} + \sigma_2^2 \sum_{j=0}^{h-1} \rho^{2j} \right), \\
&= \sigma_1^{-2} \left(\left(\beta^2 \sigma_1^2 + \sigma_2^2 \right) \frac{1 - \rho^{2h}}{1 - \rho^2} - \rho^{2(h-1)} \beta^2 \sigma_1^2 \right), \\
&= \left(1 - \rho^{2h} \right) \frac{\sigma_w^2}{\sigma_1^2} - \rho^{2(h-1)} \beta^2,
\end{aligned}$$

where the expectation follows from independence and $\mathbb{E}[\boldsymbol{\mu}_{t+s} \boldsymbol{\mu}_{1,t}] = \mathbf{0}$ for all $s > 0$.

S1.2.3 LP GLS estimator of Lusompa (2023)

Let $\widehat{\boldsymbol{\varepsilon}}_{t+1} = \mathbf{y}_{t+1} - \widehat{\mathbf{A}}\mathbf{y}_t$ be the estimated VAR error term. The LP GLS-Lu estimator $\widehat{\mathbf{B}}_h^{\text{Lu}}$ for \mathbf{B}_h in eq.(4) from the main paper can then be written as:

$$\begin{aligned}
\widehat{\mathbf{B}}_h^{\text{Lu}} &= \left(\frac{1}{T} \sum_{t=1}^{T-h} \left(\mathbf{y}_{t+h} - \sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \widehat{\boldsymbol{\varepsilon}}_{t+j} \right) \mathbf{y}'_t \right) \widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \left(\frac{1}{T} \sum_{t=1}^{T-h} \left(\mathbf{B}_h \mathbf{y}_t + \sum_{j=1}^h \mathbf{B}_{h-j} \boldsymbol{\varepsilon}_{t+j} - \sum_{j=1}^{h-1} \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \widehat{\boldsymbol{\varepsilon}}_{t+j} \right) \mathbf{y}'_t \right) \widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \mathbf{B}_{0,h} + \frac{1}{T} \sum_{t=1}^{T-h} \left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} \left(\mathbf{B}_{h-j} \boldsymbol{\varepsilon}_{t+j} - \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \widehat{\boldsymbol{\varepsilon}}_{t+j} \right) \right) \mathbf{y}'_t \widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \mathbf{B}_{0,h} + \frac{1}{T} \sum_{t=1}^{T-h} \left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} \left(\left(\mathbf{B}_{h-j} - \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \right) \boldsymbol{\varepsilon}_{t+j} + \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} (\widehat{\mathbf{A}} - \mathbf{A}) \mathbf{y}_{t+j-1} \right) \right) \mathbf{y}'_t \widehat{\boldsymbol{\Gamma}}^{-1},
\end{aligned} \tag{S1.14}$$

using $\widehat{\boldsymbol{\varepsilon}}_{t+j} = (\mathbf{A} - \widehat{\mathbf{A}})\mathbf{y}_{t+j-1} + \boldsymbol{\varepsilon}_{t+j}$.

Consider first that we can write for a $j \geq 1$ that

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\varepsilon}_{t+j} \mathbf{y}'_t &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \begin{bmatrix} \mu_{1,t+j} \\ \mu_{2,t+j} + \frac{\alpha}{\sqrt{T}} \mu_{2,t+j-1} \end{bmatrix} \begin{bmatrix} \mu_{1,t} & w_t \end{bmatrix} \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \boldsymbol{\mu}_{t+j} \mathbf{y}'_t + \mathbf{1}_{\{j=1\}} \alpha \sigma_2^2 \mathbf{e}_2 \mathbf{e}'_2 + O_p(T^{-1/2}) = O_p(1).
\end{aligned} \tag{S1.15}$$

Given the sequential dependence of $\widehat{\mathbf{B}}_h^{\text{Lu}}$ on previous horizon estimates, we can first establish the asymptotic bound using strong induction. The base case follows from $\widehat{\mathbf{B}}_1^{\text{Lu}} = \widehat{\mathbf{A}}$ such that from eq.(S1.7) and noting that $\mathbf{B}_1 = \mathbf{A}$ we have that $\widehat{\mathbf{B}}_1^{\text{Lu}} = \mathbf{B}_1 + O_p(T^{-1/2})$. Assuming then that for all $1 \leq j \leq h-1$,

$$\widehat{\mathbf{B}}_{h-j}^{\text{Lu}} = \mathbf{B}_{h-j} + O_p(T^{-1/2}), \tag{S1.16}$$

and substituting this in eq. (S1.14), together with eq.(S1.15), we obtain:

$$\widehat{\mathbf{B}}_h^{\text{Lu}} = \mathbf{B}_{0,h} + O_p(T^{-1/2}). \tag{S1.17}$$

Thus, by induction, the bound holds for all $h \geq 1$.

We next derive the asymptotic distribution of $\widehat{\theta}_h^{\text{Lu}}$. Using $\mathbf{e}_2' \mathbf{B}_{0,h} \mathbf{e}_1 = \rho^{h-1} \beta = \theta_h$ to select the relevant element in $\mathbf{B}_{0,h}$, we have from eq.(S1.14)

$$\sqrt{T} \left(\widehat{\theta}_h^{\text{Lu}} - \theta_h \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mathbf{e}_2' \left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} \left(\mathbf{B}_{h-j} - \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \right) \boldsymbol{\varepsilon}_{t+j} + \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} (\widehat{\mathbf{A}} - \mathbf{A}) \mathbf{y}_{t+j-1} \right) \mathbf{y}_t' \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1.$$

We derive each of the three terms in this expression separately.

For the first, since $\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\varepsilon}_{t+h} \mathbf{y}_t' = O_p(1)$, we can make use of eq.(S1.2) to write

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mathbf{e}_2' \boldsymbol{\varepsilon}_{t+h} \mathbf{y}_t' \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1 &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mathbf{e}_2' \boldsymbol{\varepsilon}_{t+h} \mathbf{y}_t' \boldsymbol{\Gamma}^{-1} \mathbf{e}_1 + o_p(1), \\ &= \frac{1}{\sigma_1^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \left(\mu_{2,t+h} + \frac{\alpha}{\sqrt{T}} \mu_{2,t+h-1} \right) \mu_{1,t} + o_p(1), \\ &= \frac{1}{\sigma_1^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mu_{2,t+h} \mu_{1,t} + o_p(1). \end{aligned}$$

For the second, we obtain because $h \leq H$ is a finite quantity that

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \sum_{j=1}^{h-1} \mathbf{e}_2' \left(\mathbf{B}_{h-j} - \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \right) \boldsymbol{\varepsilon}_{t+j} \mathbf{y}_t' \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1 &= \mathbf{e}_2' \sum_{j=1}^{h-1} \left(\mathbf{B}_{h-j} - \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\varepsilon}_{t+j} \mathbf{y}_t' \right) \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1, \\ &= O_p(T^{-1/2}), \end{aligned}$$

since $\widehat{\boldsymbol{\Gamma}}^{-1} = O_p(1)$ and because $\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\varepsilon}_{t+j} \mathbf{y}_t' = O_p(1)$ and $\widehat{\mathbf{B}}_{h-j}^{\text{Lu}} - \mathbf{B}_{h-j} = O_p(T^{-1/2})$ for any $1 \leq j \leq h-1$ by eqs.(S1.15) and (S1.16).

Third,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \sum_{j=1}^{h-1} \mathbf{e}_2' \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} (\widehat{\mathbf{A}} - \mathbf{A}) \mathbf{y}_{t+j-1} \mathbf{y}_t' \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1 &= \sum_{j=1}^{h-1} \mathbf{e}_2' \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \sqrt{T} (\widehat{\mathbf{A}} - \mathbf{A}) \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{y}_{t+j-1} \mathbf{y}_t' \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1, \\ &= \sum_{j=1}^{h-1} \mathbf{e}_2' \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \sqrt{T} (\widehat{\mathbf{A}} - \mathbf{A}) \frac{1}{T} \sum_{t=1}^{T-h} \left(\mathbf{B}_{j-1} \mathbf{y}_t + \sum_{l=1}^{j-1} \mathbf{B}_{j-1-l} \boldsymbol{\varepsilon}_{t+l} \right) \mathbf{y}_t' \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1, \\ &= \sum_{j=1}^{h-1} \mathbf{e}_2' \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \sqrt{T} (\widehat{\mathbf{A}} - \mathbf{A}) \frac{1}{T} \sum_{t=1}^{T-h} \left(\mathbf{B}_{j-1} \mathbf{y}_t \mathbf{y}_t' \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1 + \sum_{l=1}^{j-1} \mathbf{B}_{j-1-l} \boldsymbol{\varepsilon}_{t+l} \mathbf{y}_t' \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}_1 \right), \\ &= \sum_{j=1}^{h-1} \left(\mathbf{e}_2' \widehat{\mathbf{B}}_{h-j}^{\text{Lu}} \sqrt{T} (\widehat{\mathbf{A}} - \mathbf{A}) \left(\mathbf{B}_{j-1} \mathbf{e}_1 + O_p(T^{-1/2}) \right) \right), \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^{h-1} \mathbf{e}'_2 \mathbf{B}_{h-j} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \mathbf{y}'_t \boldsymbol{\Gamma}^{-1} + \alpha \sigma_2^2 \mathbf{e}_2 \mathbf{e}'_2 \boldsymbol{\Gamma}^{-1} \right) \mathbf{B}_{j-1} \mathbf{e}_1 + o_p(1), \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \sum_{j=1}^{h-1} \mathbf{e}'_2 \mathbf{B}_{h-j} \boldsymbol{\mu}_{t+1} \mathbf{y}'_t \boldsymbol{\Gamma}^{-1} \mathbf{B}_{j-1} \mathbf{e}_1 + \alpha \sigma_2^2 \sum_{j=2}^{h-1} \mathbf{e}'_2 \mathbf{B}_{h-j} \mathbf{e}_2 \mathbf{e}'_2 \boldsymbol{\Gamma}^{-1} \mathbf{B}_{j-1} \mathbf{e}_1 + o_p(1), \quad (\text{S1.18})
\end{aligned}$$

where use is made of eqs.(S1.2) and (S1.8), and eq.(S1.15) on the 4th equality.

The first term in eq.(S1.18) is given by

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \sum_{j=1}^{h-1} \mathbf{e}'_2 \mathbf{B}_{h-j} \boldsymbol{\mu}_{t+1} \mathbf{y}'_t \boldsymbol{\Gamma}^{-1} \mathbf{B}_{j-1} \mathbf{e}_1, \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mathbf{e}'_2 \mathbf{B}_{h-1} \boldsymbol{\mu}_{t+1} \mathbf{y}'_t \boldsymbol{\Gamma}^{-1} \mathbf{e}_1 + \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \sum_{j=2}^{h-1} \mathbf{e}'_2 \mathbf{B}_{h-j} \boldsymbol{\mu}_{t+1} \mathbf{y}'_t \boldsymbol{\Gamma}^{-1} \mathbf{B}_{j-1} \mathbf{e}_1, \\
&= \mathbf{e}'_2 \mathbf{B}_{h-1} \frac{1}{\sigma_1^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \boldsymbol{\mu}'_{1,t} + \sum_{j=2}^{h-1} \frac{\rho^{j-2} \beta}{\sigma_w^2} \mathbf{e}'_2 \mathbf{B}_{h-j} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \mathbf{y}'_t \mathbf{e}_2, \\
&= \mathbf{e}'_2 \mathbf{B}_{h-1} \frac{1}{\sigma_1^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \boldsymbol{\mu}'_{1,t} + \sum_{j=2}^{h-1} \frac{\rho^{h-3} \beta}{\sigma_w^2} \mathbf{e}'_2 \mathbf{A}_0 \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \boldsymbol{w}_t, \\
&= \mathbf{e}'_2 \mathbf{B}_{h-1} \frac{1}{\sigma_1^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \boldsymbol{\mu}'_{1,t} + (h-2) \frac{\rho^{h-3} \beta}{\sigma_w^2} \mathbf{e}'_2 \mathbf{A}_0 \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \boldsymbol{w}_t,
\end{aligned}$$

where use is made of $\mathbf{B}_0 = \mathbf{I}_2$ such that $\boldsymbol{\Gamma}^{-1} \mathbf{B}_0 \mathbf{e}_1 = \boldsymbol{\Gamma}^{-1} \mathbf{e}_1 = \sigma_1^{-2} \mathbf{e}_1$ along with $\boldsymbol{\Gamma}^{-1} \mathbf{B}_{j-1} \mathbf{e}_1 = \rho^{j-2} \beta \sigma_w^{-2} \mathbf{e}_2$ for $j > 1$ and $\mathbf{B}_{h-j} = \rho^{h-j-1} \mathbf{A}_0$.

The second term in eq.(S1.18) is given by

$$\begin{aligned}
\alpha \sigma_2^2 \sum_{j=1}^{h-1} \mathbf{e}'_2 \mathbf{B}_{h-j} \mathbf{e}_2 \mathbf{e}'_2 \boldsymbol{\Gamma}^{-1} \mathbf{B}_{j-1} \mathbf{e}_1 &= \alpha \sigma_2^2 \sum_{j=2}^{h-1} \mathbf{e}'_2 \mathbf{B}_{h-j} \mathbf{e}_2 \mathbf{e}'_2 \boldsymbol{\Gamma}^{-1} \mathbf{B}_{j-1} \mathbf{e}_1 = \beta \frac{\alpha \sigma_2^2}{\sigma_w^2} \sum_{j=2}^{h-1} \rho^{h-2} \\
&= (h-2) \rho^{h-2} \beta \frac{\alpha \sigma_2^2}{\sigma_w^2},
\end{aligned}$$

using $\mathbf{B}_0 = \mathbf{I}_2$ such that $\mathbf{e}'_2 \mathbf{B}_{h-1} \mathbf{e}_2 \mathbf{e}'_2 \boldsymbol{\Gamma}^{-1} \mathbf{B}_0 \mathbf{e}_1 = 0$.

Collecting and expanding terms yields

$$\begin{aligned}
\sqrt{T} (\hat{\theta}_h^{\text{Lu}} - \theta_h) &= (h-2) \rho^{h-2} \beta \frac{\alpha \sigma_2^2}{\sigma_w^2} + \frac{1}{\sigma_1^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{2,t+h} \boldsymbol{\mu}'_{1,t} + \mathbf{e}'_2 \mathbf{B}_{h-1} \frac{1}{\sigma_1^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \boldsymbol{\mu}'_{1,t} \\
&\quad + (h-2) \frac{\rho^{h-3} \beta}{\sigma_w^2} \mathbf{e}'_2 \mathbf{A}_0 \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \boldsymbol{w}_t + O_p(T^{-1/2}),
\end{aligned}$$

$$\begin{aligned}
&= (h-2)\rho^{h-2}\beta\frac{\alpha\sigma_2^2}{\sigma_w^2} + \frac{1}{\sigma_1^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+h}\mu_{1,t} + \frac{\rho^{h-2}\beta}{\sigma_1^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{1,t+1}\mu_{1,t} \\
&\quad + \frac{\rho^{h-1}}{\sigma_1^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+1}\mu_{1,t} + (h-2)\frac{\rho^{h-3}\beta^2}{\sigma_w^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{1,t+1}w_t \\
&\quad + (h-2)\frac{\rho^{h-2}\beta}{\sigma_w^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+1}w_t + o_p(1). \tag{S1.19}
\end{aligned}$$

Applying a standard martingale central limit theorem to Eq.(S1.19) yields:

$$\sqrt{T}\left(\widehat{\theta}_h^{\text{Lu}} - \theta_h\right) \xrightarrow{d} \mathcal{N}\left(b_h^{\text{Lu}}, V_h^{\text{Lu}}\right). \tag{S1.20}$$

Using independence, the asymptotic bias is given by

$$b_h^{\text{Lu}} = \lim_{T \rightarrow \infty} \mathbb{E}\left[\sqrt{T}\left(\widehat{\theta}_h^{\text{Lu}} - \theta_h\right)\right] = (h-2)\rho^{h-2}\beta\frac{\alpha\sigma_2^2}{\sigma_w^2}.$$

The asymptotic variance is given by

$$\begin{aligned}
V_h^{\text{Lu}} &= \lim_{T \rightarrow \infty} \text{Var}\left(\sqrt{T}\left(\widehat{\theta}_h^{\text{Lu}} - \theta_h\right)\right), \\
&= \lim_{T \rightarrow \infty} \frac{1}{\sigma_1^4}\frac{1}{T}\sum_{t=1}^{T-h}\text{Var}(\mu_{2,t+h}\mu_{1,t}) + \frac{\rho^{2(h-2)}\beta^2}{\sigma_1^4}\frac{1}{T}\sum_{t=1}^{T-h}\text{Var}(\mu_{1,t+1}\mu_{1,t}) \\
&\quad + \frac{\rho^{2(h-1)}}{\sigma_1^4}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\text{Var}(\mu_{2,t+1}\mu_{1,t}) \\
&\quad + (h-2)^2\frac{\rho^{2(h-3)}\beta^4}{\sigma_w^4}\frac{1}{T}\sum_{t=1}^{T-h}\text{Var}(\mu_{1,t+1}w_t) + (h-2)^2\frac{\rho^{2(h-2)}\beta^2}{\sigma_w^4}\frac{1}{T}\sum_{t=1}^{T-h}\text{Var}(\mu_{2,t+1}w_t) \\
&\quad + 2\frac{1}{\sigma_1^2}(h-2)\frac{\rho^{h-2}\beta}{\sigma_w^2}\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+h}\mu_{1,t}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+1}w_t\right)\right], \\
&= \frac{\sigma_2^2}{\sigma_1^2} + \rho^{2(h-2)}\beta^2 + \rho^{2(h-1)}\frac{\sigma_2^2}{\sigma_1^2} + (h-2)^2\rho^{2(h-3)}\beta^4\frac{\sigma_1^2}{\sigma_w^2} + (h-2)^2\rho^{2(h-2)}\beta^2\frac{\sigma_2^2}{\sigma_w^2} \\
&\quad + 2(h-2)\rho^{2(h-2)}\beta^2\frac{\sigma_2^2}{\sigma_w^2}, \\
&= \left(1 + \rho^{2(h-1)}\right)\frac{\sigma_2^2}{\sigma_1^2} + \left(1 + h(h-2)\frac{\sigma_2^2}{\sigma_w^2}\right)\rho^{2(h-2)}\beta^2 + (h-2)^2\rho^{2(h-3)}\beta^4\frac{\sigma_1^2}{\sigma_w^2},
\end{aligned}$$

using independence and noting that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mu_{2,t+h} \mu_{1,t} = \frac{1}{\sqrt{T}} \sum_{t=h}^{T-1} \mu_{2,t+1} \mu_{1,t-h+1}$ such that

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mu_{2,t+h} \mu_{1,t} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mu_{2,t+1} w_t \right) \right] &= \frac{1}{T} \mathbb{E} \left[\left(\sum_{t=h}^{T-1} \mu_{2,t+1} \mu_{1,t-h+1} \right) \left(\sum_{t=1}^{T-h} \mu_{2,t+1} w_t \right) \right], \\ &= \mathbb{E} [\mu_{2,t+1} \mu_{1,t-h+1} \mu_{2,t+1} w_t], \\ &= \rho^{h-2} \beta \mathbb{E} [\mu_{2,t+1}^2 \mu_{1,t-h+1}^2] = \rho^{h-2} \beta \sigma_2^2 \sigma_1^2. \end{aligned}$$

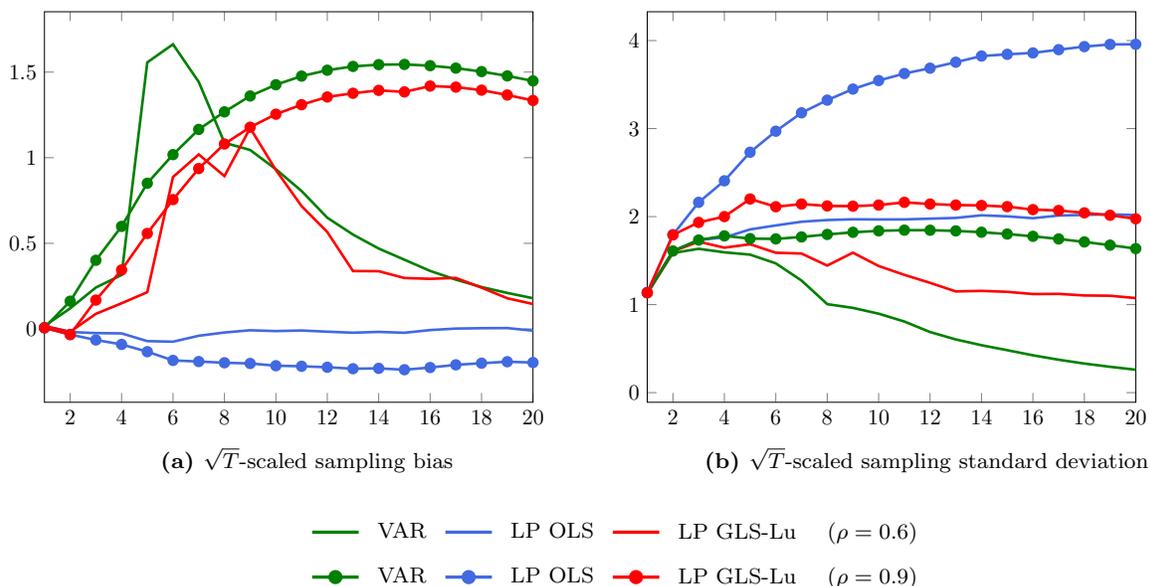
□

Supplement 2 Simulation Results for Local Misspecification with Data-Driven Lag Selection

This section reports the simulation results that correspond to Section 4.1.2 of the main paper. We evaluate estimator performance under local misspecification with data-driven lag selection. We present results for $T = 250$, two levels of persistence ($\rho = 0.6$ and $\rho = 0.9$), and three lag selection rules: AIC, the rule-of-thumb $p = \lfloor T^{1/4} \rfloor = 4$, and a larger fixed lag length $p = 8$. Each figure shows \sqrt{T} -scaled bias and standard deviation for all estimators, as well as heatmaps of weighted RMSE minima across horizons $h = 1, \dots, 20$ for a range of bias-variance weights $\lambda \in [0, 1]$.

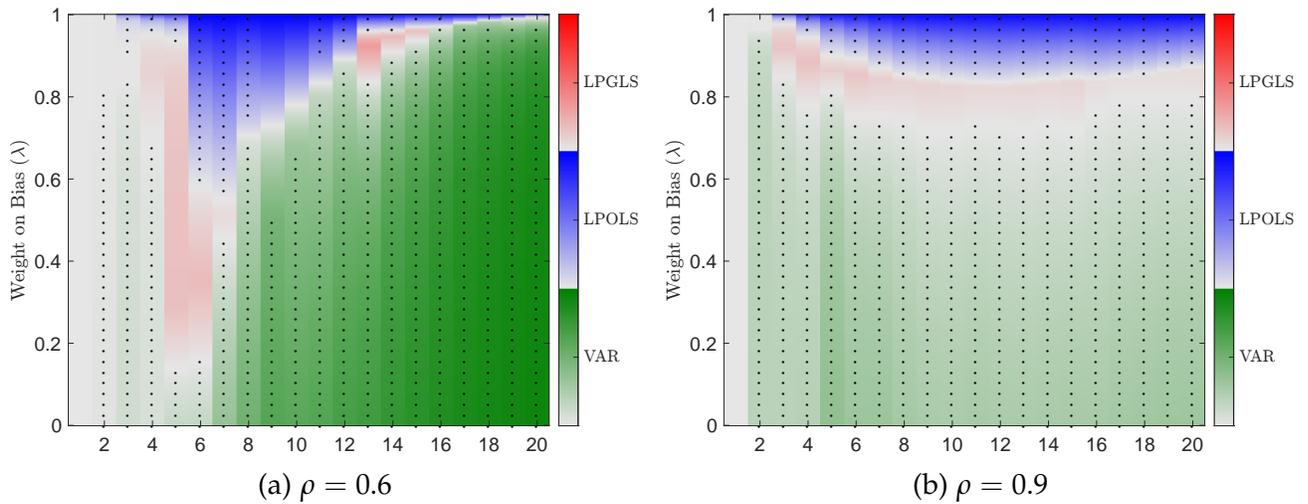
S2.1 AIC Lag Selection

Figure S2.1: Bias and Standard Deviation—Local Misspecification with AIC Lag Selection



Notes: Displayed are the \sqrt{T} -scaled bias and standard deviation of the VAR, LP OLS, and LP GLS-Lu IRs estimators under the DGP in eq.(21), based on 10,000 Monte Carlo replications. The simulation uses parameter values $\beta = \sigma_1^2 = \sigma_2^2 = 1$, $\rho \in \{0.6, 0.9\}$, $\alpha = 0.5$, and sample size $T = 250$. The VAR lag length is selected using the AIC and applied uniformly to the three estimators. The horizontal axis indicates the projection horizon $h = 1, \dots, 20$.

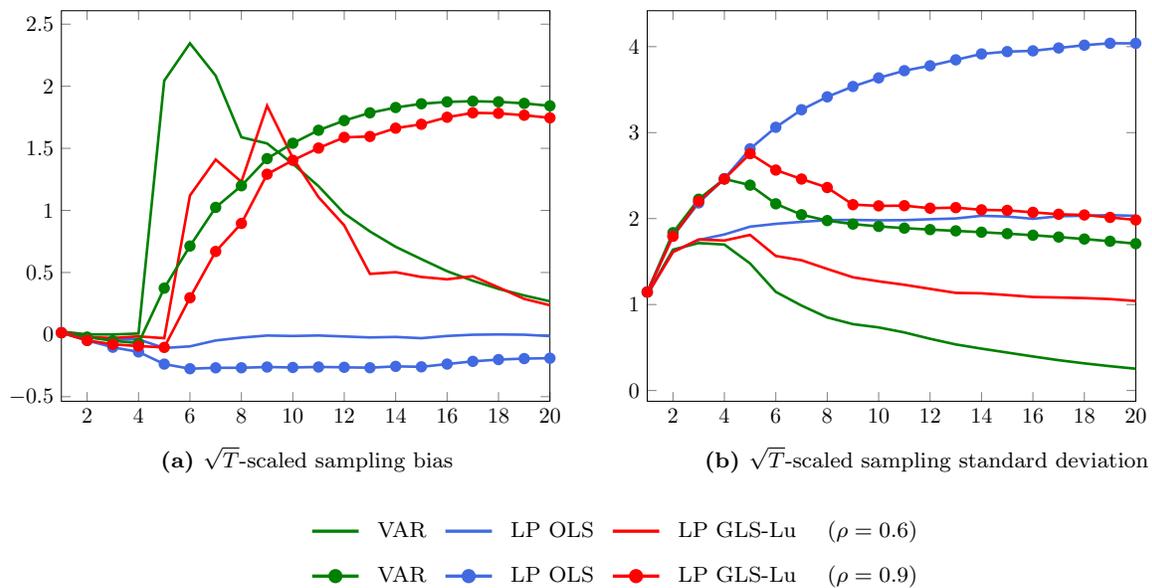
Figure S2.2: Estimator Dominance by Weighted RMSE—Local Misspecification with AIC Lag Selection



Notes: The heatmaps visualize estimator dominance across forecast horizons ($h = 1, \dots, 20$), plotted on the x-axis, and squared-bias weights ($\lambda \in [0, 1]$), plotted on the y-axis. Each cell color corresponds to the estimator—VAR IR, LP OLS, or LP GLS-Lu—minimizing the weighted RMSE defined in eq.(20), based on 10,000 Monte Carlo replications from the DGP in eq.(21) with $\beta = \sigma_1^2 = \sigma_2^2 = 1$, $\rho \in \{0.6, 0.9\}$, $\alpha = 0.5$, and sample size $T = 250$. The VAR lag length is selected using the AIC and applied uniformly to the three estimators. Color intensity reflects the RMSE reduction relative to the second-best estimator: darker shades indicate stronger dominance. Black dots highlight regions where LP GLS-Lu ranks second-best. For visual clarity, they are shown only every third weight step.

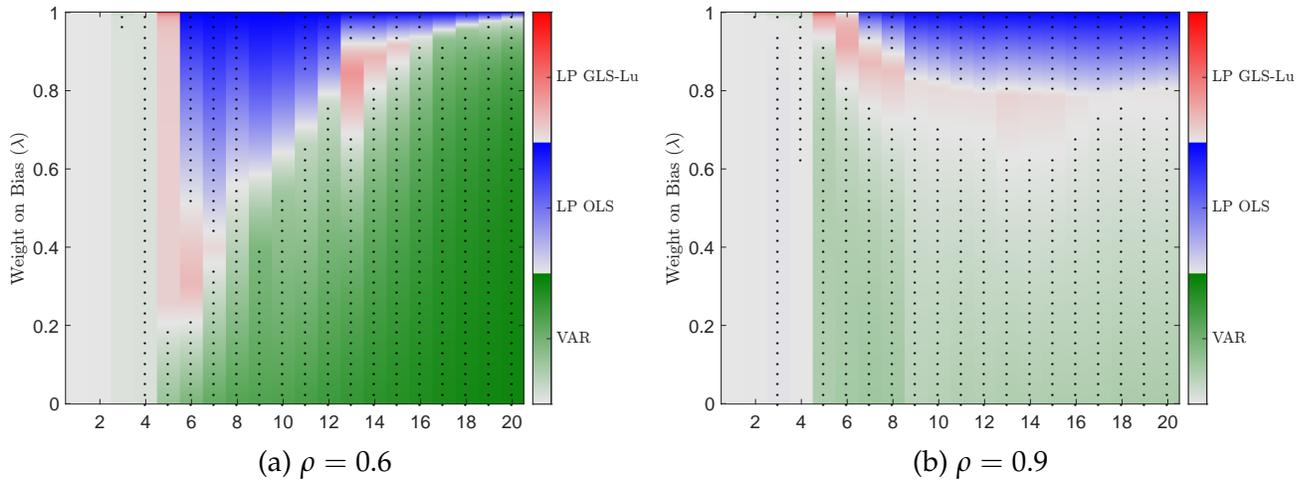
S2.2 Lag Length Set to 4

Figure S2.4: Bias and Standard Deviation—Local Misspecification with Lag Length Set to 4



Notes: See Figure S2.1, except that the lag length is fixed at $p = 4$ instead of being determined by the AIC.

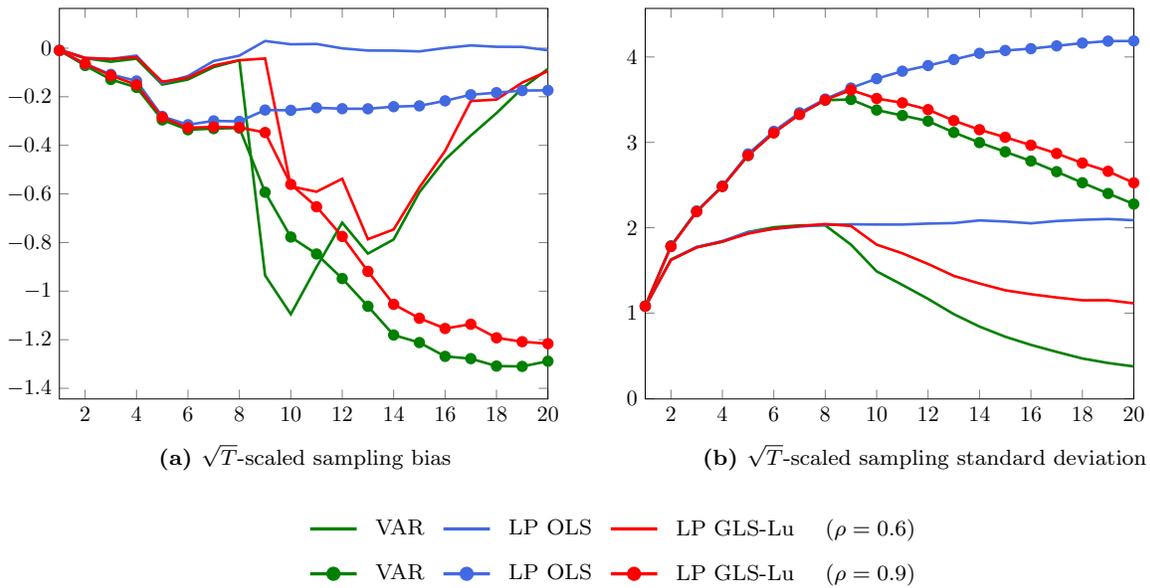
Figure S2.5: Estimator Dominance by Weighted RMSE—Local Misspecification with Lag Length Set to 4



Notes: See Figure S2.2, except that the lag length is now fixed at $p = 4$ rather than determined by the AIC.

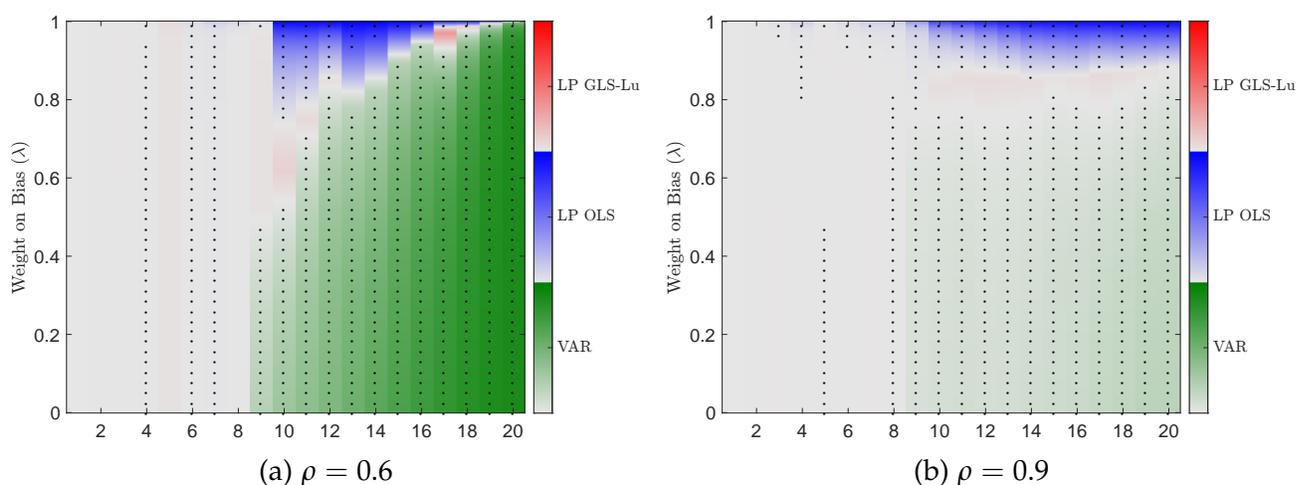
S2.3 Lag Length Set to 8

Figure S2.7: Bias and Standard Deviation—Local Misspecification with Lag Length Set to 8



Notes: See Figure S2.1, except that the lag length is fixed at $p = 4$ instead of being determined by the AIC.

Figure S2.8: Estimator Dominance by Weighted RMSE—Local Misspecification with Lag Length Set to 8



Notes: See Figure S2.2, except that the lag length is now fixed at $p = 8$ rather than determined by the AIC.

References

- Li, D., Plagborg-Møller, M., and Wolf, C. K. (2022). Local projections vs. vars: Lessons from thousands of dgps. NBER Working Paper 30207, National Bureau of Economic Research.
- Lusompa, A. (2023). Local Projections, Autocorrelation, and Efficiency. *Quantitative Economics*, 14(4):1199–1220.