WORKING PAPER

GLS ESTIMATION OF LOCAL PROJECTIONS: TRADING ROBUSTNESS FOR EFFICIENCY

Ignace De Vos
Gerdie Everaert

December 2025 Revised
2024/1095

GHENT UNIVERSITY

Department of Economics

Faculty of Economics and Business Administration
Campus Tweekerken, St.-Pietersplein 5, 9000 Ghent - BELGIUM

GLS Estimation of Local Projections: Trading Robustness for Efficiency

Ignace De Vos[a,b] and Gerdie Everaert[c,*]

[a]VU Amsterdam, Department of Econometrics and Data Science
[b]Tinbergen Institute
[c]Ghent University, Department of Economics

Original version: October 10, 2024
This version: December 27, 2025

Abstract

Local projections (LPs) are widely used for estimating impulse responses (IRs) as they are 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 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, which leads them to align with VAR IRs, while the latter preserve the unrestricted nature of LPs but end up replicating LP OLS estimates. Consequently, the intended efficiency gains are either not achieved or come at the expense of the very robustness that motivates the use of LPs.
JEL-codes: C22, C32, C52
Keywords: impulse responses, local projections, VAR models, GLS estimation, efficiency, robustness

1 Introduction

Since the seminal work of Jordà (2005), local projections (LPs) have become a widely used method for estimating impulse responses (IRs). Unlike VAR-based approaches that extrapolate IRs from a fixed number of sample autocovariances, LPs estimate them directly at each horizon, thereby

*Corresponding author. E-mail: gerdie.everaert@ugent.be
We thank two anonymous referees for their detailed comments and suggestions, which greatly improved the paper. We are also grateful to the associate editor and the managing editor for their supportive guidance throughout the revision process. We further benefited from helpful feedback during presentations at various workshops and seminars. All remaining errors are our own. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government — department EWI.

imposing fewer restrictions on the data's dynamics. This flexibility is often argued to enhance robustness relative to VARs. Olea et al. (2024) indeed show that LP confidence intervals achieve correct asymptotic coverage under broad conditions, while even mild misspecification can cause severe undercoverage in VAR-based intervals. Similarly, Kolesár and Plagborg-Møller (2024) demonstrate that LPs retain robustness in nonlinear settings. At the same time, LP estimates are typically more variable and can appear more erratic than VAR-based IRs (Ramey, 2016). Because each horizon is estimated separately, forecast errors accumulate across horizons, inducing a moving-average (MA) structure in the LP residuals and inflating estimator variance, especially at longer horizons.

This highlights a broader bias–variance trade-off. Plagborg-Møller and Wolf (2021) show that LPs and VARs yield identical IRs in population when the VAR includes a sufficiently long lag length. In finite samples, however, low-order VARs achieve lower variance but may suffer from misspecification bias, while LPs are more robust but less precise. Simulation evidence supports this view: Kilian and Kim (2011) find that when the data follow a finite-order VAR, LP confidence intervals are excessively wide, while bias-adjusted VAR bootstrap intervals are considerably narrower. Extending this to thousands of designs, Li et al. (2024) document that LPs generally deliver lower bias but higher variance than VAR estimators.

Concerns over the finite-sample variability of LPs have motivated refinements aimed at improving efficiency. Jordà (2005) already suggested recursively incorporating previous-horizon projection errors, which later inspired Generalized Least Squares (GLS) transformations for LPs. Lusompa (2023), for instance, shows that the autocorrelation structure of LP errors reflects the dynamics of an auxiliary VAR, which motivates a GLS transformation based on VAR residuals. A similar approach is explored by Breitung and Brüggemann (2023). Monte Carlo evidence in Bruns and Lütkepohl (2022) indicates that such GLS variants often outperform LP OLS and other LP refinements in terms of root mean squared error (RMSE), particularly at longer horizons.

However, these GLS transformations are motivated by settings where the reduced-form dynamics are fully captured by a VAR, either exactly in a finite-order specification or asymptotically in a sieve VAR as the lag length p grows with the sample size. In applied work, by contrast, researchers must choose a finite lag length. Even when a sieve representation exists, the working finite-order VAR can be locally misspecified. In this empirically relevant fixed-p regime, VAR-residual–based GLS transformations may reduce variance but risk importing misspecification bias, leaving the net effect unclear.

This paper builds on this notion and evaluates how LP GLS estimators perform relative to the benchmark LP OLS and VAR IRs. In doing so, we impose only weak regularity conditions that allow for local reduced-form VAR misspecification in a fixed-p setting, ensuring broad empirical relevance and enabling us to assess the effectiveness of GLS-based LP estimators in managing the bias–variance trade-off. We distinguish between two broad strands of GLS implementations. The first strand exploits the fact that under a correct reduced-form VAR specification, whether

exact or asymptotic, LP errors follow a Vector Moving Average (VMA) process in terms of VAR projection errors and IR coefficients (Lusompa, 2023). This motivates GLS transformations that condition on auxiliary VAR residuals. However, we find that once all available VAR residuals are used, the LP GLS estimator reproduces the VAR IRs mechanically, as it fully imposes the VAR dynamics on the LPs, regardless of whether the VAR is correctly specified. The closely related variant of Breitung and Brüggemann (2023) likewise coincides with the VAR IRs, either numerically or up to $\sqrt{T}$-equivalence. These VAR-residual–based GLS variants should therefore not be viewed as refinements of LPs, but rather as rebranded implementations of the VAR model, with the same sensitivities as the latter. The second strand allows for local reduced-form VAR misspecification. In such cases, we show that LP errors follow a VMA process involving iteratively re-centered VAR projection errors and pseudo-true IRs. This motivates an alternative GLS variant that instead conditions on previous-horizon LP residuals, as originally suggested by Jordà (2005). However, conditioning on LP residuals replicates the LP OLS estimator, preserving robustness and LP flexibility but forgoing efficiency gains.

Taken together, our results imply that none of these LP GLS variants generate distinct asymptotic distributions. They are either numerically identical to VAR IRs or to LP OLS, or are $\sqrt{T}$–equivalent to them. These conclusions are established in a fixed-p framework, which reflects the empirically relevant case where applied researchers must work with a finite lag length, but they continue to hold pointwise along a sieve sequence in which the lag length grows with the sample size (i.e., $p = p(T) \to \infty$ as $T \to \infty$, under standard sieve regularity conditions).

The only exception is the LP GLS estimator of Lusompa (2023), which excludes the current-horizon VAR residual from the conditioning set. As a result, it imposes most — but not all — of the VAR dynamics on the LP, preventing full collapse to the VAR IRs and yielding a genuinely distinct limiting distribution. Because closed-form expressions are analytically intractable under our general framework, we complement the general results with structured data-generating processes (DGPs) that allow sharper analytical and simulation-based comparisons. Specifically, we analyze (i) a stylized example with shrinking local misspecification that delivers explicit expressions for asymptotic bias and variance, (ii) a similar DGP but with misspecification induced by data-driven lag selection, (iii) data simulated from the empirically calibrated DSGE model of Smets and Wouters (2005), and (iv) data generated by the dynamic factor model of Stock and Watson (2016), following the Monte Carlo design of Li et al. (2024), which yields a large set of empirically realistic DGPs. Across these designs, the Lusompa (2023) estimator typically lies between LP OLS and VAR IRs along the bias–variance frontier, but seldom improves upon either benchmark in terms of weighted RMSE.

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 various GLS transformations. Section 3 analyzes the bias–variance trade-off, showing how LP GLS estimators align with, or deviate from, the benchmark methods in this general setting. Section 4 then focuses on the Lusompa (2023) estimator and illustrates its

behavior relative to VAR and LP OLS across a wide range of DGPs. Section 5 concludes. Proofs and supplementary material are provided in Appendices A and B.

2 Assumptions, Autocorrelation Processes, and Estimators

This section establishes the framework for evaluating LP and VAR impulse response estimators. We begin by introducing a general reduced-form regularity condition and the Wold representation of the data. To ensure broad applicability, our framework is deliberately general and does not restrict the process to any particular parametric or structural model. We then define the benchmark VAR and LP estimators and analyze the autocorrelation structure of LP errors, building on the results of Lusompa (2023) under correct reduced-form VAR specification. These results motivate the first strand of LP GLS estimators, which condition on auxiliary VAR residuals to mitigate the accumulation of projection errors. Finally, we extend the analysis to local reduced-form VAR misspecification and introduce a new theorem that justifies a second strand of LP GLS estimators based on previous-horizon LP residuals.

2.1 Assumptions and Reduced-Form 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.

This assumption provides a general reduced-form foundation for our analysis, ensuring that results apply to a wide class of stationary multivariate processes. 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 canonical Wold representation (see, e.g., Hamilton, 1994, Ch. 4):

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

where $\boldsymbol{\epsilon}_t$ denotes the $k \times 1$ vector of one-step-ahead linear prediction errors (innovations) with $\mathbb{E}(\boldsymbol{\epsilon}_t) = \mathbf{0}_{k\times 1}$, a positive definite covariance matrix $\boldsymbol{\Sigma}_\epsilon$, and $\mathbb{E}(\boldsymbol{\epsilon}_t \mathbf{y}'_{t-j}) = \mathbf{0}_{k\times k}$ for all $j \geq 1$.

The goal is to estimate population reduced-form IRs, which quantify the dynamic effects of each shock in $\boldsymbol{\epsilon}_t$ on $\mathbf{y}_{t+h}$, over the horizons $h = 1, \ldots, H$, where H is fixed, finite, and satisfies $H < T$.

Based on the Wold representation in eq.(1), these IRs are given by the sequence of $(k \times k)$ matrices $\{\mathbf{\Theta}_h\}_{h=1}^{H}$.

Remark 1. (Reduced-form IRs and structural identification). Assumption 1 imposes mild regularity conditions on the reduced-form Wold representation and does not restrict the causal (structural) model. Consequently, the IRs $\mathbf{\Theta}_h$ 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 emphasized by Plagborg-Møller and Wolf (2021), structural identification is a population concept, logically separate from the choice of reduced-form estimation method. Our analysis therefore focuses on estimators of reduced-form IRs. When structural IRs are of interest, they can be obtained by post-multiplying the reduced-form IRs $\mathbf{\Theta}_h$ by an appropriate identification matrix $\boldsymbol{B}^{-1}$, such that the structural IRs are given by $\mathbf{\Theta}_h \boldsymbol{B}^{-1}$. Specific identification schemes — such as IV approaches and contemporaneous or long-run restrictions — amount to imposing conditions that allow one or more columns of $\boldsymbol{B}^{-1}$ to be recovered (see, e.g., Kilian and Lütkepohl, 2017, Stock and Watson, 2018, Plagborg-Møller and Wolf, 2021). Issues of global misspecification — such as structural non-invertibility or underidentification — lie outside our maintained reduced-form framework and do not affect the validity of the reduced-form Wold representation.

2.2 Benchmark Estimators: VAR and LP OLS

VAR. Consider the VAR model:

$$\mathbf{y}_{t+1} = \boldsymbol{A}\mathbf{y}_t + \boldsymbol{\eta}_{t+1}, \qquad \text{for} \quad t = 1, \ldots, T-1, \tag{2}$$

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

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

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

such that the corresponding VAR IR estimator for $\mathbf{\Theta}_h$ is given by $\widehat{\boldsymbol{A}}_{(-a)}^{h}$. 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 $\widehat{\boldsymbol{A}}_{(-1)} = \widehat{\boldsymbol{A}}$.

LP OLS. Local projections estimate the IRs $\mathbf{\Theta}_h$ directly at each horizon through separate regressions:

$$\mathbf{y}_{t+h} = \boldsymbol{B}_h \mathbf{y}_t + \boldsymbol{e}_{t+h,h}, \qquad \text{for} \quad h = 1, \ldots, H, \tag{4}$$

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

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

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

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 $\widehat{\boldsymbol{B}}_{h,(-h)} = \widehat{\boldsymbol{B}}_h$. Alternatively, setting $a = H$ uses the same dataset for the explanatory variable $\mathbf{y}_t$ (i.e., $\mathbf{y}_1, \ldots, \mathbf{y}_{T-H}$) across all horizons $h = 1, \ldots, H$. This approach is often used in practice to ensure a uniform sample size and composition, reducing variability that may arise from differing sample periods at each horizon.

Remark 2 (Companion-form shorthand, fixed lag length, and sieve asymptotics)**.** The first-order models in eqs.(2) and (4) should be understood as companion-form representations of VAR and LP models with a fixed lag length p. The companion form is purely a notational shorthand: any VAR or LP with a fixed number of lags can be written this way with appropriately defined state variables (cf. Lütkepohl, 2005, Ch. 2), so the notation does not restrict the choice of p, beyond assuming it is fixed. We adopt a fixed-p framework because it naturally captures local misspecification, which arises when the chosen lag length is too short to approximate the reduced-form dynamics. Moreover, it yields exact population and finite-sample identities without invoking asymptotic approximations. Where asymptotic arguments are used, they serve only to attach explicit remainder orders arising from estimation-window misalignment in this fixed-lag setting. Links to sieve asymptotics, where the lag length grows with the sample size, are provided where relevant.

2.3 Autocorrelation Structures of LP Errors

Because projection errors accumulate in $\boldsymbol{e}_{t+h,h}$, the LP errors are serially correlated, and the variance of the LP OLS estimator increases with the horizon h. This 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, it is helpful to forward iterate the VAR in eq.(2). This yields the decomposition:

$$\mathbf{y}_{t+h} = \boldsymbol{A}^h \mathbf{y}_t + \sum_{j=1}^{h} \boldsymbol{A}^{h-j} \boldsymbol{\eta}_{t+j}. \tag{6}$$

Since the LP error $\boldsymbol{e}_{t+h,h}$ is defined as the projection error from regressing $\mathbf{y}_{t+h}$ onto $\mathbf{y}_t$, eq. (6) shows that its behavior is determined by the autocorrelation structure in the forward VMA term $\sum_{j=1}^{h} \boldsymbol{A}^{h-j} \eta_{t+j}$ and by how this term relates to $\mathbf{y}_t$. The properties of the VAR projection errors $\boldsymbol{\eta}_t$ are therefore central for understanding the autocorrelation structure of LP errors. Because these properties differ under correct and locally misspecified reduced-form VARs, we first formalize this distinction and then characterize the resulting VMA representations of LP errors.

2.3.1 Correct versus Locally Misspecified Reduced-Form VAR Specification

By Assumption 1, $\mathbf{y}_t$ admits a canonical invertible Wold representation, so the process can always be expressed as a projection VAR(∞) (cf. Hamilton, 1994, Plagborg-Møller and Wolf, 2021). In practice, however, researchers estimate finite-order VARs, truncating the infinite-order representation at some lag length p. This truncation is the source of *local reduced-form misspecification*: if p is too short to capture the reduced-form dynamics, the finite-order VAR only approximates the projection VAR(∞). Local misspecification is common when rich autoregressive dynamics or moving-average components are compressed into a short VAR, or when degrees-of-freedom constraints force the use of a limited number of lags (see, e.g., Braun and Mittnik, 1993; Stock and Watson, 2018).

By contrast, a *correctly specified* reduced-form VAR requires a lag length p that is sufficiently long to capture the reduced-form dynamics. This holds exactly when the true process is itself a finite-order VAR(p), and more generally asymptotically under a sieve where the lag length tends to infinity in function of T, i.e. $p = p(T) \to \infty$ as $T \to \infty$, under standard sieve regularity conditions.

Our analysis of the autocorrelation structure of LP errors therefore distinguishes between two cases: the empirically relevant scenario of local reduced-form VAR misspecification, where the VAR lag order is insufficient, and the case of correct reduced-form VAR specification. For brevity, we will often refer to these simply as local misspecification and correct specification, respectively.

2.3.2 VMA Representation of LP Errors under Correct Specification

Lusompa (2023) shows that, under Assumption 1 and correct reduced-form VAR specification, the VAR projection errors $\boldsymbol{\eta}_t$ coincide with the Wold innovations $\boldsymbol{\epsilon}_t$. Under this condition, the forward-iterated VAR representation in eq. (6) implies that the LP projection error in eq.(4) equals the forward VMA term. In population, the LP error $\boldsymbol{e}_{t+h,h}$ can therefore be expressed as a VMA

process of these innovations and the IRs:

$$e_{t+h,h}^{\text{VAR}} = \boldsymbol{\Theta}_{h-1}\boldsymbol{\eta}_{t+1} + \boldsymbol{\Theta}_{h-2}\boldsymbol{\eta}_{t+2} + \cdots + \boldsymbol{\Theta}_1\boldsymbol{\eta}_{t+h-1} + \boldsymbol{\eta}_{t+h}, \tag{7}$$

where the superscript 'VAR' indicates that this representation underlies the GLS transformations based on VAR residuals, as discussed in Subsection 2.4.1.

2.3.3 VMA Representation of LP Errors under Local Misspecification

Under local misspecification, the decomposition in eq. (6) continues to hold, but the VAR projection errors $\boldsymbol{\eta}_t$ are no longer orthogonal to $\mathbf{y}_t$. Consequently, the forward VMA term $\sum_{j=1}^{h} \boldsymbol{A}^{h-j}\boldsymbol{\eta}_{t+j}$ exhibits forward dependence, such that terms as $\mathbb{E}[\boldsymbol{\eta}_{t+j}\mathbf{y}_t']$ are generally nonzero for $j \geq 1$. This forward dependence implies two important departures from the correctly specified case. First, the VAR projection errors $\boldsymbol{\eta}_t$ no longer coincide with the Wold innovations $\boldsymbol{\epsilon}_t$ and become serially correlated. Second, the LP coefficients $\boldsymbol{B}_h$ no longer equal the true impulse responses $\boldsymbol{\Theta}_h$ (i.e., $\boldsymbol{B}_h \neq \boldsymbol{\Theta}_h$). However, following Galvao and Kato (2014), $\boldsymbol{B}_h$ 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, $\boldsymbol{B}_h$ provides the optimal linear approximation to the true h-period-ahead response $\boldsymbol{\Theta}_h$, even if the underlying model is locally misspecified. Nonetheless, it is important to recognize that $\widehat{\boldsymbol{B}}_h$ is not necessarily closer to $\boldsymbol{\Theta}_h$ than $\widehat{\boldsymbol{A}}^h$, 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 $\boldsymbol{B}_h$:

Lemma 1. *Under Assumption 1, allowing for local reduced-form VAR misspecification, pseudo-true IRs $\boldsymbol{B}_h$ are given by:*

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

where $\boldsymbol{C}_j = \boldsymbol{\phi}_j\boldsymbol{\Gamma}^{-1}$, $\boldsymbol{\phi}_j = \mathbb{E}[\boldsymbol{\eta}_{t+j}\mathbf{y}_t']$ and $\boldsymbol{\Gamma} = \mathbb{E}[\mathbf{y}_t\mathbf{y}_t']$.
Proof. See Appendix A.

Lemma 1 shows that the LP coefficients $\boldsymbol{B}_h$ deviate from the VAR IRs $\boldsymbol{A}^h$ whenever $\boldsymbol{C}_j \neq \mathbf{0}$ for some $j \leq h$. This occurs under local misspecification, where future VAR errors $\boldsymbol{\eta}_{t+j}$ are correlated with current regressors $\mathbf{y}_t$. These forward dependence terms $\boldsymbol{\phi}_j = \mathbb{E}[\boldsymbol{\eta}_{t+j}\mathbf{y}_t']$ give rise to misspecification terms $\boldsymbol{C}_j = \boldsymbol{\phi}_j\boldsymbol{\Gamma}^{-1}$, 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, under local reduced-form VAR misspecification, the LP errors

$\boldsymbol{e}_{t+h,h}$ cannot be expressed purely as a VMA process of Wold innovations and IRs. Instead, the misspecification terms $\boldsymbol{C}_j$ generate additional dependence, giving rise to the following structure:

Theorem 1. *Under Assumption 1, in a fixed-p setting allowing for local reduced-form VAR misspecification, the horizon-h LP errors $\boldsymbol{e}_{t+h,h}$ follow a VMA process of order $(h-1)$, expressed as:*

$$\boldsymbol{e}_{t+h,h} = \boldsymbol{B}_{h-1}\boldsymbol{\nu}_{t+1,1} + \boldsymbol{B}_{h-2}\boldsymbol{\nu}_{t+2,2} + \ldots + \boldsymbol{B}_1\boldsymbol{\nu}_{t+h-1,h-1} + \boldsymbol{\nu}_{t+h,h}, \tag{8}$$

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

$$\boldsymbol{\nu}_{s,j} = \boldsymbol{\eta}_s - \boldsymbol{C}_j\mathbf{y}_{s-j} - \sum_{\ell=1}^{j-1}\boldsymbol{C}_{j-\ell}\boldsymbol{\nu}_{s-j+\ell,\ell}. \tag{9}$$

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.

The VMA representation in Theorem 1 nests the result of Lusompa (2023). Under correct reduced-form VAR specification with finite p, the misspecification terms $\boldsymbol{C}_j$ are zero, while under a sieve framework with $p = p(T) \to \infty$, they vanish asymptotically as $T \to \infty$. In both cases, the re-centered errors $\boldsymbol{\nu}_{s,j}$ in eq.(7) reduce to the VAR projection errors $\boldsymbol{\eta}_s$, and the pseudo-true coefficients $\boldsymbol{B}_h$ in Lemma 1 converge to the VAR IRs $\boldsymbol{A}^h$. Consequently, under these conditions, the general VMA representation in Theorem 1 simplifies to the form given in eq.(7).

2.4 LP GLS Estimators

In this section, we present the different GLS transformations of LPs. These estimators share a common structure and differ only in the choice of residuals used for conditioning, so they are nested in the following general expression:

$$\widehat{\boldsymbol{B}}^{\text{GLS}}_{h,(-a)} = \left(\sum_{t=1}^{T-a}\left(\mathbf{y}_{t+h} - \boldsymbol{\Psi}_{t,h}\right)\mathbf{y}_t'\right)\left(\sum_{t=1}^{T-a}\mathbf{y}_t\mathbf{y}_t'\right)^{-1}, \quad \text{for} \quad h \leq a \leq T-1, \tag{10}$$

initialized with $\widehat{\boldsymbol{B}}^{\text{GLS}}_{1,(-a)} = \widehat{\boldsymbol{B}}_{1,(-a)}$ and maintaining the convention $\widehat{\boldsymbol{B}}^{\text{GLS}}_{0,(-a)} = \boldsymbol{I}_k$. The specification of the transformation term $\boldsymbol{\Psi}_{t,h}$ 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 $\widehat{\boldsymbol{B}}^{\text{GLS}}_{h,(-h)} = \widehat{\boldsymbol{B}}^{\text{GLS}}_h$ when the maximum available sample is used at each horizon (such that $a = h$).

2.4.1 LP GLS Estimators Based on Auxiliary VAR Residuals

Under correct reduced-form VAR specification, the VMA expression in eq.(7) allows the LP in eq.(4) to be written as:

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

with $\boldsymbol{\Theta}_0 = \boldsymbol{I}_k$. Estimates of $\boldsymbol{\eta}_{t+j}$ are given by $\widehat{\boldsymbol{\eta}}_{t+j,(-a)} = \mathbf{y}_{t+j} - \widehat{\boldsymbol{A}}_{(-a)} \mathbf{y}_{t+j-1}$ and are readily available from the VAR in eq.(2), while the IRs $(\boldsymbol{\Theta}_{h-1}, \ldots, \boldsymbol{\Theta}_1)$ 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 $(\boldsymbol{\eta}_{t+1}, \ldots, \boldsymbol{\eta}_{t+h-1})$ at horizon h, while excluding $\boldsymbol{\eta}_{t+h}$. The corresponding feasible GLS estimator, $\widehat{\boldsymbol{B}}_{h,(-a)}^{\text{Lu}}$, is constructed iteratively by setting $\boldsymbol{\Psi}_{t,h} = \sum_{j=1}^{h-1} \widehat{\boldsymbol{B}}_{h-j,(-a)}^{\text{Lu}} \widehat{\boldsymbol{\eta}}_{t+j,(-a)}$ in eq.(10).

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

Since there is no compelling reason to exclude either $\boldsymbol{\eta}_{t+1}$ or $\boldsymbol{\eta}_{t+h}$ from the conditioning set at horizon h, we also consider an extended LP GLS estimator that naturally conditions on the full set of VAR residuals $(\boldsymbol{\eta}_{t+1}, \ldots, \boldsymbol{\eta}_{t+h})$, thereby using all available information from the VAR. The corresponding feasible GLS estimator, $\widehat{\boldsymbol{B}}_{h,(-a)}^{\eta}$, is constructed iteratively by setting $\boldsymbol{\Psi}_{t,h} = \sum_{j=1}^{h} \widehat{\boldsymbol{B}}_{h-j,(-a)}^{\eta} \widehat{\boldsymbol{\eta}}_{t+j,(-a)}$.

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 AR(∞) 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.

[1] Breitung and Brüggemann (2023) propose transforming eq.(4) by moving $\widehat{\boldsymbol{\eta}}_{t+h}$ to the left-hand side and including $(\widehat{\boldsymbol{\eta}}_{t+2}, \ldots, \widehat{\boldsymbol{\eta}}_{t+h-1})$ 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.

2.4.2 LP GLS Based on Previous-Horizon LP Residuals

Under local reduced-form VAR specification, 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:

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

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

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 low-variance VAR estimator or retain the robustness of LP OLS. All results are derived under Assumption 1 in a fixed-p setting, which allows for local reduced-form VAR misspecification. We also indicate how each result nests the correct-specification case, including the sieve setting with $p = p(T) \to \infty$ as $T \to \infty$. We distinguish between two strands of LP GLS estimators — those based on auxiliary VAR residuals and those based on previous-horizon LP residuals — which reflect fundamentally different properties.

3.1 Equivalence Properties of LP GLS Based on Auxiliary VAR Residuals

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

Proposition 1. *Under Assumption 1, in a fixed-p setting allowing for local reduced-form VAR misspecification, the LP GLS estimator $\widehat{\boldsymbol{B}}^{\eta}_h$, which fully incorporates VAR residuals in the GLS transformation, is numerically identical to the VAR IR estimator $\widehat{\boldsymbol{A}}^h$: $\widehat{\boldsymbol{B}}^{\eta}_h = \widehat{\boldsymbol{A}}^h$ for all $h = 1, \ldots, H$.*
Proof. See Appendix A.

Intuitively, the estimated h-step-ahead forward iterated VAR

$$\mathbf{y}_{t+h} = \widehat{\boldsymbol{A}}^h \mathbf{y}_t + \sum_{j=1}^{h} \widehat{\boldsymbol{A}}^{h-j} \widehat{\boldsymbol{\eta}}_{t+j},$$

shows that conditioning on $(\widehat{\boldsymbol{\eta}}_{t+1}, \ldots, \widehat{\boldsymbol{\eta}}_{t+h})$ eliminates all error terms on the right-hand side of the LP equation. The GLS transformation $\mathbf{y}_{t+h} - \sum_{j=1}^{h} \widehat{\boldsymbol{B}}_{h-j} \widehat{\boldsymbol{\eta}}_{t+j}$ therefore collapses to the VAR law of motion, ensuring by construction that $\widehat{\boldsymbol{B}}_h^{\eta} = \widehat{\boldsymbol{A}}^h$. As a result, implementing the LP GLS estimator $\widehat{\boldsymbol{B}}_h^{\eta}$ offers no additional value, as it simply reproduces the VAR IRs $\widehat{\boldsymbol{A}}^h$.

This numerical equivalence holds for any finite sample size T and lag length p, regardless of whether the reduced-form VAR is correctly specified. Consequently, if the lag length grows with T (a sieve, $p = p(T) \to \infty$), the result continues to hold pointwise in T, since for each sample size the same numerical equivalence applies.

The LP GLS estimator proposed by Lusompa (2023) — henceforth LP GLS-Lu — conditions on all available VAR residuals except the current-horizon residual $\widehat{\boldsymbol{\eta}}_{t+h}$. The following corollary establishes its relationship with the LP OLS estimator $\widehat{\boldsymbol{B}}_h$ and the VAR IR estimator $\widehat{\boldsymbol{A}}^h$.

Corollary 1. *Under Assumption 1, in a fixed-p setting allowing for local reduced-form VAR misspecification, the LP GLS estimator $\widehat{\boldsymbol{B}}_h^{\mathrm{Lu}}$ of Lusompa (2023) deviates from the LP OLS and VAR IR estimators $\widehat{\boldsymbol{B}}_h$ and $\widehat{\boldsymbol{A}}^h$ as $T \to \infty$ for $h = 2, \ldots, H$:*

$$\widehat{\boldsymbol{B}}_h^{\mathrm{Lu}} = \widehat{\boldsymbol{B}}_h + \boldsymbol{\psi}_h^{\mathrm{B}} + O_p(T^{-1/2}), \tag{11}$$

$$\widehat{\boldsymbol{B}}_h^{\mathrm{Lu}} = \widehat{\boldsymbol{A}}^h + \boldsymbol{\psi}_h^{\mathrm{A}} + O_p(T^{-1/2}), \tag{12}$$

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

Proof. See Appendix A.

Corollary 1 shows that with fixed lag length p, under local reduced-form VAR misspecification, the LP GLS-Lu estimator $\widehat{\boldsymbol{B}}_h^{\mathrm{Lu}}$ asymptotically differs from both LP OLS and VAR IRs by the deterministic deviation terms $\boldsymbol{\psi}_h^{\mathrm{B}}$ and $\boldsymbol{\psi}_h^{\mathrm{A}}$, plus a standard $O_p(T^{-1/2})$ remainder. The deviation from LP OLS arises because $\widehat{\boldsymbol{B}}_h^{\mathrm{Lu}}$ conditions on the intermediate-horizon VAR projection errors $(\boldsymbol{\eta}_{t+1}, \ldots, \boldsymbol{\eta}_{t+h-1})$ in its GLS transformation, thereby imposing most, but not all, of the VAR's dynamic structure. Under local misspecification, some of these projection errors may be correlated with the regressors $\mathbf{y}_t$, which induces a nonzero deviation $\boldsymbol{\psi}_h^{\mathrm{B}}$ from LP OLS. The deviation from the VAR estimator arises because $\widehat{\boldsymbol{B}}_h^{\mathrm{Lu}}$ does not condition on $\boldsymbol{\eta}_{t+h}$, which prevents it from fully replicating the VAR dynamics. If $\boldsymbol{\eta}_{t+h}$ is correlated with $\mathbf{y}_t$, as can occur under local misspecification, its omission from the conditioning set yields a deviation $\boldsymbol{\psi}_h^{\mathrm{A}}$ from the VAR IRs.

Under correct reduced-form VAR specification — either exactly for finite p or along a sieve with $p = p(T) \to \infty$ — the misspecification terms $\boldsymbol{C}_j$ vanish, implying that the LP GLS–Lu, LP OLS, and VAR IR estimators are all consistent for the same impulse responses. Their asymptotic distributions, however, need not coincide. First, residual differences typically remain of order $O_p(T^{-1/2})$, leading to different asymptotic variances. Second, under a locally misspecified sieve, the deviation terms themselves may be $O(T^{-1/2})$, inducing nonzero mean shifts in the asymptotic distribution of the LP GLS-Lu estimator relative to VAR IRs and LP OLS.

Given the generality of Assumption 1, which imposes only reduced-form regularity, a general characterization of the asymptotic distribution is not feasible at this stage: the limiting laws depend on DGP-specific features (e.g., the decay of the VAR(∞) tail) and on the growth rate of $p(T)$. We therefore confine attention here to the fixed-p results established in Corollary 1, and turn in Section 4 to both stylized examples and empirically relevant DGPs that allow us to analytically derive or simulate the bias and variance of VAR IRs, LP OLS, and LP GLS-Lu estimators.

Nonetheless, based on the structure of the GLS transformation — which imposes most but not all of the VAR dynamics — we conjecture that the LP GLS-Lu estimator typically lies between LP OLS and VAR in terms of bias and efficiency: it is likely to be (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 $\widehat{\boldsymbol{B}}_h^{\text{Lu}}$ is uniformly more efficient than LP OLS, the paper does illustrate potential efficiency gains for an AR(1) model.

Remark 4. Corollary 1 reflects a key nonparametric result from Plagborg-Møller and Wolf (2021), showing that when a VAR(p) is estimated and the same p lags are included as controls in the LP, the VAR and LP IR estimands coincide for horizons $h \leq p$, even under local reduced-form VAR misspecification. In line with this result, Corollary 1 shows that for $h \leq p$, the deviation terms $\boldsymbol{\psi}_h^{\text{A}}$ and $\boldsymbol{\psi}_h^{\text{B}}$ 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 $\boldsymbol{\eta}_{t+j}$ are, by construction, orthogonal to $\mathbf{y}_t$ for all $j \leq p$, which implies that the misspecification terms $\boldsymbol{C}_j$ 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.

The LP GLS estimator proposed by Breitung and Brüggemann (2023) conditions on all available VAR residuals except the horizon-1 residual, $\widehat{\boldsymbol{\eta}}_{t+1}$. The following corollary establishes its equivalence to the VAR IR estimator, $\widehat{\boldsymbol{A}}^h$.

Corollary 2. *Under Assumption 1 in a fixed-p setting allowing for local reduced-form VAR misspecification, the LP GLS estimator $\widehat{\boldsymbol{B}}_{h,(-a)}^{\text{BB}}$ proposed by Breitung and Brüggemann (2023) exhibits the following properties relative to the VAR IR estimate $\widehat{\boldsymbol{A}}_{(-a)}^h$, depending on the sample:*

(i) $a = H$: $\widehat{\boldsymbol{B}}_{h,(-H)}^{\text{BB}} = \widehat{\boldsymbol{A}}_{(-H)}^h$ *for all* $h = 1, \ldots, H$.

(ii) $a = h$: $\widehat{\boldsymbol{B}}_h^{\text{BB}} = \widehat{\boldsymbol{A}}^h + O_p(T^{-1})$ *as* $T \to \infty$, *for all* $h = 2, \ldots, H$.

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 $\widehat{\boldsymbol{A}}^h$ when both are computed on the same aligned

sample of $T - H$ observations. This equivalence arises because the transformed disturbance in the LP regression is precisely $\widehat{\boldsymbol{\eta}}_{t+1,(-H)}$ from that window, which is orthogonal to $\mathbf{y}_t$ by the OLS normal equations. Hence, $\widehat{\boldsymbol{B}}_h^{\mathrm{BB}}$ is numerically identical to $\widehat{\boldsymbol{A}}^h$. If instead the VAR residuals $\widehat{\boldsymbol{\eta}}_{t+1}$ are estimated on the full VAR sample $t = 1, \ldots, T-1$ while the horizon-h LP uses the shorter window $t = 1, \ldots, T-h$, orthogonality is broken only by the tail trimming over $t = T-h+1, \ldots, T-1$. Given that h is fixed, this induces a negligible sample–window misalignment of order $O_p(T^{-1})$. The convergence rate is therefore sufficiently fast to ensure that $\widehat{\boldsymbol{B}}_h^{\mathrm{BB}}$ shares the same asymptotic distribution as $\widehat{\boldsymbol{A}}^h$, as already established by Breitung and Brüggemann (2023). Consequently, $\widehat{\boldsymbol{B}}_h^{\mathrm{BB}}$ will typically exhibit a lower variance than the LP OLS estimator $\widehat{\boldsymbol{B}}_h$, but its equivalence to $\widehat{\boldsymbol{A}}^h$ highlights that this variance reduction is achieved by fully imposing the VAR specification across the entire forecast horizon.

The results above hold irrespective of whether the reduced-form VAR is correctly specified. Accordingly, under correct specification — either exact for finite p or along a sieve with $p = p(T) \to \infty$ — the equivalences in Corollary 2 continue to apply. Because the statements are finite-sample identities (or involve only an $O_p(T^{-1})$ sample–window misalignment remainder, independent of p), they also hold pointwise along a sieve, so $\widehat{\boldsymbol{B}}_h^{\mathrm{BB}}$ and $\widehat{\boldsymbol{A}}^h$ remain asymptotically equivalent with identical limit distributions.

3.2 Equivalence Properties of LP GLS Based on Previous-Horizon LP Residuals

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

Proposition 2. *Under Assumption 1, in a fixed-p setting allowing for local reduced-form misspecification, the LP GLS estimator $\widehat{\boldsymbol{B}}_{h,(-a)}^{\nu}$, which uses previous-horizon LP residuals in the GLS transformation, satisfies the following properties relative to LP OLS $\widehat{\boldsymbol{B}}_{h,(-a)}$, depending on the employed sample:*

(i) $a = H$: $\widehat{\boldsymbol{B}}_{h,(-H)}^{\nu} = \widehat{\boldsymbol{B}}_{h,(-H)}$ for all $h = 1, \ldots, H$.

(ii) $a = h$: $\widehat{\boldsymbol{B}}_h^{\nu} = \widehat{\boldsymbol{B}}_h + O_p(T^{-1})$ as $T \to \infty$, for all $h = 2, \ldots, H$.

Proof. See Appendix A.

The key insight of Proposition 2 is that the LP GLS estimator $\widehat{\boldsymbol{B}}_h^{\nu}$, which uses previous-horizon LP residuals, replicates the LP OLS estimator $\widehat{\boldsymbol{B}}_h$. This result holds irrespective of whether the reduced-form VAR is correctly specified. When all LPs are estimated on exactly the same reduced sample of $T - H$ observations across the projection horizon, the orthogonality property of OLS ensures exact numerical equivalence. When instead the longest available sample of $T - h$ observations is used, the difference between the two estimators again arises only from a sample–window misalignment of order $O_p(T^{-1})$. Hence $\widehat{\boldsymbol{B}}_h^{\nu}$ and $\widehat{\boldsymbol{B}}_h$ are asymptotically equivalent

and share the same asymptotic distribution. These equivalences persist under correct specification, including along a sieve, so $\widehat{\boldsymbol{B}}_h^{\nu}$ and $\widehat{\boldsymbol{B}}_h$ remain asymptotically indistinguishable. As a result, GLS estimation using LP residuals offers no practical advantage over standard LP OLS.

4 Illustrative Examples

This section focuses on the Lusompa (2023) estimator, the only LP GLS variant in our analysis with a genuinely distinct asymptotic distribution, and studies its behavior relative to VAR and LP OLS benchmarks under different illustrative designs. As shown in Section 3, all other LP GLS variants are either numerically identical or $\sqrt{T}$–equivalent to VAR IRs or LP OLS. Because their simulated behavior was correspondingly indistinguishable from their benchmark counterparts, we do not report them here to avoid redundancy.

Our evaluation focuses on point estimators, reporting bias, variance, and a weighted RMSE of the IR point estimators. We do not report coverage or interval length, which also reflect properties of the inference procedure (point estimator plus standard errors) rather than the pure bias–variance trade-off of the estimators themselves. Large-scale Monte Carlo evidence on coverage and interval length is already provided by Lusompa (2023). Our results are complementary, isolating the bias–variance mechanisms that help explain those patterns.

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:

$$w_{t+1} = \rho w_t + \beta \mu_{1,t} + \mu_{2,t+1} + \frac{\alpha}{\sqrt{T}} \mu_{2,t}, \tag{13}$$

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

remains latent. The same locally misspecified model — up to timing and normalization conventions — is also used by Li et al. (2022) to compare LP OLS to VAR IRs.

As $T \to \infty$, the $\alpha/\sqrt{T}$ term shrinks to zero, causing the MA component $\mu_{2,t}$ to vanish asymptotically. In the limit, the DGP in eq.(13) thus 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

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

This local misspecification setup captures the idea that finite-order dynamic models provide useful but imperfect representations of the true DGP in finite samples. By introducing a vanishing deviation from the AR(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.(13) 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 $\widehat{\theta}_h^{\text{VAR}} = \mathbf{e}_2' \widehat{\mathbf{A}}^h \mathbf{e}_1 = \widehat{\rho}^{h-1}\widehat{\beta}$, the LP OLS estimator by $\widehat{\theta}_h^{\text{LP}} = \mathbf{e}_2' \widehat{\mathbf{B}}_h \mathbf{e}_1$ and the LP GLS-Lu estimator by $\widehat{\theta}_h^{\text{Lu}} = \mathbf{e}_2' \widehat{\mathbf{B}}_h^{\text{Lu}} \mathbf{e}_1$, with $\widehat{\mathbf{A}}$, $\widehat{\mathbf{B}}_h$ and $\widehat{\mathbf{B}}_h^{\text{Lu}}$ defined in Section 2. We then obtain the following result:

Proposition 3. *Consider the DGP in eq.*(13)*, with* $|\rho| < 1$ *and* $\alpha \in \mathbb{R}$*. Assume* $\sigma_j^2 > 0$ *and* $\mathbb{E}(\mu_{j,t}^4) < \infty$ *for* $j = 1, 2$*, and define* $\sigma_w^2 = (\beta^2\sigma_1^2 + \sigma_2^2)/(1 - \rho^2)$*. Then, as* $T \to \infty$

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

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* $h \geq 1$*,*

$$b_h^{\text{VAR}} = (h-1)\,\rho^{h-2}\beta\frac{\alpha\sigma_2^2}{\sigma_w^2}, \qquad V_h^{\text{VAR}} = \rho^{2(h-1)}\frac{\sigma_2^2}{\sigma_1^2} + (h-1)^2\,\frac{\sigma_2^2}{\sigma_w^2}\rho^{2(h-2)}\beta^2,$$

$$b_h^{\text{LP}} = 0, \qquad V_h^{\text{LP}} = \left(1 - \rho^{2h}\right)\frac{\sigma_w^2}{\sigma_1^2} - \rho^{2(h-1)}\beta^2.$$

For all $h \geq 2$,

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

For $h = 1, 2$, *we have* $\widehat{\theta}_h^{\mathrm{Lu}} = \widehat{\theta}_h^{\mathrm{LP}}$.
The ranking in terms of bias magnitude and variance is as follows for $h > 2$*:*

$$|b_h^{\mathrm{VAR}}| > |b_h^{\mathrm{Lu}}| > |b_h^{\mathrm{LP}}| = 0, \quad \text{for } \alpha \neq 0,\ \rho \neq 0,\ \beta \neq 0,$$

$$V_h^{\mathrm{LP}} > V_h^{\mathrm{Lu}} > V_h^{\mathrm{VAR}}, \quad \text{for } (\rho, \beta) \neq (0, 0).$$

If any of $\alpha = 0$, $\beta = 0$, *or* $\rho = 0$, *then* $b_h^{\mathrm{VAR}} = b_h^{\mathrm{Lu}} = b_h^{\mathrm{LP}} = 0$. *If* $\rho = \beta = 0$, *then* $V_h^{\mathrm{VAR}} = 0$ *and* $V_h^{\mathrm{Lu}} = V_h^{\mathrm{LP}} = \sigma_2^2/\sigma_1^2$.

Proof. See Appendix B (Online Supplementary Material).

The bias ranking reflects the fact 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 imposes most — but not all — of the VAR dynamics on the LP, resulting in a bias that is typically closer to that of the VAR than to LP OLS. Similarly, the variance ranking mirrors the extent to which model structure is imposed: the VAR achieves the lowest variance, LP GLS–Lu attains an intermediate level by partially exploiting VAR dynamics, and LP OLS exhibits the highest variance owing to its minimal structure. Figure 1 plots the asymptotic bias and standard deviation across horizons under low and high persistence ($\rho = 0.6$ and $\rho = 0.9$), visualizing these trade-offs.

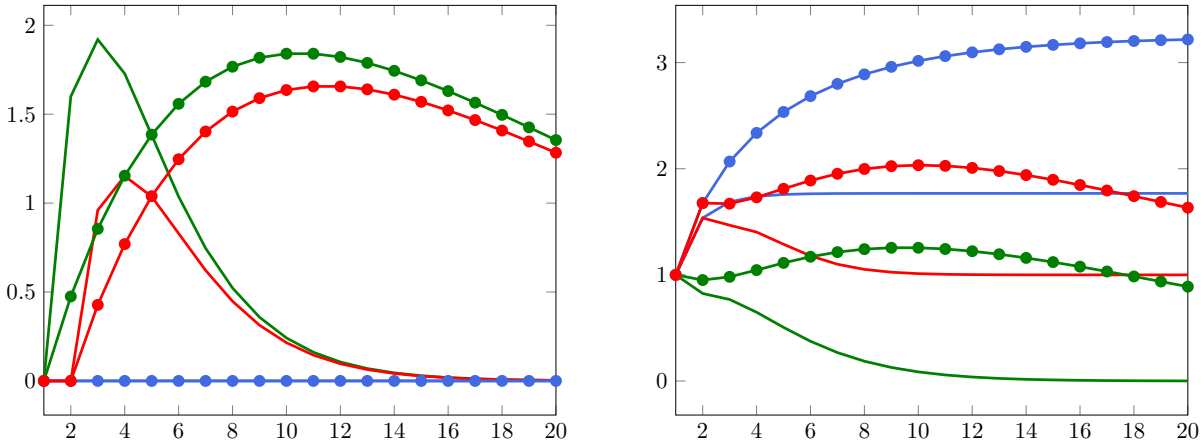
To assess overall performance, we compute a weighted RMSE:

$$\mathrm{RMSE}_{\lambda,h}^{\mathrm{est}} = \sqrt{\lambda\left(b_h^{\mathrm{est}}\right)^2 + (1-\lambda)V_h^{\mathrm{est}}}, \tag{14}$$

for each estimator $\mathrm{est} \in \{\mathrm{VAR}, \mathrm{LP}, \mathrm{Lu}\}$, 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 horizons $h = 1, \ldots, 20$ 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

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



(a) Asymptotic Bias **(b)** Asymptotic Standard Deviation

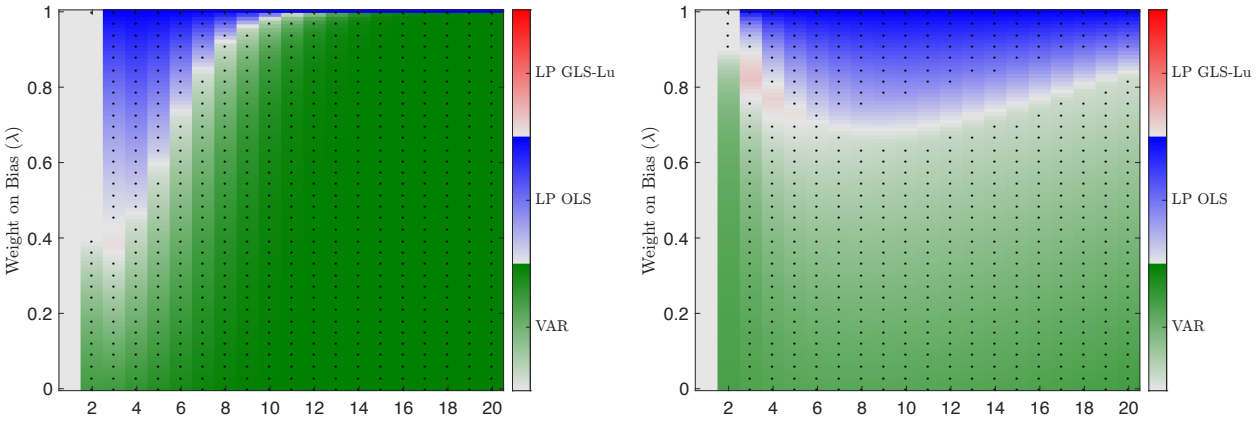
VAR LP OLS LP GLS-Lu $(\rho = 0.6)$

VAR LP OLS LP GLS-Lu $(\rho = 0.9)$

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.(13), 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 $h = 1, \ldots, 20$.

those cases. Overall, under this stylized local misspecification, the LP GLS-Lu estimator offers no improvement over the benchmark VAR and LP OLS estimators.

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



(a) $\rho = 0.6$ (b) $\rho = 0.9$

Notes: The heatmaps visualize estimator dominance across forecast horizons ($h = 1, \ldots, 20$), 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.(14), computed from data simulated from eq.(13) 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.

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:

$$w_{t+1} = \rho w_t + \beta \mu_{1,t} + \mu_{2,t+1} + \alpha \mu_{2,t-4}, \tag{15}$$

which replaces the vanishing misspecification term $\alpha/\sqrt{T}$ from eq.(13) with a fixed α. As a result, the DGP no longer converges to a finite-order AR model, 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 $p = \lfloor T^{1/4} \rfloor$.

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 report the bias and standard deviation of each estimator and provide heatmaps indicating, for each horizon $h = 1, \ldots, 20$ and each bias weight $\lambda \in [0, 1]$, which method minimizes the weighted RMSE. 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 more 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. business 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 we treat the shock as observed and include it 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.

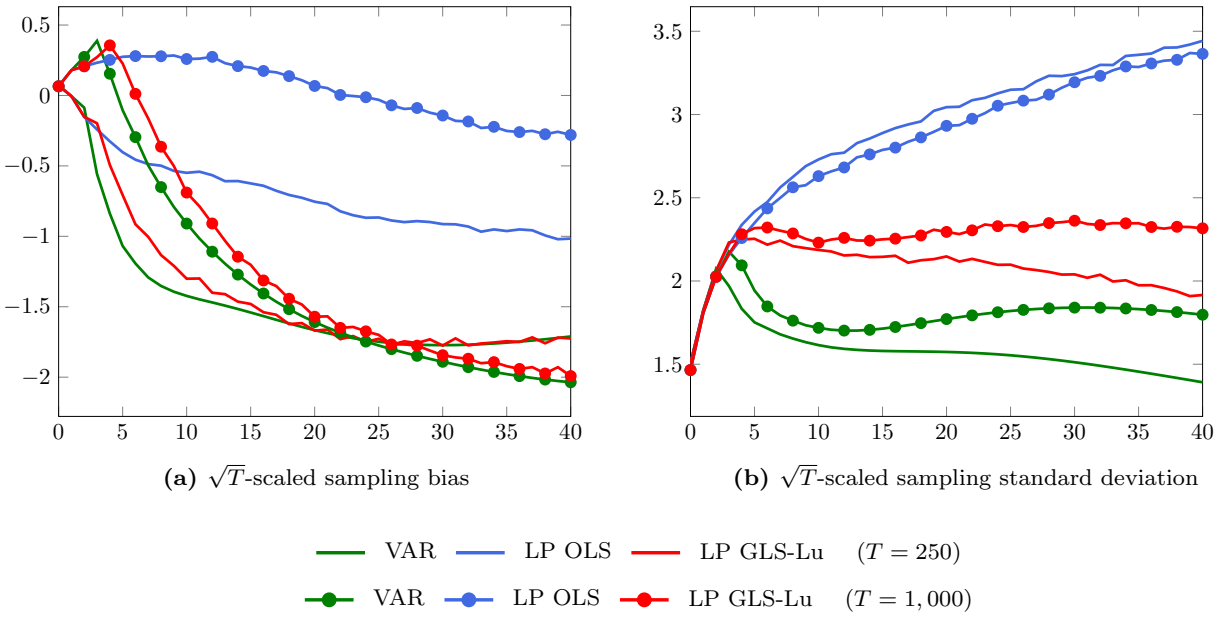
We consider two sample sizes, $T \in \{250, 1000\}$. The smaller sample ($T = 250$) reflects typical macroeconomic applications, while the larger one ($T = 1000$) allows us to assess how estimator performance evolves with increased sample size. The VAR lag order p is selected using the AIC, subject to 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 reported 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. We scale these quantities by $\sqrt{T}$ so that the Monte Carlo summaries align with the objects appearing in the asymptotic distributions, thereby allowing meaningful comparison across 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 $h = 0, \ldots, 40$ 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

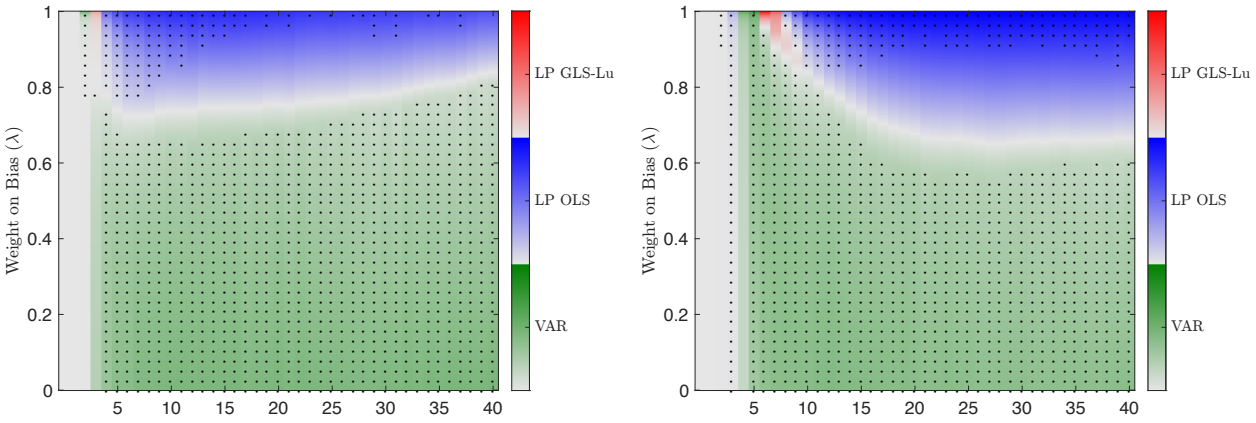


(a) $\sqrt{T}$-scaled sampling bias

(b) $\sqrt{T}$-scaled sampling standard deviation

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 $h = 0, \ldots, 40$.

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



(a) $T = 250$

(b) $T = 1{,}000$

Notes: Displayed are heatmaps of the estimator minimizing the weighted RMSE defined in eq.(14), 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 $h = 0, \ldots, 40$; 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 Simulations Based on the Stock and Watson (2016) Dynamic Factor Model

Our final simulation study generates data from the DFM of Stock and Watson (2016), estimated on 207 quarterly U.S. time series for 1959Q1–2014Q4. This DFM is considered to be sufficiently rich to capture the key time series properties of macroeconomic data and has become the benchmark for large-scale Monte Carlo designs (see e.g. Lazarus et al., 2018; Li et al., 2024). We follow the setup in Li et al. (2024), replicating their framework for comparing VAR IR and LP OLS estimators and extending the analysis by including the LP GLS–Lu estimator.[2] We briefly summarize the main elements of this design below.

The DFM is specified with six latent factors. We consider both a specification in differences and in levels. In the differenced version, both the factor process and idiosyncratic components are estimated with two lags, whereas in the levels version, the factor process is estimated as a VECM with four lags and the idiosyncratic components as AR(4) processes. Since our theoretical framework assumes stationarity, the differenced specification aligns most closely with our analysis, but we also include the levels specification for comparability with Li et al. (2024).

We then define the set of DGPs by randomly selecting subsets of variables from the 207-series DFM. For each specification (differences and levels), we construct 6,000 DGPs: 3,000 fiscal policy VARs, which always include government spending, and 3,000 monetary policy VARs, which always include the federal funds rate. In addition to the policy variable, four other distinct series are randomly drawn, subject to the restriction that at least one series must measure real activity and at least one must measure prices. One of these four is then randomly chosen as the response variable of interest. This procedure yields a total of 12,000 DGPs across the two specifications.

For each DGP, the corresponding true impulse responses $\{\theta_h\}_{h=0}^{H}$ are computed up to horizon $H = 20$ from the state-space representation of the DFM. In line with our focus on reduced-form estimation, we adopt the observed shock identification scheme, which treats the observed fiscal or monetary policy shock as directly entering the VAR and allows impulse responses to be defined without imposing additional structural assumptions.

In each Monte Carlo replication, we simulate $T = 200$ observations from the full 207-variable DFM and, for each DGP, retain the relevant subset of variables. The observed fiscal or monetary policy shock is constructed from the simulated data following the procedure in Li et al. (2024), and placed as the first variable in the system. The sampling distribution of the estimators around the true impulse responses is approximated using 5,000 Monte Carlo replications. We consider two lag-length choices: a fixed lag length of $p = 4$ and selection by AIC. The AIC almost always chooses a very short lag length (one or two), whereas practitioners working with quarterly data typically include at least four lags. For this reason, as in Li et al. (2024), we report the results for $p = 4$ in the main text and relegate the AIC results to Appendix B

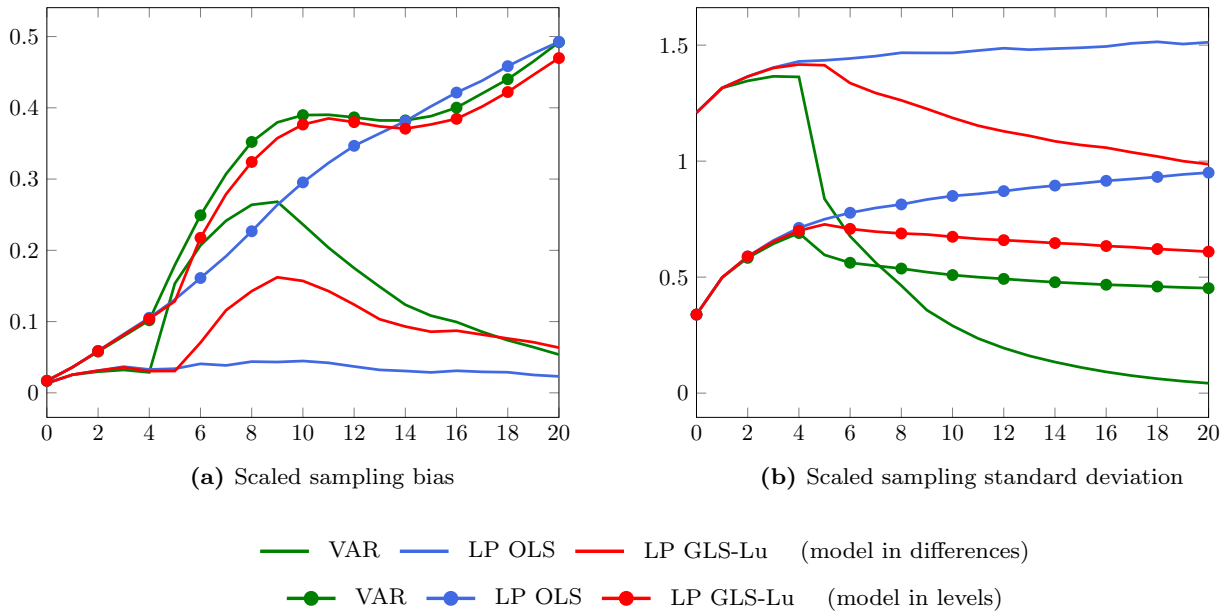
[2]We gratefully use the replication package of Li et al. (2024) to implement the Stock and Watson (2016) DFM and Monte Carlo design, extending it to incorporate the LP GLS-Lu estimator.

Figure 5 depicts the bias–variance trade-off across horizons. Displayed are the medians across the 6,000 DGPs of the absolute bias $|\mathbb{E}(\hat{\theta}_h) - \theta_h|$ and the standard deviation of $\hat{\theta}_h$ for the different estimation procedures, scaled by $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$ to remove the units of the response variable. As before, a clear bias–variance trade-off emerges: LP OLS typically exhibits lower bias than VAR IRs, but this comes at the cost of substantially higher variance at intermediate and longer horizons. The LP GLS–Lu estimator tends to fall between these two benchmarks in terms of both bias and variance.

Figure 6 summarizes overall performance in terms of weighted RMSE across horizons and loss-function weights. The heatmaps show which estimator dominates most frequently across the 6,000 DGPs, with shading indicating the strength of dominance and dots marking cases where LP GLS–Lu is the runner-up. Note that the dark-green regions indicate cases where the VAR uniformly dominates across all horizons and weights, leaving no meaningful runner-up. This comparison confirms that GLS–Lu almost never provides a clear improvement in overall performance, reinforcing the patterns already documented in the preceding designs.

Similar patterns emerge when lag length is selected by AIC. These results, reported in Appendix B, are somewhat more favorable to LP OLS since the AIC typically chooses a shorter lag length (one or two rather than four). Nevertheless, LP GLS–Lu again fails to deliver a consistent improvement in overall performance.

Figure 5: Scaled Bias and Standard Deviation — Stock–Watson DFM (lag length fixed at 4)

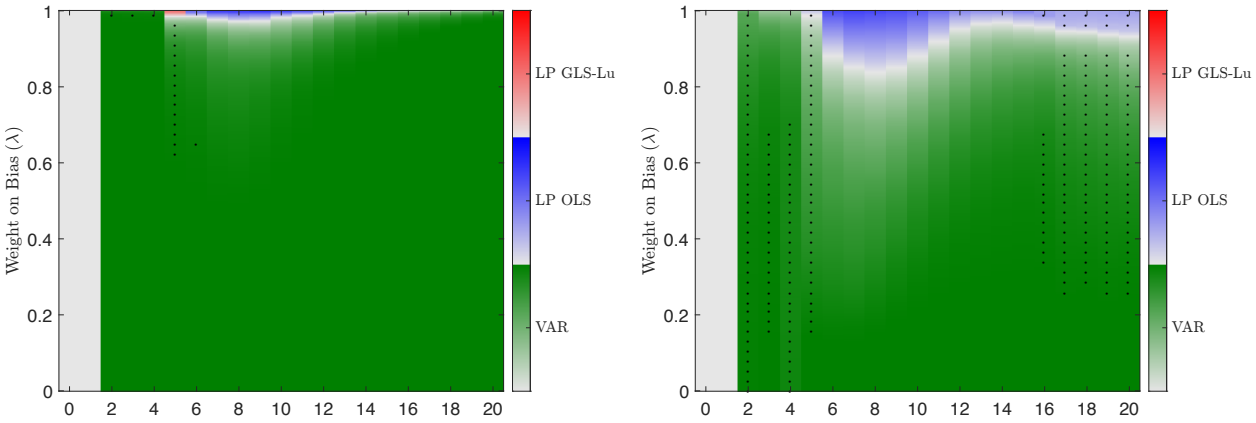


(a) Scaled sampling bias

(b) Scaled sampling standard deviation

Notes: Displayed are the medians (across 6,000 DGPs) of the absolute bias $|\mathbb{E}(\hat{\theta}_h) - \theta_h|$ and the standard deviation of $\hat{\theta}_h$ for the different estimation procedures, scaled by $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$, i.e., the root mean squared value of the true impulse responses $\{\theta_h\}_{h=0}^{20}$. For each DGP, bias and standard deviation are computed from 5,000 Monte Carlo replications based on data simulated from the Stock and Watson (2016) DFM, as implemented in Li et al. (2024), with $T = 200$. The VAR lag length is fixed at $p = 4$ and applied uniformly across the VAR, LP OLS, and LP GLS–Lu estimators. The horizontal axis indicates the projection horizon $h = 0, \ldots, 20$.

Figure 6: Estimator Dominance by Weighted RMSE — Stock–Watson DFM (lag length fixed at 4)



(a) Model in differences (b) Model in levels

Notes: Displayed are heatmaps of the estimator that attains the lowest weighted RMSE defined in eq.(14). For each (h, λ) combination, the "winner" is the estimator that most frequently minimizes the loss across the 6,000 DGPs simulated from the Stock and Watson (2016) DFM, following the design of Li et al. (2024). Color shading indicates the strength of dominance, with darker colors corresponding to higher frequencies and the darkest shade indicating that the estimator always wins. Dots mark cases where the LP GLS–Lu estimator is the second-best procedure. Results are based on 5,000 Monte Carlo replications per DGP with $T = 200$. The VAR lag length is fixed at $p = 4$ for the VAR, LP OLS, and LP GLS–Lu estimators. The horizontal axis indicates the projection horizon $h = 0, \ldots, 20$; the vertical axis varies the squared-bias weight $\lambda \in [0, 1]$.

5 Conclusion

This paper 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, the second uses previous-horizon LP residuals.

The central insight of our analysis is that these two strands of GLS implementations do not produce estimators that are genuinely distinct from the benchmark approaches; instead, they tend to align with VAR IRs or LP OLS. Conditioning on all VAR residuals imposes the dynamic structure of the VAR onto the LP framework, causing the LP GLS estimator to collapse to the VAR IRs — gaining precision under correct specification but losing robustness. Conditioning on LP residuals, by contrast, retains the flexibility of the LP framework but yields estimators that are equivalent to LP OLS. These equivalence results are derived under minimal reduced-form assumptions and therefore hold for a broad class of stationary multivariate processes, regardless of whether the auxiliary VAR is correctly specified or locally misspecified.

The only exception is the LP GLS estimator proposed by Lusompa (2023), which conditions only 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 root mean squared error. These patterns

persist in simulations based on the Smets and Wouters (2005) DSGE model and the Stock and Watson (2016) DFM.

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.

Bruns, M. and Lütkepohl, H. (2022). Comparison of Local Projection Estimators For Proxy Vector Autoregressions. *Journal of Economic Dynamics and Control*, 134:104272.

Faust, J. and Wright, J. H. (2013). Efficient prediction of excess returns. *The Review of Economics and Statistics*, 95(4):1223–1233.

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.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press, New Jersey.

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.

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

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.

Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018). Har inference: Recommendations for practice. *Journal of Business & Economic Statistics*, 36(4):541–559.

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.

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.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

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. (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.

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. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2A, pages 415–525. Elsevier, Amsterdam.

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

Throughout the proofs we maintain Assumption 1. 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 $h/T \to 0$ as $T \to \infty$ for all $h = 1, \ldots, H$.

A.1 Preliminaries: Notation and Useful Results

Backward iteration identities. For the VAR(1) in eq.(2), estimated on $t = 1, \ldots, T-a$, repeated substitution yields, for any $h \geq 1$,

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

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

$$= \widehat{\boldsymbol{A}}_{(-a)}^{h} \mathbf{y}_{t} + \sum_{j=1}^{h} \widehat{\boldsymbol{A}}_{(-a)}^{h-j} \widehat{\boldsymbol{\eta}}_{t+j,(-a)}. \tag{A-1c}$$

Here $\widehat{\mathbf{A}}_{(-a)}^{0} = \mathbf{I}_k$. 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).

Sample second moments and inverse. Under Assumption 1 (stationarity, finite fourth moments, absolutely summable Wold coefficients), it holds for fixed h (see, e.g., Hamilton, 1994; Lütkepohl, 2005),

$$\widehat{\boldsymbol{\Gamma}}_{T-h} = \frac{1}{T} \sum_{t=1}^{T-h} \boldsymbol{y}_t \boldsymbol{y}_t' = \boldsymbol{\Gamma} + O_p(T^{-1/2}), \qquad \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} = \boldsymbol{\Gamma}^{-1} + O_p(T^{-1/2}), \tag{A-2}$$

where $\boldsymbol{\Gamma} = \mathbb{E}(\boldsymbol{y}_t \boldsymbol{y}_t')$ is positive definite with eigenvalues bounded away from 0 and ∞.

OLS $\sqrt{T}$–consistency. For fixed $h \geq 1$, the OLS normal equations give

$$\widehat{\boldsymbol{B}}_h = \Big(\tfrac{1}{T} \sum_{t=1}^{T-h} \boldsymbol{y}_{t+h} \boldsymbol{y}_t'\Big) \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} = \boldsymbol{B}_h + \Big(\tfrac{1}{T} \sum_{t=1}^{T-h} \boldsymbol{e}_{t+h,h} \boldsymbol{y}_t'\Big) \widehat{\boldsymbol{\Gamma}}_{T-h}^{-1},$$

where, under Assumption 1, $\frac{1}{T} \sum_{t=1}^{T-h} \boldsymbol{e}_{t+h,h} \boldsymbol{y}_t' = O_p(T^{-1/2})$ and, by eq.(A-2), $\widehat{\boldsymbol{\Gamma}}_{T-h}^{-1} = O_p(1)$. Hence

$$\widehat{\boldsymbol{B}}_h = \boldsymbol{B}_h + O_p(T^{-1/2}), \qquad \widehat{\boldsymbol{A}} = \boldsymbol{A} + O_p(T^{-1/2}), \tag{A-3}$$

with the second relation obtained by setting $h = 1$ so that $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{B}}_1$.

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

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

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

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

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 $\widehat{\boldsymbol{A}}_{(-a)}$, we have:

$$\widehat{\boldsymbol{A}}_{(-a)} = \widehat{\boldsymbol{\gamma}}_{T-a}\widehat{\boldsymbol{\Gamma}}_{T-a}^{-1} = \widehat{\boldsymbol{\gamma}}_{T-1}\widehat{\boldsymbol{\Gamma}}_{T-1}^{-1} + O_p(T^{-1}) = \widehat{\boldsymbol{A}} + O_p(T^{-1}). \tag{A-4}$$

VAR residual forward moments. Using $\widehat{\boldsymbol{\eta}}_{t+j} = \boldsymbol{\eta}_{t+j} - (\widehat{\boldsymbol{A}} - \boldsymbol{A})\,\mathbf{y}_{t+j-1}$, for any $j \ge 1$,

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

By (A-3), $\widehat{\boldsymbol{A}} - \boldsymbol{A} = O_p(T^{-1/2})$, and under Assumption 1 $\frac{1}{T}\sum_{t=1}^{T-h}\mathbf{y}_{t+j-1}\mathbf{y}_t' = O_p(1)$. Similarly, with $\boldsymbol{\phi}_j = \mathbb{E}[\boldsymbol{\eta}_{t+j}\mathbf{y}_t']$ and by standard results for series with finite 4th moments and absolutely summable Wold coefficients, we have $\frac{1}{T}\sum_{t=1}^{T-h}\boldsymbol{\eta}_{t+j}\mathbf{y}_t' = \boldsymbol{\phi}_j + O_p(T^{-1/2})$. Therefore

$$\widehat{\boldsymbol{\phi}}_j = \boldsymbol{\phi}_j + O_p(T^{-1/2}), \qquad j \ge 1, \tag{A-5}$$

and since $\boldsymbol{\phi}_1 = \mathbf{0}$ we have $\widehat{\boldsymbol{\phi}}_1 = O_p(T^{-1/2})$.

A.2 Proof of Lemma 1

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

$$\mathbb{E}\left[\mathbf{y}_{t+h} \mid \mathbf{y}_t\right] = \boldsymbol{B}_h\mathbf{y}_t. \tag{A-6}$$

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

$$\mathbb{E}\left[\mathbf{y}_{t+h} \mid \mathbf{y}_t\right] = \boldsymbol{A}^h\mathbf{y}_t + \sum_{j=1}^{h}\boldsymbol{A}^{h-j}\mathbb{E}\left[\boldsymbol{\eta}_{t+j} \mid \mathbf{y}_t\right] = \left(\boldsymbol{A}^h + \sum_{j=1}^{h}\boldsymbol{A}^{h-j}\boldsymbol{C}_j\right)\mathbf{y}_t. \tag{A-7}$$

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

Hence, equating (A-7) to the projection definition in eq.(A-6) reveals that:

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

□

A.3 Proof of Theorem 1

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

$$\boldsymbol{e}_{t+h,h} = \mathbf{y}_{t+h} - \boldsymbol{B}_h \mathbf{y}_t = \mathbf{y}_{t+h} - \left(\boldsymbol{A}^h + \sum_{j=1}^{h} \boldsymbol{A}^{h-j} \boldsymbol{C}_j \right) \mathbf{y}_t = \sum_{j=1}^{h} \boldsymbol{A}^{h-j} (\boldsymbol{\eta}_{t+j} - \boldsymbol{C}_j \mathbf{y}_t), \tag{A-8}$$

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

$$\boldsymbol{A}^{h-j} = \boldsymbol{B}_{h-j} - \sum_{\ell=j+1}^{h} \boldsymbol{A}^{h-\ell} \boldsymbol{C}_{\ell-j}. \tag{A-9}$$

Accordingly, setting $j = 1$ in eq.(A-9) allows substitution of $\boldsymbol{A}^{h-1}$ into eq.(A-8) to obtain

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

where $\boldsymbol{\nu}_{t+1,1} = \boldsymbol{\eta}_{t+1} - \boldsymbol{C}_1 \mathbf{y}_t$. In similar fashion, substituting $\boldsymbol{A}^{h-2}$ using eq.(A-9) and continuing recursively in this way yields

$$\boldsymbol{e}_{t+h,h} = \boldsymbol{B}_{h-1} \boldsymbol{\nu}_{t+1,1} + \boldsymbol{B}_{h-2} \boldsymbol{\nu}_{t+2,2} + \ldots + \boldsymbol{B}_1 \boldsymbol{\nu}_{t+h-1,h-1} + \boldsymbol{\nu}_{t+h,h},$$

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

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

□

A.4 Proofs for the Results in the Propositions

A.4.1 Proof of Proposition 1

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

$$\mathbf{y}_{t+h} = \widehat{\boldsymbol{A}}^{h}\mathbf{y}_t + \sum_{j=1}^{h}\widehat{\boldsymbol{A}}^{h-j}\widehat{\boldsymbol{\eta}}_{t+j}. \tag{A-10}$$

Substituting this VAR expansion into the implemented GLS transformation $\mathbf{y}^{\eta}_{t+h} = \mathbf{y}_{t+h} - \boldsymbol{\Psi}_{t,h}$ of eq.(10), with $\boldsymbol{\Psi}_{t,h} = \sum_{j=1}^{h}\widehat{\boldsymbol{B}}^{\eta}_{h-j}\widehat{\boldsymbol{\eta}}_{t+j}$, gives in turn:

$$\mathbf{y}^{\eta}_{t+h} = \widehat{\boldsymbol{A}}^{h}\mathbf{y}_t + \sum_{j=1}^{h}\left(\widehat{\boldsymbol{A}}^{h-j} - \widehat{\boldsymbol{B}}^{\eta}_{h-j}\right)\widehat{\boldsymbol{\eta}}_{t+j}. \tag{A-11}$$

We can then proceed by mathematical induction to show that $\widehat{\boldsymbol{B}}^{\eta}_{h} = \widehat{\boldsymbol{A}}^{h}$ for all $h = 1,\ldots,H$. Note that from the initialization in eq.(10), it follows directly for $h = 1$ that $\widehat{\boldsymbol{B}}^{\eta}_{1} = \widehat{\boldsymbol{B}}_{1} = \widehat{\boldsymbol{A}}$, and eq.(A-11) 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 $\widehat{\boldsymbol{B}}^{\eta}_{h-j} = \widehat{\boldsymbol{A}}^{h-j}$ for all $j = 1,\ldots,h$ and any $h \geq 2$, and note that in this case the summation term in eq.(A-11) cancels, leaving $\mathbf{y}^{\eta}_{t+h} = \widehat{\boldsymbol{A}}^{h}\mathbf{y}_t$. Accordingly, substituting this into the expression for the LP GLS estimator in eq.(10) gives as the result for h:

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

We can therefore conclude that $\widehat{\boldsymbol{B}}^{\eta}_{h} = \widehat{\boldsymbol{A}}^{h}$ for all $h = 1,\ldots,H$. □

A.4.2 Proof of Proposition 2

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

$$\widehat{\boldsymbol{B}}^{\nu}_{h,(-a)} - \widehat{\boldsymbol{B}}_{h,(-a)} = -\sum_{j=1}^{h-1}\widehat{\boldsymbol{B}}^{\nu}_{h-j,(-a)}\left(\frac{1}{T}\sum_{t=1}^{T-a}\widehat{\boldsymbol{\nu}}_{t+j,j,(-a)}\mathbf{y}_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T-a}\mathbf{y}_t\mathbf{y}_t'\right)^{-1}. \tag{A-12}$$

Case (i): $a = H$. When a reduced sample of $T - H$ observations is used, $\widehat{\boldsymbol{\nu}}_{t+j,j,(-H)}$ 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 $\boldsymbol{\nu}_{t+j,j,(-H)}$ is the error term for $j = 1,\ldots,h-1$. Making use of this orthogonality in eq.(A-12) shows that $\widehat{\boldsymbol{B}}^{\nu}_{h,(-H)} = \widehat{\boldsymbol{B}}_{h,(-H)}$, proving part (i).

Case (ii): $a = h$. When the longest available sample of $T - h$ observations is used at each horizon

h, $\widehat{\boldsymbol{v}}_{t+j,j}$ and $\mathbf{y}_t$ are orthogonal by construction over the sample period $t = 1, \ldots, T-j$, but the summation in the numerator of eq.(A-12) runs over $t = 1, \ldots, T-h$ with $h > j$. Therefore, we can decompose the summation as follows:

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

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

$$\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\boldsymbol{v}}_{t+j,j}\mathbf{y}_t' = O_p(T^{-1}). \tag{A-13}$$

From eq.(A-2) we know that $\left(\frac{1}{T}\sum_{t=1}^{T-h}\mathbf{y}_t\mathbf{y}_t'\right)^{-1} = \boldsymbol{\Gamma}^{-1} + O_p(T^{-1/2})$, with $\boldsymbol{\Gamma} = \mathbb{E}(\boldsymbol{y}_t\boldsymbol{y}_t') = O(1)$. Substituting this into eq.(A-12) and using the initial conditions $\widehat{\boldsymbol{B}}_0^{\nu} = \mathbf{I}_k$ and $\widehat{\boldsymbol{B}}_1^{\nu} = \widehat{\boldsymbol{B}}_1$, we iteratively find from using eq.(A-13) that:

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

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

A.5 Proofs for the Results in the Corollaries

A.5.1 Proof of eq.(11) in Corollary 1

By setting $\boldsymbol{\Psi}_{t,h} = \sum_{j=1}^{h-1}\widehat{\boldsymbol{B}}^{\mathrm{Lu}}_{h-j}\widehat{\boldsymbol{\eta}}_{t+j}$ in eq.(10), and the shorthand $\widehat{\boldsymbol{\Gamma}}_{T-h} = \frac{1}{T}\sum_{t=1}^{T-h}\mathbf{y}_t\mathbf{y}_t'$, the $\widehat{\boldsymbol{B}}^{\mathrm{Lu}}_h$ estimator can be written as:

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

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

For $h = 1$, $\widehat{\boldsymbol{B}}^{\mathrm{Lu}}_1 = \widehat{\boldsymbol{B}}_1$, so by eq.(A-3)

$$\widehat{\boldsymbol{B}}^{\mathrm{Lu}}_1 = \boldsymbol{B}_1 + O_p(T^{-1/2}). \tag{A-15}$$

For $h = 2$, eq.(A-14) gives

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

Using eq.(A-15), eq.(A-2) and eq.(A-5) (with $j = 1$ and $\boldsymbol{\phi}_1 = \mathbb{E}[\boldsymbol{\eta}_{t+1}\mathbf{y}_t'] = \mathbf{0}$) yields

$$\widehat{\boldsymbol{B}}_2^{\text{Lu}} = \widehat{\boldsymbol{B}}_2 + O_p(T^{-1/2}) = \boldsymbol{B}_2 + O_p(T^{-1/2}). \tag{A-16}$$

For $h = 3$, eq.(A-14) gives

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

where substituting in the result for $\widehat{\boldsymbol{B}}_2^{\text{Lu}}$ and proceeding as above results in

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

Accordingly, by iterating the steps above for general $h > 1$, the recursive structure of eq.(A-14) implies that

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

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

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

with initialization $\boldsymbol{\psi}_1^{\text{B}} = \mathbf{0}$ and where $\boldsymbol{\psi}_2^{\text{B}} = \mathbf{0}$ because $\boldsymbol{\phi}_1 = \mathbb{E}[\boldsymbol{\eta}_{t+1}\mathbf{y}_t'] = \mathbf{0}$. □

A.5.2 Proof of eq.(12) in Corollary 1

Substituting eq.(A-1c) into the expression for $\widehat{\boldsymbol{B}}_h^{\text{Lu}}$, we have for $h \geq 2$:

$$\begin{aligned}
\widehat{\boldsymbol{B}}_h^{\text{Lu}} &= \left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\mathbf{y}_{t+h} - \sum_{j=1}^{h-1}\widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}}\widehat{\boldsymbol{\eta}}_{t+j}\right)\mathbf{y}_t'\right)\widehat{\boldsymbol{\Gamma}}_{T-h}^{-1}, \\
&= \left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\widehat{\boldsymbol{A}}^h\mathbf{y}_t + \sum_{j=1}^{h-1}\left(\widehat{\boldsymbol{A}}^{h-j} - \widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}}\right)\widehat{\boldsymbol{\eta}}_{t+j} + \widehat{\boldsymbol{\eta}}_{t+h}\right)\mathbf{y}_t'\right)\widehat{\boldsymbol{\Gamma}}_{T-h}^{-1}, \\
&= \widehat{\boldsymbol{A}}^h + \left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\sum_{j=1}^{h-1}\left(\widehat{\boldsymbol{A}}^{h-j} - \widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}}\right)\widehat{\boldsymbol{\eta}}_{t+j} + \widehat{\boldsymbol{\eta}}_{t+h}\right)\mathbf{y}_t'\right)\widehat{\boldsymbol{\Gamma}}_{T-h}^{-1}, \\
&= \widehat{\boldsymbol{A}}^h + \left(\widehat{\boldsymbol{\phi}}_h - \sum_{j=1}^{h-1}\left(\widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}} - \widehat{\boldsymbol{A}}^{h-j}\right)\widehat{\boldsymbol{\phi}}_j\right)\widehat{\boldsymbol{\Gamma}}_{T-h}^{-1},
\end{aligned} \tag{A-17}$$

where $\widehat{\boldsymbol{\phi}}_j = \frac{1}{T}\sum_{t=1}^{T-h}\widehat{\boldsymbol{\eta}}_{t+j}\mathbf{y}_t'$ 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 $\widehat{\boldsymbol{B}}_1^{\text{Lu}} = \widehat{\boldsymbol{A}}$ by definition, applying eqs.(A-2) and (A-5) with $\boldsymbol{\phi}_1 = \mathbf{0}$

yields:

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

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

For $h = 3$, we can substitute the result for $h = 2$ into the recursion, and apply eqs.(A-2) and (A-5) as above:

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

Hence, by recursively substituting back into the expression (A-17), we obtain the following result for general $h \geq 2$:

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

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

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

□

A.6 Proof of Corollary 2

Substituting eq.(A-1b) into the expression for $\widehat{\boldsymbol{B}}_{h,(-a)}^{\text{BB}}$, we have:

$$\widehat{\boldsymbol{B}}_{h,(-a)}^{\text{BB}} = \left(\sum_{t=1}^{T-a} \left(\widehat{\boldsymbol{A}}_{(-a)}^{h-1} \mathbf{y}_{t+1} - \sum_{j=2}^{h-2} \left(\widehat{\boldsymbol{B}}_{h-j,(-a)}^{\text{BB}} - \widehat{\boldsymbol{A}}_{(-a)}^{h-j} \right) \widehat{\boldsymbol{\eta}}_{t+j,(-a)} \right) \mathbf{y}_t' \right) \left(\sum_{t=1}^{T-a} \mathbf{y}_t \mathbf{y}_t' \right)^{-1}, \qquad \text{(A-18)}$$

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

Case 1 $(a = H)$: The LP is estimated over the fixed sample $t = 1, \ldots, T - H$.
We shall proceed by induction. For $h = 1$, we have by construction that $\widehat{\boldsymbol{B}}_{1,(-H)}^{\text{BB}} = \widehat{\boldsymbol{A}}_{(-H)}$. Assume then as the strong inductive hypothesis for $h - 1$ that $\widehat{\boldsymbol{B}}_{h-j,(-H)}^{\text{BB}} = \widehat{\boldsymbol{A}}_{(-H)}^{h-j}$ holds for all $j = 2, \ldots, h-2$ and $h > 1$. Substituting this hypothesis into eq.(A-18) gives for h that:

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

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

Case 2 $(a = h)$: The LP is estimated over the longest possible sample $t = 1, \ldots, T - h$.
In this case, $\boldsymbol{A}$ is estimated using the full sample $t = 1, \ldots, T - 1$, so that $\widehat{\boldsymbol{A}}_{(-a)}^{j} = \widehat{\boldsymbol{A}}_{(-1)}^{j}$. By our

adopted convention, $\widehat{\boldsymbol{A}}^{j}_{(-1)} = \widehat{\boldsymbol{A}}^{j}$, the $(-a)$ subscripts can be omitted from eq.(A-18). For $h = 1$, we have by construction that $\widehat{\boldsymbol{B}}^{\text{BB}}_{1} = \widehat{\boldsymbol{A}}$. For $h = 2$, substituting into eq.(A-18), we obtain:

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

with the simplification in the last step following from $\widehat{\boldsymbol{A}}_{(-2)} = \widehat{\boldsymbol{A}} + O_{p}(T^{-1})$ by eq.(A-4).

Now assume for $h - 1$ and $h \geq 3$ that $\widehat{\boldsymbol{B}}^{\text{BB}}_{h-j} = \widehat{\boldsymbol{A}}^{h-j} + O_{p}(T^{-1})$ holds for all $j = 2, \ldots, h-2$, and substitute the hypothesis into eq.(A-18). Using eq.(A-4), we thus obtain for h that:

$$\widehat{\boldsymbol{B}}^{\text{BB}}_{h} = \widehat{\boldsymbol{A}}^{h-1} \widehat{\boldsymbol{A}}_{(-h)} + O_{p}(T^{-1}) = \widehat{\boldsymbol{A}}^{h-1} \left(\widehat{\boldsymbol{A}} + O_{p}(T^{-1}) \right) = \widehat{\boldsymbol{A}}^{h} + O_{p}(T^{-1}).$$

Hence, by mathematical induction, the result $\widehat{\boldsymbol{B}}^{\text{BB}}_{h} = \widehat{\boldsymbol{A}}^{h} + O_{p}(T^{-1})$ holds for all $h = 2, \ldots, H$. □

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,*]

[a]VU Amsterdam, Department of Econometrics and Data Science
[b]Tinbergen Institute
[c]Ghent 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 presents additional simulation results. In Section S2.1, we report results for the local misspecification setting with data-driven lag selection, complementing Section 4.1.2 of the main text. Section S2.2 reports additional simulation results related to the design based on the Stock and Watson (2016) dynamic factor model introduced in Section 4.3 of the main text.

Supplement 1 Proof of Proposition 3 from the Main Paper

The derivation of the asymptotic distributions for the LP OLS and VAR IR estimators follows arguments similar to those in Li et al. (2022), which analyzes the same locally misspecified framework up to timing and normalization.

For convenience, we restate the DGP from eq.(13):

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

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 $\text{Var}(\boldsymbol{\mu}_t) = \text{diag}(\sigma_1^2, \sigma_2^2)$ and finite fourth

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

moments. 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 $j = 1, 2$.

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

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

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

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

with $\sigma_w^2 = (\beta^2 \sigma_1^2 + \sigma_2^2)/(1 - \rho^2)$. This result is unaffected by the local misspecification term $\alpha/\sqrt{T}$, 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.(13), 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:

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

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

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

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

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

$$\begin{aligned}
\widehat{\mathbf{A}} &= \mathbf{A}_0 + \left(\frac{1}{T}\sum_{t=1}^{T-1}\left(\begin{bmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} + \frac{\alpha}{\sqrt{T}}\mu_{2,t} \end{bmatrix}\begin{bmatrix} \mu_{1,t} & w_t \end{bmatrix}\right)\right)\widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \mathbf{A}_0 + \left(\frac{1}{T}\sum_{t=1}^{T-1}\boldsymbol{\mu}_{t+1}\mathbf{y}_t' + \frac{\alpha\sigma_2^2}{\sqrt{T}}\mathbf{e}_2\mathbf{e}_2' + O_p(T^{-1})\right)\widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \mathbf{A}_0 + \left(\frac{1}{T}\sum_{t=1}^{T-1}\boldsymbol{\mu}_{t+1}\mathbf{y}_t' + \frac{\alpha\sigma_2^2}{\sqrt{T}}\mathbf{e}_2\mathbf{e}_2'\right)\widehat{\boldsymbol{\Gamma}}^{-1} + O_p(T^{-1}), \qquad \text{(S1.6)} \\
&= \mathbf{A}_0 + O_p(T^{-1/2}). \qquad \text{(S1.7)}
\end{aligned}$$

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

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

$$\sqrt{T}\left(\widehat{\mathbf{A}} - \mathbf{A}_0\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\boldsymbol{\mu}_{t+1}\mathbf{y}_t'\boldsymbol{\Gamma}^{-1} + \alpha\sigma_2^2\mathbf{e}_2\mathbf{e}_2'\boldsymbol{\Gamma}^{-1} + o_p(1), \tag{S1.8}$$

which given $\mathbb{E}(\boldsymbol{\mu}_{t+1}\mathbf{y}_t') = \mathbf{0}$ 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 difference central limit theorem to the first term in Eq.(S1.8) yields:

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

where

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

from $\mathbb{E}\left[\boldsymbol{\mu}_{t+1}\mathbf{y}_t'\right] = \mathbf{0}$.

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

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

where the expectation follows from independence:

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

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

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

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 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:

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

where we use $\mathbf{B}_{0,h-1}\mathbf{e}_2\mathbf{e}_2' = \begin{pmatrix} 0 & 0 \\ \rho^{h-2}\beta & \rho^{h-1} \end{pmatrix}\mathbf{e}_2\mathbf{e}_2' = \rho^{h-1}\mathbf{e}_2\mathbf{e}_2'$ 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 $\mathbf{B}_{0,h}$, we have from eq. (S1.12)

$$\sqrt{T}\big(\widehat{\theta}_h^{\mathrm{LP}} - \theta_h\big) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} S_{t,h} + o_p(1), \qquad S_{t,h} = \sigma_1^{-2}\mu_{1,t}\sum_{j=0}^{h-1}\mathbf{e}_2'\mathbf{B}_{0,j}\,\boldsymbol{\mu}_{t+h-j}, \qquad \text{(S1.13)}$$

where use is made of $\mathbf{y}_t'\mathbf{\Gamma}^{-1}\mathbf{e}_1 = \sigma_1^{-2}\mu_{1,t}$ and the deterministic term $\rho^{h-1}\alpha\sigma_2^2\,\mathbf{e}_2\mathbf{e}_2'\mathbf{\Gamma}^{-1}$ vanishes after left–right selection because $\mathbf{e}_2'\mathbf{\Gamma}^{-1}\mathbf{e}_1 = 0$.

Since $\{\boldsymbol{\mu}_t\}$ are i.i.d. with independent components, $\mathbb{E}[\boldsymbol{\mu}_{t+s}\,\mu_{1,t}] = \mathbf{0}$ for all $s > 0$, and we have that $\mathbb{E}(S_{t,h}) = 0$. In addition, since each $S_{t,h}$ depends on $\mu_{1,t}$ and the future innovation block $(\boldsymbol{\mu}_{t+1}, \ldots, \boldsymbol{\mu}_{t+h})$, the sequence is h-dependent. Therefore, with $\mathbb{E}(S_{t,h}) = 0$, fixed h and finite fourth moments for $\{\boldsymbol{\mu}_t\}$, the CLT for m-dependent sequences (e.g., Billingsley, 1995, Thm. 27.4) applies, yielding

$$\sqrt{T}\left(\widehat{\theta}_h^{\mathrm{LP}} - \theta_h\right) \xrightarrow{d} \mathcal{N}\left(0, V_h^{\mathrm{LP}}\right). \qquad \text{(S1.14)}$$

such that $b_h^{\mathrm{LP}} = 0$, and with

$$V_h^{\mathrm{LP}} = \mathbb{E}[S_{t,h}^2] = \sigma_1^{-4}\mathbb{E}\left[\left(\sum_{j=0}^{h-1}\mathbf{e}_2'\mathbf{B}_{0,j}\boldsymbol{\mu}_{t+h-j}\mu_{1,t}\right)^2\right],$$

$$
\begin{aligned}
&= \sigma_1^{-4}\mathbb{E}\left[\sum_{j=1}^{h-1}\rho^{j-1}\beta\mu_{1,t+h-j}\mu_{1,t} + \sum_{j=0}^{h-1}\rho^{j}\mu_{2,t+h-j}\mu_{1,t}\right]^2, \\
&= \sigma_1^{-2}\left(\beta^2\sigma_1^2\sum_{j=1}^{h-1}\rho^{2(j-1)} + \sigma_2^2\sum_{j=0}^{h-1}\rho^{2j}\right), \\
&= \sigma_1^{-2}\left(\left(\beta^2\sigma_1^2+\sigma_2^2\right)\frac{1-\rho^{2h}}{1-\rho^2} - \rho^{2(h-1)}\beta^2\sigma_1^2\right), \\
&= \left(1-\rho^{2h}\right)\frac{\sigma_w^2}{\sigma_1^2} - \rho^{2(h-1)}\beta^2,
\end{aligned}
$$

where the expectation follows from independence and $\mathbb{E}[\boldsymbol{\mu}_{t+s}\mu_{1,t}] = \mathbf{0}$ for all $s > 0$. The result uses the fact that the variance of the sum $(1/\sqrt{T})\sum_{t=1}^{T-h} S_{t,h}$ contains no cross–time covariance contributions. Specifically, each score satisfies $S_{t,h} = \mu_{1,t}G_{t,h}$, with $G_{t,h}$ depending only on future innovations; since $\mu_{1,t}$ is independent of $(\mu_{1,s}, \mu_{2,s})_{s\neq t}$ and has mean zero, we have $\mathbb{E}[S_{t,h}S_{t+\ell,h}] = 0$ for all $\ell \geq 1$. Thus the score sequence is serially uncorrelated.

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:

$$
\begin{aligned}
\widehat{\boldsymbol{B}}_h^{\text{Lu}} &= \left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\mathbf{y}_{t+h} - \sum_{j=1}^{h-1}\widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}}\widehat{\boldsymbol{\varepsilon}}_{t+j}\right)\mathbf{y}_t'\right)\widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\mathbf{B}_h\mathbf{y}_t + \sum_{j=1}^{h}\boldsymbol{B}_{h-j}\boldsymbol{\varepsilon}_{t+j} - \sum_{j=1}^{h-1}\widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}}\widehat{\boldsymbol{\varepsilon}}_{t+j}\right)\mathbf{y}_t'\right)\widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \mathbf{B}_{0,h} + \frac{1}{T}\sum_{t=1}^{T-h}\left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1}\left(\boldsymbol{B}_{h-j}\boldsymbol{\varepsilon}_{t+j} - \widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}}\widehat{\boldsymbol{\varepsilon}}_{t+j}\right)\right)\mathbf{y}_t'\widehat{\boldsymbol{\Gamma}}^{-1}, \\
&= \boldsymbol{B}_{0,h} + \frac{1}{T}\sum_{t=1}^{T-h}\left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1}\left(\left(\boldsymbol{B}_{h-j} - \widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}}\right)\boldsymbol{\varepsilon}_{t+j} + \widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}}(\widehat{\mathbf{A}} - \mathbf{A})\mathbf{y}_{t+j-1}\right)\right)\mathbf{y}_t'\widehat{\boldsymbol{\Gamma}}^{-1},
\end{aligned}
\tag{S1.15}
$$

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

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\boldsymbol{\varepsilon}_{t+j}\mathbf{y}_t' &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\begin{bmatrix}\mu_{1,t+j}\\ \mu_{2,t+j}+\frac{\alpha}{\sqrt{T}}\mu_{2,t+j-1}\end{bmatrix}\begin{bmatrix}\mu_{1,t} & w_t\end{bmatrix},\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\boldsymbol{\mu}_{t+j}\mathbf{y}_t' + \mathbf{1}_{\{j=1\}}\alpha\sigma_2^2\mathbf{e}_2\mathbf{e}_2' + O_p(T^{-1/2}) = O_p(1). \qquad \text{(S1.16)}\end{aligned}$$

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

$$\widehat{\boldsymbol{B}}_{h-j}^{\text{Lu}} = \boldsymbol{B}_{h-j} + O_p(T^{-1/2}), \qquad \text{(S1.17)}$$

and substituting this in eq. (S1.15), together with eq.(S1.16), we obtain:

$$\widehat{\boldsymbol{B}}_h^{\text{Lu}} = \boldsymbol{B}_{0,h} + O_p(T^{-1/2}). \qquad \text{(S1.18)}$$

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.(S1.15)

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

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

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

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 $\widehat{\boldsymbol{B}}_{h-j}^{\mathrm{Lu}}-\boldsymbol{B}_{h-j}=O_p(T^{-1/2})$ for any $1\leq j\leq h-1$ by eqs.(S1.16) and (S1.17).

Third,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\sum_{j=1}^{h-1}\mathbf{e}_2'\widehat{\boldsymbol{B}}_{h-j}^{\mathrm{Lu}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{y}_{t+j-1}\mathbf{y}_t'\widehat{\boldsymbol{\Gamma}}^{-1}\mathbf{e}_1&=\sum_{j=1}^{h-1}\mathbf{e}_2'\widehat{\boldsymbol{B}}_{h-j}^{\mathrm{Lu}}\sqrt{T}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\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{\boldsymbol{B}}_{h-j}^{\mathrm{Lu}}\sqrt{T}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\boldsymbol{B}_{j-1}\mathbf{y}_t+\sum_{l=1}^{j-1}\boldsymbol{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{\boldsymbol{B}}_{h-j}^{\mathrm{Lu}}\sqrt{T}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\boldsymbol{B}_{j-1}\mathbf{y}_t\mathbf{y}_t'\widehat{\boldsymbol{\Gamma}}^{-1}\mathbf{e}_1+\sum_{l=1}^{j-1}\boldsymbol{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{\boldsymbol{B}}_{h-j}^{\mathrm{Lu}}\sqrt{T}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\left(\boldsymbol{B}_{j-1}\mathbf{e}_1+O_p(T^{-1/2})\right)\right),\\
&=\sum_{j=1}^{h-1}\mathbf{e}_2'\boldsymbol{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)\boldsymbol{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'\boldsymbol{B}_{h-j}\boldsymbol{\mu}_{t+1}\mathbf{y}_t'\boldsymbol{\Gamma}^{-1}\boldsymbol{B}_{j-1}\mathbf{e}_1+\alpha\sigma_2^2\sum_{j=2}^{h-1}\mathbf{e}_2'\boldsymbol{B}_{h-j}\mathbf{e}_2\mathbf{e}_2'\boldsymbol{\Gamma}^{-1}\boldsymbol{B}_{j-1}\mathbf{e}_1+o_p(1),\quad\text{(S1.19)}
\end{aligned}$$

where use is made of eqs.(S1.2) and (S1.8), and eq.(S1.16) on the 4th equality.

The first term in eq.(S1.19) is given by

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

$$= \mathbf{e}_2' \boldsymbol{B}_{h-1} \frac{1}{\sigma_1^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} \mu_{1,t} + (h-2) \frac{\rho^{h-3}\beta}{\sigma_w^2} \mathbf{e}_2' \boldsymbol{A}_0 \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \boldsymbol{\mu}_{t+1} w_t,$$

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

The second term in eq.(S1.19) is given by

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

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

Collecting and expanding terms yields

$$\begin{aligned}
\sqrt{T}\left(\widehat{\theta}_h^{\mathrm{Lu}} - \theta_h\right) &= (h-2)\rho^{h-2}\beta\frac{\alpha\sigma_2^2}{\sigma_w^2} + \frac{1}{\sigma_1^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+h}\mu_{1,t} + \mathbf{e}_2'\boldsymbol{B}_{h-1}\frac{1}{\sigma_1^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\boldsymbol{\mu}_{t+1}\mu_{1,t} \\
&\quad + (h-2)\frac{\rho^{h-3}\beta}{\sigma_w^2}\mathbf{e}_2'\boldsymbol{A}_0\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\boldsymbol{\mu}_{t+1}w_t + O_p(T^{-1/2}), \\
&= (h-2)\rho^{h-2}\beta\frac{\alpha\sigma_2^2}{\sigma_w^2} + \frac{1}{\sigma_1^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+h}\mu_{1,t} + \frac{\rho^{h-2}\beta}{\sigma_1^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{1,t+1}\mu_{1,t} \\
&\quad + \frac{\rho^{h-1}}{\sigma_1^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+1}\mu_{1,t} + (h-2)\frac{\rho^{h-3}\beta^2}{\sigma_w^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{1,t+1}w_t \\
&\quad + (h-2)\frac{\rho^{h-2}\beta}{\sigma_w^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mu_{2,t+1}w_t + o_p(1). \qquad \text{(S1.20)}
\end{aligned}$$

Applying a standard martingale central limit theorem to Eq.(S1.20) yields:

$$\sqrt{T}\left(\widehat{\theta}_h^{\mathrm{Lu}} - \theta_h\right) \xrightarrow{d} \mathcal{N}\left(b_h^{\mathrm{Lu}}, V_h^{\mathrm{Lu}}\right). \qquad \text{(S1.21)}$$

Using independence, the asymptotic bias is given by

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

The asymptotic variance is given by

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

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

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

□

S1.2.4 Proof of Bias and Variance Rankings

Proof of the bias ranking. For $h > 2$ and $\alpha, \beta, \rho \neq 0$,

$$
b_h^{\text{LP}} = 0, \qquad |b_h^{\text{VAR}}| = (h-1)|\rho|^{h-2}|\beta|\frac{|\alpha|\sigma_2^2}{\sigma_w^2}, \qquad |b_h^{\text{Lu}}| = (h-2)|\rho|^{h-2}|\beta|\frac{|\alpha|\sigma_2^2}{\sigma_w^2},
$$

which are strictly positive; since $(h-1) > (h-2)$, we obtain $|b_h^{\text{VAR}}| > |b_h^{\text{Lu}}| > |b_h^{\text{LP}}| = 0$. If any of α, β, ρ equals zero, then directly from the formulas $b_h^{\text{VAR}} = b_h^{\text{Lu}} = b_h^{\text{LP}} = 0$.

Proof of the variance ranking.

Step 1: Proof that $V_h^{\text{Lu}} > V_h^{\text{VAR}}$.

Subtracting the closed forms yields

$$V_h^{\text{Lu}} - V_h^{\text{VAR}} = \underbrace{\frac{\sigma_2^2}{\sigma_1^2}}_{>0} + \underbrace{\left(1 - \tfrac{\sigma_2^2}{\sigma_w^2}\right)\rho^{2(h-2)}\beta^2}_{\geq 0} + \underbrace{(h-2)^2\rho^{2(h-3)}\beta^4\frac{\sigma_1^2}{\sigma_w^2}}_{\geq 0} > 0,$$

where $\sigma_w^2 \geq \sigma_2^2$ implies $1 - \sigma_2^2/\sigma_w^2 \geq 0$ and the remaining factors are nonnegative (even powers and squares). Therefore $V_h^{\text{Lu}} > V_h^{\text{VAR}}$.

Step 2: Proof that $V_h^{\text{LP}} > V_h^{\text{Lu}}$ for $(\rho, \beta) \neq (0, 0)$.

First rewrite V_h^{LP} using $\frac{\sigma_w^2}{\sigma_1^2} = \frac{\beta^2}{1-\rho^2} + \frac{\sigma_2^2}{(1-\rho^2)\sigma_1^2}$ and the geometric sum:

$$V_h^{\text{LP}} = (1-\rho^{2h})\frac{\sigma_w^2}{\sigma_1^2} - \rho^{2(h-1)}\beta^2 = \frac{\sigma_2^2}{\sigma_1^2}\sum_{j=0}^{h-1}\rho^{2j} + \beta^2\sum_{j=0}^{h-2}\rho^{2j}.$$

Next, rewrite V_h^{Lu} using $\beta^2\frac{\sigma_1^2}{\sigma_w^2} = (1-\rho^2) - \frac{\sigma_2^2}{\sigma_w^2}$ and combining terms:

$$V_h^{\text{Lu}} = \left(1+\rho^{2(h-1)}\right)\tfrac{\sigma_2^2}{\sigma_1^2} + \rho^{2(h-2)}\beta^2\left(1+2(h-2)\tfrac{\sigma_2^2}{\sigma_w^2}\right) + (h-2)^2\rho^{2(h-3)}(1-\rho^2)\beta^2\left(1-\tfrac{\sigma_2^2}{\sigma_w^2}\right).$$

Subtracting and grouping terms,

$$V_h^{\text{LP}} - V_h^{\text{Lu}} = \Delta_{1,h} + \Delta_{2,h},$$

with

$$\Delta_{1,h} = \frac{\sigma_2^2}{\sigma_1^2}\left(\sum_{j=1}^{h-2}\rho^{2j} - (h-2)\rho^{2(h-2)}\frac{\beta^2\sigma_1^2}{\sigma_w^2}\right), \qquad \Delta_{2,h} = \beta^2\left(\sum_{j=0}^{h-3}\rho^{2j} - (h-2)\rho^{2(h-3)}g(s)\right),$$

where $s = \sigma_2^2/\sigma_w^2$ and

$$g(s) = (h-2)(1-\rho^2) + \big((h-1)\rho^2 - (h-2)\big)s.$$

For $\Delta_{1,h}$, since $\frac{\beta^2\sigma_1^2}{\sigma_w^2} \in [0,1)$ and ρ^{2j} is strictly decreasing for $0 < |\rho| < 1$, we have

$$\Delta_{1,h} \;\geq\; \frac{\sigma_2^2}{\sigma_1^2}\Big(\sum_{j=1}^{h-2}\rho^{2j} - (h-2)\rho^{2(h-2)}\Big) \;>\; 0 \quad \text{for } 0 < |\rho| < 1,$$

while if $\rho = 0$ then $\Delta_{1,h} = 0$.
For $\Delta_{2,h}$, since $s \in (0, 1-\rho^2]$ and g is affine with slope $(h-1)\rho^2 - (h-2)$, we may bound $g(s)$ by its maximum on the closed interval $[0, 1-\rho^2]$:

$$\max_{s\in[0,1-\rho^2]} g(s) = \begin{cases} g(0) = (h-2)(1-\rho^2), & \rho^2 \leq \dfrac{h-2}{h-1}, \\ g(1-\rho^2) = (h-1)\rho^2(1-\rho^2), & \rho^2 \geq \dfrac{h-2}{h-1}. \end{cases}$$

Thus $g(s) \leq \max_{s\in[0,1-\rho^2]} g(s)$ gives the lower bounds

$$\Delta_{2,h} \;\geq\; \begin{cases} \beta^2\Big(\sum_{j=0}^{h-3}\rho^{2j} - (h-2)^2(1-\rho^2)\rho^{2(h-3)}\Big), & \rho^2 \leq \frac{h-2}{h-1}, \\ \beta^2\Big(\sum_{j=0}^{h-3}\rho^{2j} - (h-2)(h-1)(1-\rho^2)\rho^{2(h-2)}\Big), & \rho^2 \geq \frac{h-2}{h-1}. \end{cases}$$

In the second case, $(1-\rho^2) \leq \frac{1}{h-1}$, so

$$\sum_{j=0}^{h-3}\rho^{2j} - (h-2)(h-1)(1-\rho^2)\rho^{2(h-2)} \;\geq\; \sum_{j=0}^{h-3}\rho^{2j} - (h-2)\rho^{2(h-2)} \;>\; 0,$$

since ρ^{2j} decreases in j. In the first case, writing $r = (1-\rho^2)/\rho^2 > 0$ and $n = h-2$,

$$\sum_{j=0}^{h-3}\rho^{2j} = \frac{1-\rho^{2n}}{1-\rho^2} > \; n^2(1-\rho^2)\rho^{2(n-1)} \iff \big((1+r)^n - 1\big)(1+r) - n^2r^2 > 0,$$

which holds because by the binomial lower bound

$$(1+r)^{n+1} - (1+r) - n^2r^2 \;\geq\; nr\Big(1 - \tfrac{n-1}{2}r + \tfrac{(n-1)(n+1)}{6}r^2\Big),$$

and the bracket is a convex quadratic with negative discriminant (hence no real roots) and value 1 at $r = 0$, so it is > 0 for all $r > 0$. Thus $\Delta_{2,h} > 0$ for $\beta \neq 0$.
Combining, $\Delta_{1,h} \geq 0$ (strict if $\rho \neq 0$) and $\Delta_{2,h} \geq 0$ (strict if $\beta \neq 0$) give $V_h^{\mathrm{LP}} - V_h^{\mathrm{Lu}} \geq 0$ for $h > 2$, with the equality only if $\rho = \beta = 0$. □

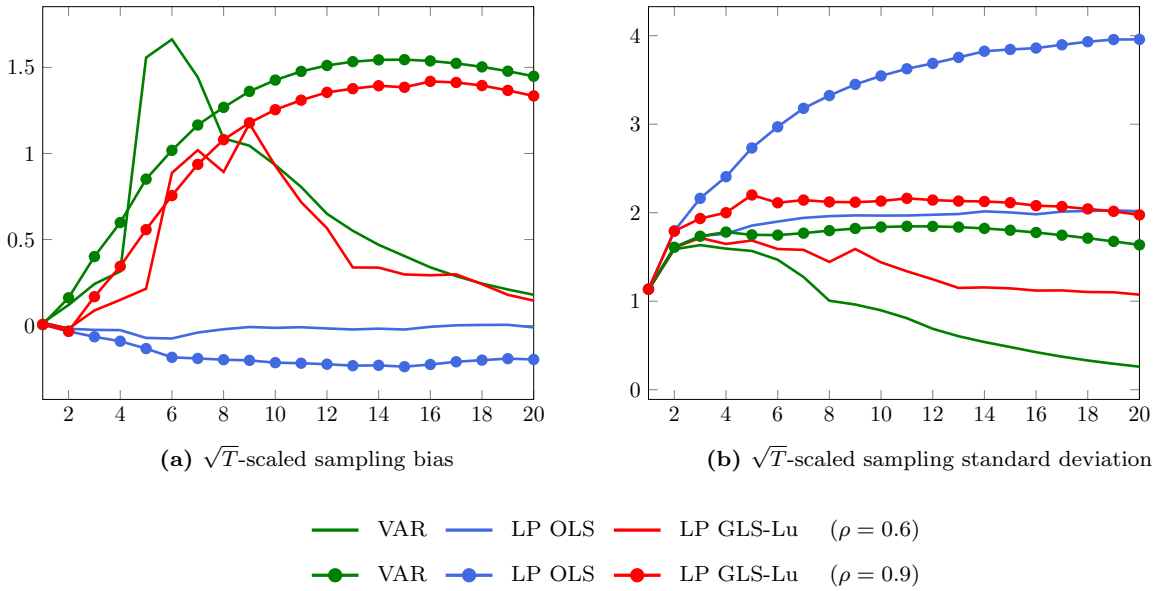
Supplement 2 Additional Simulation Results

S2.1 Simulation Results for Local Misspecification with Data-Driven Lag Selection

This section reports the simulation results that correspond to 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 $h = 1, \ldots, 20$ for a range of bias-variance weights $\lambda \in [0, 1]$.

AIC Lag Selection

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

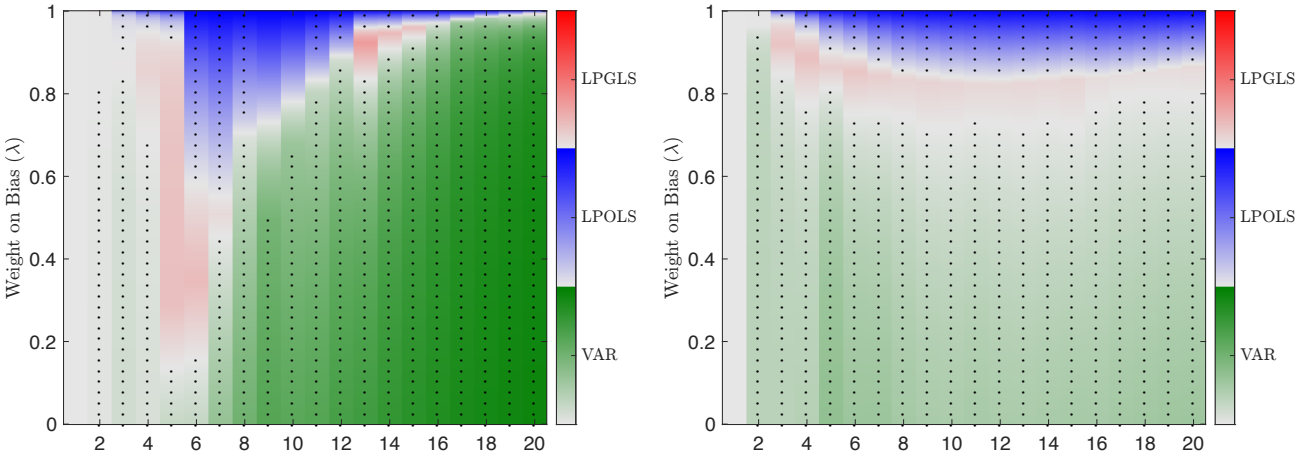


(a) $\sqrt{T}$-scaled sampling bias

(b) $\sqrt{T}$-scaled sampling standard deviation

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.(15), 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 $h = 1, \ldots, 20$.

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

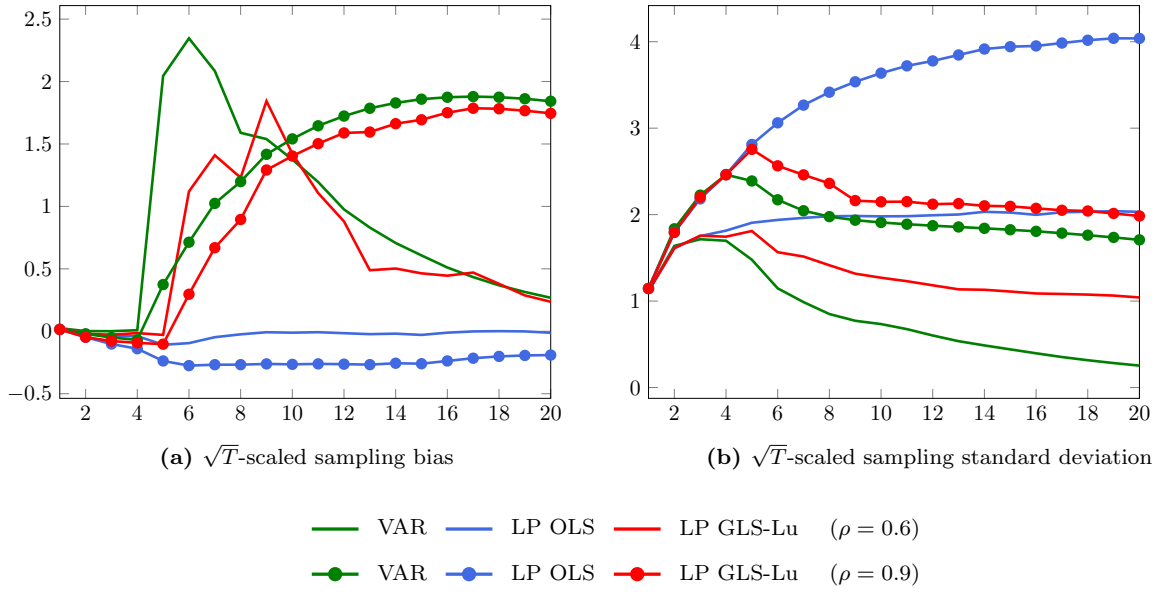


(a) $\rho = 0.6$ (b) $\rho = 0.9$

Notes: The heatmaps visualize estimator dominance across forecast horizons ($h = 1, \ldots, 20$), 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.(14), based on 10,000 Monte Carlo replications from the DGP in eq.(15) 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.

Lag Length Set to 4

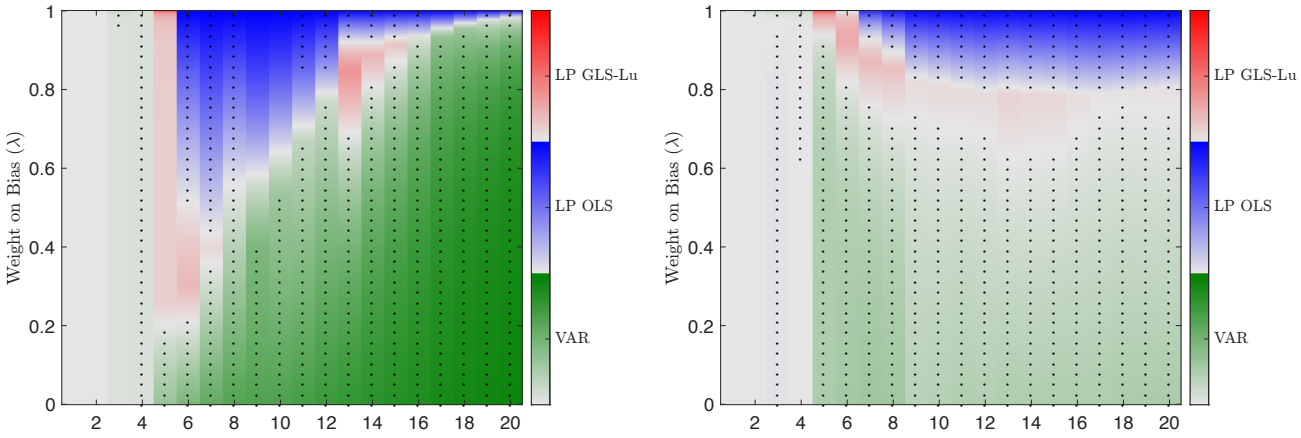
Figure S2.3: Bias and Standard Deviation — Local Misspecification with Lag Length Set to 4



(a) $\sqrt{T}$-scaled sampling bias **(b)** $\sqrt{T}$-scaled sampling standard deviation

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

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

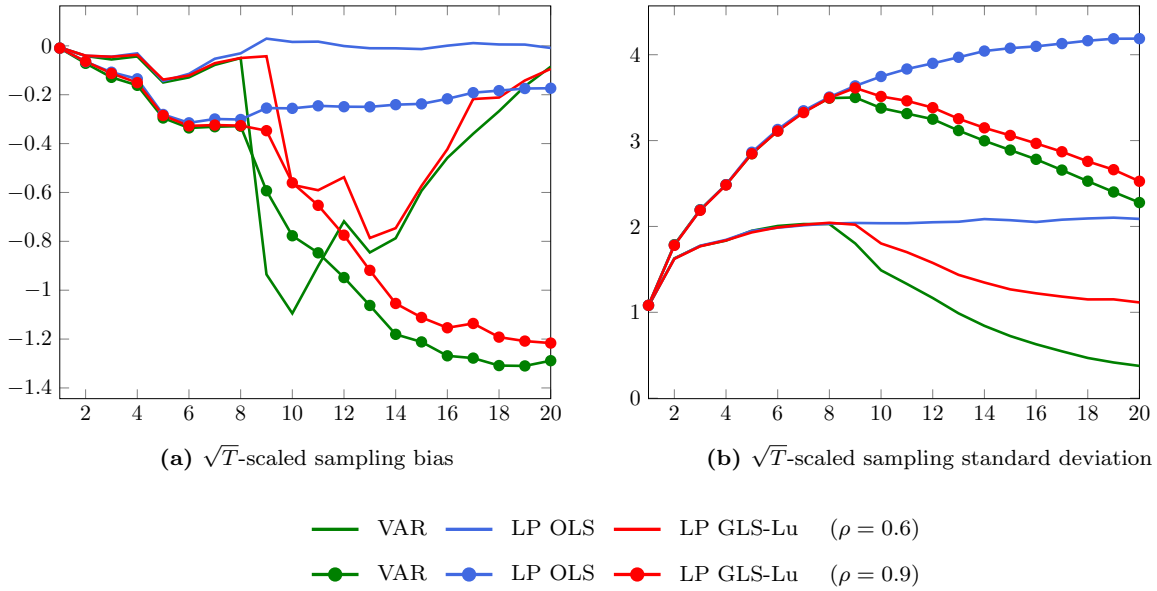


(a) $\rho = 0.6$ (b) $\rho = 0.9$

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

Lag Length Set to 8

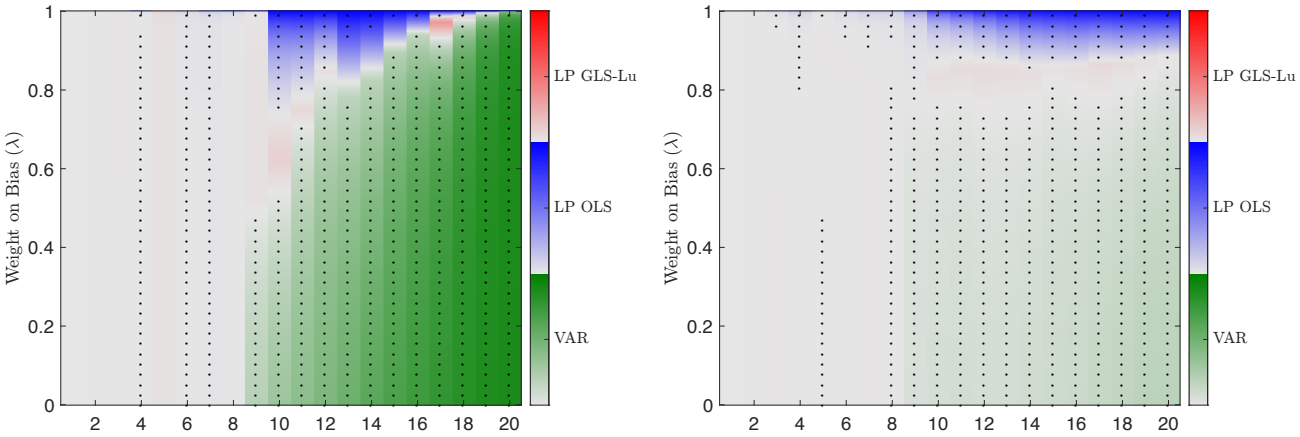
Figure S2.5: Bias and Standard Deviation — Local Misspecification with Lag Length Set to 8



(a) $\sqrt{T}$-scaled sampling bias **(b)** $\sqrt{T}$-scaled sampling standard deviation

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

Figure S2.6: Estimator Dominance by Weighted RMSE — Local Misspecification with Lag Length Set to 8



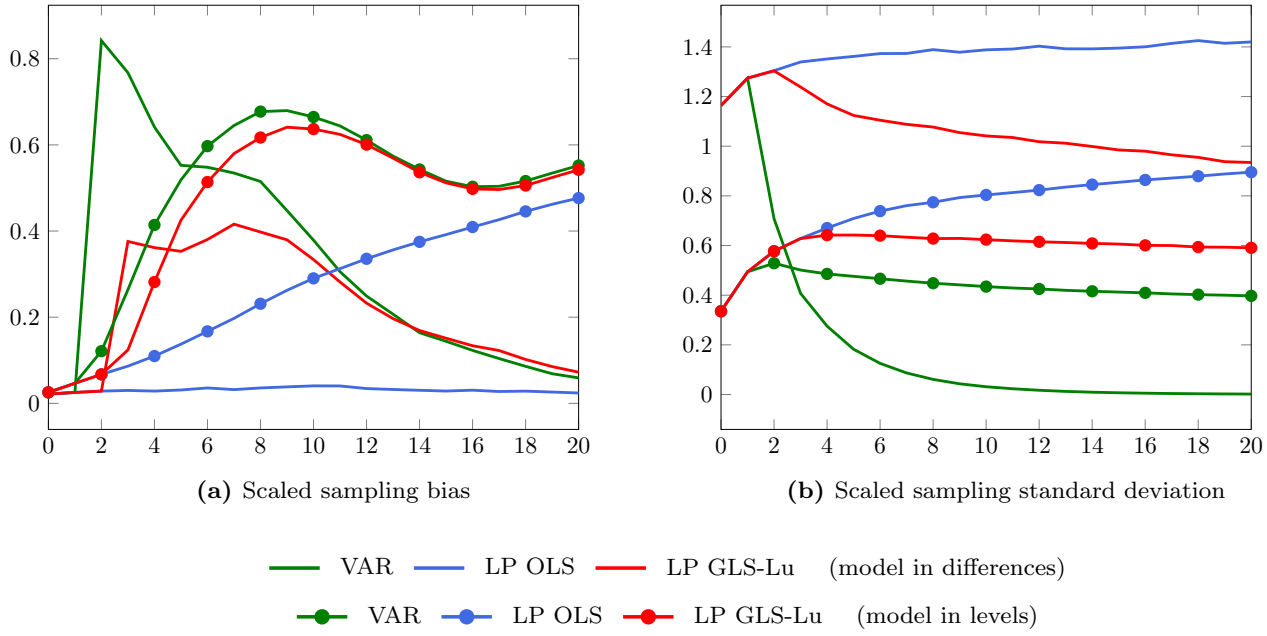
(a) $\rho = 0.6$ (b) $\rho = 0.9$

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

S2.2 Additional Simulations Based on Stock and Watson (2016) Dynamic Factor Model

This subsection complements Section 4.3 of the main text by reporting simulation results for the case where the VAR lag length is selected by AIC rather than fixed at four. Figure S2.7 displays the median scaled bias and scaled standard deviation across DGPs, while Figure S2.8 summarizes estimator dominance by weighted RMSE.

Figure S2.7: Scaled Bias and Standard Deviation — Stock-Watson DFM (AIC Lag Selection)

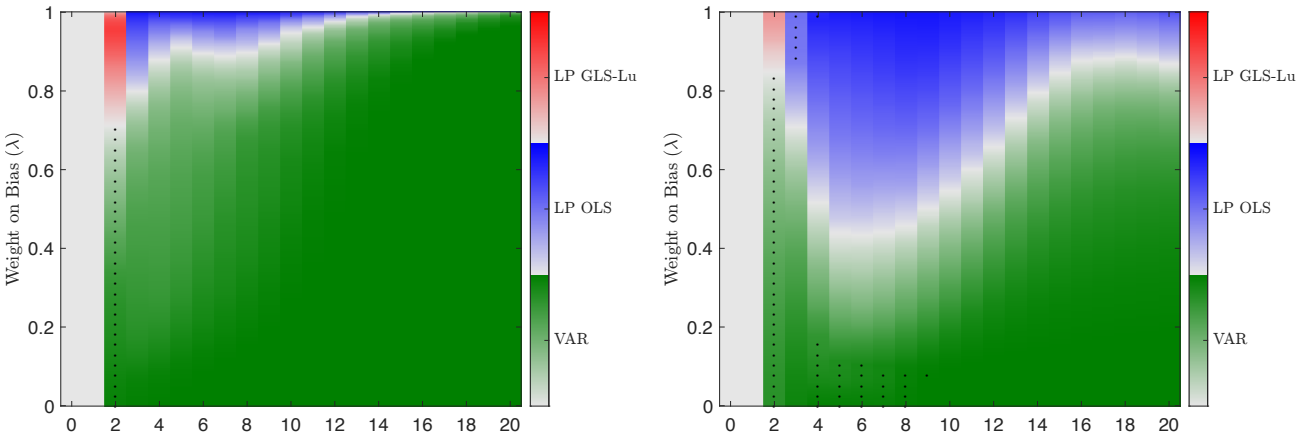


(a) Scaled sampling bias

(b) Scaled sampling standard deviation

Notes: Displayed are the medians (across 6,000 DGPs) of the absolute bias $|\mathbb{E}(\hat{\theta}_h) - \theta_h|$ and the standard deviation of $\hat{\theta}_h$ for the different estimation procedures, scaled by $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$, i.e. the root mean squared value of the true impulse responses $\{\theta_h\}_{h=0}^{20}$. For each DGP, bias and standard deviation are computed from 5,000 Monte Carlo replications based on data simulated from the Stock and Watson (2016) DFM, as implemented in Li et al. (2024), with $T = 200$. The VAR lag length is selected by AIC and applied uniformly across the VAR, LP OLS, and LP GLS–Lu estimators. The horizontal axis indicates the projection horizon $h = 0, \ldots, 20$.

Figure S2.8: Estimator Dominance by Weighted RMSE — Stock–Watson DFM (AIC Lag Selection)



(a) Model in differences

(b) Model in levels

Notes: Displayed are heatmaps of the estimator that attains the lowest weighted RMSE defined in eq.(14). For each (h, λ) combination, the "winner" is the estimator that most frequently minimizes the loss across the 6,000 DGPs simulated from the Stock and Watson (2016) DFM, following the design of Li et al. (2024). Color shading indicates the strength of dominance, with darker colors corresponding to higher frequencies and the darkest shade indicating that the estimator always wins. Dots mark cases where the LP GLS–Lu estimator is the second-best procedure. Results are based on 5,000 Monte Carlo replications per DGP with $T = 200$. The VAR lag length is selected by AIC and applied uniformly across the VAR, LP OLS, and LP GLS–Lu estimators. The horizontal axis indicates the projection horizon $h = 0, \ldots, 20$; the vertical axis varies the squared-bias weight $\lambda \in [0, 1]$.

References

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, New York, 3rd edition.

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.

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.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2A, pages 415–525. Elsevier, Amsterdam.