WORKING PAPER

GLS ESTIMATION OF LOCAL PROJECTIONS: TRADING ROBUSTNESS FOR EFFICIENCY

Ignace De Vos
Gerdie Everaert

October 2024
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

October 10, 2024

Abstract

Local projections (LPs) are often regarded as more robust to model misspecification than impulse responses (IRs) derived from forward-iterated dynamic model estimates, as LPs impose fewer restrictions on the underlying dynamics. However, because forecast errors accumulate in the LP errors over the projection horizon, this robustness comes at the price of an increase in variance. To address this, several Generalized Least Squares (GLS) estimators have been proposed to reduce error accumulation and enhance efficiency. We demonstrate, however, that the implied conditioning on dynamic model (horizon-one LP) residuals imposes strong restrictions on the underlying data generating process, undermining the very robustness to misspecification that LPs are valued for. In fact, we show that these GLS LP estimators tend to align more closely with forward-iterated IRs from potentially misspecified models, than with OLS-estimated LPs. Furthermore, we find that conditioning on previous horizon LP residuals fails to deliver efficiency improvements over OLS-estimated LPs.

JEL-codes: C22, C13, C53

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

1 Introduction

Since the influential work of Jordà (2005), local projections (LPs) have become a widely used tool for estimating impulse responses (IRs). Unlike traditional methods that rely on forward iteration of dynamic model estimates to compute IRs over extended horizons (as in the VAR approach), LPs directly estimate IR coefficients at each forecast horizon. This feature makes LPs particularly

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

robust to model misspecification, as they employ fewer restrictions on the underlying dynamics of the data. However, this robustness comes at the cost of lower estimation accuracy (see for instance Ramey, 2016; Li et al., 2024). This is not only because each IR coefficient is estimated separately, but also because forecast errors accumulate in the LP error terms over the projection horizon. Consequently, LP OLS residuals tend to be serially correlated, necessitating either lag augmentation (Montiel Olea and Plagborg-Møller, 2021) or the use of autocorrelation-robust standard errors to ensure valid inference.

In his seminal 2005 paper, Jordà proposed a potentially significant improvement to the efficiency of the LP OLS estimator by suggesting that residuals from previous-horizon projections could be recursively included as regressors in the horizon h projection. The full details of this enhancement were left for future exploration, which inspired Lusompa (2023) and Breitung and Brüggemann (2023) to introduce various Generalized Least Squares (GLS) methods. These approaches condition on the residuals from the dynamic model (i.e., the horizon-one LP) to mitigate the accumulation of forecast errors in the LP error terms. Given the clear potential for improving estimation accuracy, these methods are poised to be widely adopted in the econometrics community. Indeed, Lusompa's (2023) approach has already seen practical application in the recent work of Clark et al. (2024).

In this paper, we analyze the properties of the LP GLS estimators proposed by Lusompa (2023) and Breitung and Brüggemann (2023). We demonstrate that the efficiency gains achieved by these estimators come from incorporating more of the assumed model's structure into the estimation process, which undermines the robustness to misspecification that is central to the LP methodology. In fact, fully utilizing the information from the dynamic model residuals results in an LP GLS estimator that exactly replicates the iterated impulse responses, which are considered more sensitive to model misspecification. As a result, these LP GLS estimators should be viewed as an alternative method for imposing model-based restrictions, rather than as a means of improving LP efficiency without sacrificing robustness. Furthermore, we show that Jordà's (2005) suggestion to condition on residuals from previous LP horizons is asymptotically equivalent to the standard LP OLS estimator and thus offers no improvement.

The remainder of this note is structured as follows. Section 2 introduces the data-generating process and presents two standard methods for estimating impulse responses: forward iteration of the dynamic model estimates and LPs estimated via OLS. Section 2.2 explores how the various GLS-transformed LP estimates compare to both the iterated IRs and the LP OLS estimates. For expositional purposes, all estimators will be based on a simple AR(1) specification without deterministic terms, although the key points can easily be generalized to more complex models. Note that the considered estimators are all consistent when the errors of the AR(1) specification are i.i.d., but not when they are persistent. We will also allow for model misspecification with moving average (MA) errors, as in the absence of misspecification, the optimal approach would be to iteratively calculate IRs from the correctly specified dynamic model (or use an LP GLS estimator that replicates these IRs). While LPs are considered more robust to misspecification, this note does not aim to prove that point. Instead, we focus on how misspecification influences the position of the various LP GLS estimators relative to iterated IR and LP OLS. Section 3 illustrates the bias-variance trade-off between the various estimators using a simple example. Section 4 concludes. Proofs and simulation results are presented in the Appendix.

2 Estimating impulse responses in the dynamic model

Consider the following stylized univariate dynamic model

$$y_{t+1} = \alpha y_t + \varepsilon_{t+1}^{q}, \qquad \text{for} \qquad t = 1, \dots, T-1, \tag{1}$$

where $|\alpha| < 1$ and $\varepsilon_{t+1}^{q} = \beta(q)\mu_{t+1}$, with $\beta(q) = \sum_{j=0}^{q} \beta_j L^j$ a stable (invertible) lag polynomial of order $q \geq 0$, with $\beta_0 = 1$, and μ_t is an i.i.d. process with $\mathbb{E}(\mu_t) = 0$, $\mathbb{E}(\mu_t^2) = \sigma^2 > 0$ and $\mathbb{E}(\mu_t^4) < \infty$ for all t. Note that q controls the degree of autocorrelation in ε_{t+1}^{q}.

The aim is to estimate the response δ_h of y_{t+h} to a unit impulse in μ_t over the horizon $h = 1, \dots, H$ for a fixed and finite $H (< T)$. In the analysis below, δ_h will be estimated using various approaches that start from a simple AR(1) specification assuming i.i.d. errors, which is correct if $q = 0$ but misspecified if $q > 0$. In particular, if $q = 0$ then $\delta_h = \alpha^h$ and the IRs can be consistently estimated using AR(1)-based methods. If $q > 0$, then $\delta_h = \alpha^h + B_{q,h}$ so that the IRs will also depend on a function of the MA coefficients $B_{q,h} = f_h(\beta(q))$, and an AR(1) specification will in general lead to inconsistent IR estimators.[1] Thus, q provides a straightforward way to control the degree of misspecification in model (1), with $B_{0,h} = 0$ when $q = 0$.

2.1 Benchmark impulse response estimators

2.1.1 Forward-iterated estimates

Let $\widehat{\alpha} = (\sum_{t=1}^{T-1} y_t^2)^{-1} \sum_{t=1}^{T-1} y_t y_{t+1}$ denote the OLS estimate of α in eq.(1), obtained from regressing y_{t+1} on y_t. The forward-iterated estimate of δ_h is then given by $\widehat{\delta}_h^{iter} = \widehat{\alpha}^h$. This is essentially the VAR approach to estimating IRs.

If the model is correctly specified $(q = 0)$, the limiting distribution of $\widehat{\delta}_h^{iter}$, for finite h, can be derived from the limiting distribution of $\widehat{\alpha}$ with the delta method

$$\sqrt{T}(\widehat{\delta}_h^{iter} - \delta_h) \xrightarrow{d} \mathcal{N}(0, (1-\alpha^2)h^2\alpha^{2h-2}). \tag{2}$$

In this case, $\widehat{\delta}_h^{iter}$ is an optimal estimate for δ_h in an efficiency sense. If $q > 0$, then $\widehat{\delta}_h^{iter}$ is inconsistent for δ_h and the asymptotic distribution will differ from eq.(2).

[1]For instance, if $q = 1$ then $\delta_h = \alpha^h + \alpha^{h-1}\beta_1$.

2.1.2 Local Projections with OLS

Jordà (2005) proposes to directly estimate δ_h with the following local projections

$$y_{t+h} = b_h y_t + e^h_{t+h}, \qquad \text{for} \qquad h = 1, \dots, H, \tag{3}$$

where backward iteration of eq.(1) reveals that $b_h = \alpha^h$ and $e^h_{t+h} = \sum_{j=1}^h \alpha^{h-j} \varepsilon^q_{t+j}$, irrespective of whether the model is correctly specified.

Regressing y_{t+h} on y_t yields the corresponding LP OLS estimator:

$$\widehat{b}_h^{OLS} = \left(\sum_{t=1}^{T-h} y_t^2 \right)^{-1} \sum_{t=1}^{T-h} y_t y_{t+h}. \tag{4}$$

If $q = 0$, the AR(1) model is correctly specified and y_t is uncorrelated with $e^h_{t+h} = \sum_{j=1}^h \alpha^{h-j} \mu_{t+j}$, which implies that $\widehat{b}_h^{OLS}$ is consistent for $\delta_h = \alpha^h$. The limiting distribution in this case is given by (see also Bhansali, 1997; Lusompa, 2023)

$$\sqrt{T}(\widehat{b}_h^{OLS} - \delta_h) \xrightarrow{d} \mathcal{N}\left(0, (1-\alpha^2)^{-1}\left(1 + \alpha^2 - (2h+1)\alpha^{2h} + (2h-1)\alpha^{2h+2}\right)\right), \tag{5}$$

where we recall that h is treated as a fixed finite quantity. Since $|\alpha| < 1$, the asymptotic variance of $\widehat{b}_h^{OLS}$ in eq.(5) exceeds that of $\widehat{\delta}_h^{iter}$ in eq.(2), particularly for larger values of h as more forecast errors then accumulate in e^h_{t+h}. Thus, for non-zero h, LP OLS is less efficient than iterated IRs. To address this inefficiency, several GLS transformations have been proposed, which we will explore in the next section.

When the AR(1) model in eq.(1) is misspecified ($q > 0$), $\widehat{b}_h^{OLS}$ is in general not consistent for δ_h. However, following Galvao and Kato (2014), we note that $\widehat{b}_h^{OLS}$ then consistently estimates the pseudo-true IR parameter δ_h^p:

$$\delta_h^p = \alpha^h + \frac{\mathbb{C}ov\left(y_t, \sum_{j=1}^h \alpha^{h-j}\varepsilon^q_{t+j}\right)}{\mathbb{V}ar(y_t)} = \alpha^h + \sum_{j=1}^h \alpha^{h-j} c_j, \tag{6}$$

with $c_j = \mathbb{C}ov(y_t, \varepsilon^q_{t+j})/\mathbb{V}ar(y_t)$. This pseudo-true IR represents the best partial linear approximation to the true h-period ahead response δ_h, which underlies the commonly cited robustness of LPs to model misspecification and is a key advantage of LPs over forward-iterated estimates. Setting $b_h = \delta_h^p$ in eq.(3), the LP errors e^h_{t+h} are now given by $e^h_{t+h} = \sum_{j=1}^h \alpha^{h-j}(\varepsilon^q_{t+j} - c_j y_t)$, or equivalently following some more algebra:

$$e^h_{t+h} = \sum_{j=1}^h \delta^p_{h-j} \nu^j_{t+j}, \qquad \text{with} \quad \nu^j_{t+j} = \varepsilon^q_{t+j} - c_j y_t - \sum_{\ell=1}^{h-1} c_{h-\ell} \nu^\ell_{t+\ell}, \tag{7}$$

where $\delta_0^p = 1$. Note that the e^h_{t+h} are by construction uncorrelated with y_t for all h, and $\delta_h^p = \delta_h$ if

$q = 0$.

2.2 Local projections with GLS

The LP error term e^h_{t+h} in eq.(3) is a moving average of forecast errors, and this structure can be used to enhance efficiency through a GLS transformation. We will explore two alternative approaches: (i) using residuals from the dynamic model (horizon-one LP OLS) and (ii) using residuals from previous-horizon LP GLS estimates.

Approach suggested by Lusompa (2023)

Lusompa (2023) suggests to transform eq.(3) for $h \geq 2$ by bringing the forecast errors $\varepsilon^q_{t+1}, \ldots, \varepsilon^q_{t+h-1}$ to the left-hand side:

$$y_{t+h} - \sum_{j=1}^{h-1} b_{h-j}\varepsilon^q_{t+j} = b_h y_t + e^{h,Lu}_{t+h}. \tag{8}$$

When $q = 0$, we have that the remainder $e^{h,Lu}_{t+h} = \varepsilon^q_{t+h} = \mu_{t+h}$ is an i.i.d. process such that this approach indeed removes autocorrelation from the LP errors. When $q > 0$, however, the structure of the LP error e^h_{t+h} in eq.(7) shows that the GLS correction is not able to fully remove the $\nu^1_{t+1}, \ldots, \nu^{h-1}_{t+h-1}$ from the LP GLS errors $e^{h,Lu}_{t+h}$. Moreover, the suggested GLS correction will result in correlation between y_t and $e^{h,Lu}_{t+h}$, such that the resulting GLS estimator will also no longer be consistent for δ^p_h. We will elaborate on this below.

In practice, eq.(8) is implemented by substituting the unobserved ε^q_{t+j} on the left-hand side with their estimates $\widehat{\varepsilon}^q_{t+j} = y_{t+j} - \widehat{\alpha} y_{t+j-1}$ obtained from the assumed dynamic model (LP at horizon $h = 1$), and by using previous-horizon LP GLS estimates for the associated LP coefficients b_{h-j}. Thus, b_h from eq.(8) is estimated by iteratively transforming the data with the $h = 1$ residuals over the full forecast horizon:

$$\widehat{b}^{GLS,Lu}_h = \left(\sum_{t=1}^{T-h} y_t^2\right)^{-1} \sum_{t=1}^{T-h} y_t \left(y_{t+h} - \sum_{j=1}^{h-1} \widehat{b}^{GLS,Lu}_{h-j}\widehat{\varepsilon}^q_{t+j}\right), \qquad \text{for} \qquad h = 2, \ldots, H, \tag{9}$$

where $\widehat{b}^{GLS,Lu}_1 = \widehat{b}^{OLS}_1 = \widehat{\alpha}$.

If $q = 0$, the correct dynamic specification is an AR(1) model, so that the estimated $\widehat{\varepsilon}^q_{t+j}$ correspond asymptotically to the errors required for the transformation in (8), and $\widehat{b}^{GLS,Lu}_h$ is consistent for δ_h. Lusompa (2023) shows that the limiting distribution is then given by

$$\sqrt{T}(\widehat{b}^{GLS,Lu}_h - \delta_h) \xrightarrow{d} \mathcal{N}\left(0, (1-\alpha^2)\left(1 + (h^2 - 1)\alpha^{2h-2}\right)\right). \tag{10}$$

A comparison of the asymptotic variance expressions in eqs.(2), (5), and (10) reveals that, under correct specification, the LP GLS estimator is more efficient than LP OLS, though less efficient than the optimal iterated IRs. The efficiency gains of the LP GLS estimator are particularly notable when persistence (α) is high, as the accumulation of forecast errors in the LP error term becomes more pronounced, leading to significant improvements over LP OLS.

Under misspecification, the estimated $\widehat{\varepsilon}^{q}_{t+j}$ are not consistent for the errors in (7), meaning that (9) transforms the data for each h according to a model that does not reflect the true underlying dynamics. To assess the consequences for LP GLS in such cases, we compare it with two benchmark IR estimators: $\widehat{b}_h^{OLS}$, generally considered the most robust to misspecification, and the forward iterated IR $\widehat{\delta}_h^{iter}$, typically the least robust. In the Appendix, we demonstrate that the differences between these estimators are given by:

$$\widehat{b}_h^{GLS,Lu} - \widehat{b}_h^{OLS} = B_h^{Lu} + O_p(T^{-1/2}), \qquad B_h^{Lu} = -\Gamma^{-1} \sum_{j=1}^{h-1} \left(\delta^{p}_{h-j} + B^{Lu}_{h-j} \right) \phi_j, \tag{11}$$

$$\widehat{b}_h^{GLS,Lu} - \widehat{\delta}_h^{iter} = C_h^{Lu} + O_p(T^{-1/2}), \qquad C_h^{Lu} = \Gamma^{-1} \left(\phi_h - \sum_{j=1}^{h-1} C^{Lu}_{h-j} \phi_j \right), \tag{12}$$

as $T \to \infty$, where $\phi_j = \mathbb{E}(y_t \varepsilon^{q}_{t+j}) - c_1 \mathbb{E}(y_t y_{t+j-1})$ and $B_1^{Lu} = C_1^{Lu} = 0$. When eq.(1) is correctly specified, y_t is uncorrelated with ε^{q}_{t+j} for all j, leading to $B_h^{Lu} = C_h^{Lu} = 0$ such that all approaches are consistent for δ_h. When $q > 0$, however, y_t is correlated with ε^{q}_{t+j}, resulting in $B_h^{Lu} \neq 0$ and $C_h^{Lu} \neq 0$. As a result, $\widehat{b}_h^{GLS,Lu}$ will differ from both $\widehat{b}_h^{OLS}$ and $\widehat{\delta}_h^{iter}$ for $h \geq 2$, even asymptotically, indicating that the use of the $h = 1$ residuals induces a loss of some of the robustness to misspecification inherent in LP OLS. The extent of this deviation, and whether it aligns more closely with one estimator or the other, depends on both the underlying data-generating process and the projection horizon.

Approach suggested by Breitung and Brüggemann (2023)

Breitung and Brüggemann (2023) suggest an alternative GLS transformation that conditions on a different set of errors $\varepsilon^{q}_{t+2}, \ldots, \varepsilon^{q}_{t+h}$.[2] Bringing these errors to the left-hand side of eq.(3) yields

$$y_{t+h} - \sum_{j=2}^{h} b_{h-j} \varepsilon^{q}_{t+j} = b_h y_t + e^{h,BB}_{t+h}. \tag{13}$$

When $q = 0$, $e^{h,BB}_{t+h} = \alpha^{h-1} \varepsilon^{q}_{t+1} = \alpha^{h-1} \mu_{t+1}$ is an i.i.d. process, meaning that also this transformation prevents the accumulation of forecast errors and eliminates autocorrelation in the LP errors. However, similar to the approach by Lusompa (2023), this is no longer the case when the model is misspecified.

[2] Note that Breitung and Brüggemann (2023) suggest to transform eq.(3) by bringing ε^{q}_{t+h} to the left-hand side and include $\varepsilon^{q}_{t+2}, \ldots, \varepsilon^{q}_{t+h-1}$ as additional regressors. However, it is not needed to re-estimate the coefficients on these forecasting errors as they were already estimated in the previous LP horizons. For expositional purposes, we therefore also brought $\varepsilon^{q}_{t+2}, \ldots, \varepsilon^{q}_{t+h-1}$ to the left-hand side of eq.(3).

Substituting the unobserved ε^q_{t+j} with the $h = 1$ dynamic model estimates $\widehat{\varepsilon}^q_{t+j}$, and using previous horizon estimates for the associated LP coefficients b_{h-j}, the b_h from eq.(13) is iteratively estimated as

$$\widehat{b}_h^{GLS,BB} = \left(\sum_{t=1}^{T-h} y_t^2\right)^{-1} \sum_{t=1}^{T-h} y_t \left(y_{t+h} - \sum_{j=2}^{h} \widehat{b}_{h-j}^{GLS,BB} \widehat{\varepsilon}^q_{t+j}\right), \qquad \text{for} \qquad h = 2, \ldots, H, \tag{14}$$

where $\widehat{b}_1^{GLS,BB} = \widehat{\alpha}$ and $\widehat{b}_0^{GLS,BB} = 1$.

Since the $\widehat{\varepsilon}^q_{t+j}$ are again used to transform the data for all $h \geq 2$, this effectively extrapolates the model for $h = 1$ across the entire forecast horizon, similar to the approach behind $\widehat{\delta}_h^{iter}$. In fact, as we confirm in the Appendix, the statement by Breitung and Brüggemann (2023) that $\widehat{b}_h^{GLS,BB}$ only differs from $\widehat{\delta}_h^{iter}$ by an asymptotically negligible term holds true,

$$\widehat{b}_h^{GLS,BB} = \widehat{\delta}_h^{iter} + O_p(T^{-1}), \tag{15}$$

such that both estimators are equivalent as $T \to \infty$, regardless of whether the model is correctly specified. The rate of convergence is also sufficiently fast to see that $\widehat{b}_h^{GLS,BB}$ has the same asymptotic distribution as $\widehat{\delta}_h^{iter}$. Consequently, $\widehat{b}_h^{GLS,BB}$ has a lower variance than $\widehat{b}_h^{OLS}$, but its asymptotic equivalence to $\widehat{\delta}_h^{iter}$ indicates that this variance reduction is achieved by imposing the AR(1) specification over the entire forecast horizon. This, in turn, eliminates the robustness of LPs and reintroduces the well-known sensitivity of forward-iterated estimators like $\widehat{\delta}_h^{iter}$.

Full GLS transformation

The argument that the GLS transformation eliminates the robustness of the LP estimator can be pushed further by also removing the forecast errors ε^q_{t+h} and ε^q_{t+1}, which are still present in the LP errors of the GLS transformations proposed by Lusompa (2023) and Breitung and Brüggemann (2023), respectively. In fact, there is no compelling reason to retain these errors, as their estimates can also be readily obtained from the assumed dynamic model. Specifically, the full set of forecast errors $\varepsilon^q_{t+1}, \ldots, \varepsilon^q_{t+h}$ can be subtracted from the left-hand side of eq.(3) from $h = 2$ onward. Upon replacing unknown quantities by sample estimates, a $\widehat{b}_h^{GLS,full}$ estimate for b_h can be iteratively obtained from the following transformed model

$$y_{t+h} - \sum_{j=1}^{h} \widehat{b}_{h-j}^{GLS,full} \widehat{\varepsilon}^q_{t+j} = b_h y_t + e_{t+h}^{h,full}, \quad \text{for} \quad h \geq 2, \tag{16}$$

with $\widehat{b}_0^{GLS,full} = 1$ and $\widehat{b}_1^{GLS,full} = \widehat{\alpha}$. It is easy to show that $y_{t+h} - \sum_{j=1}^{h} \widehat{b}_{h-j}^{GLS,full} \widehat{\varepsilon}^q_{t+j} = \widehat{\alpha}^h y_t$, such that the LP GLS estimator $\widehat{b}_h^{GLS,full}$ in eq.(16) is given by

$$\widehat{b}_h^{GLS,full} = \left(\sum_{t=1}^{T-h} y_t^2\right)^{-1} \sum_{t=1}^{T-h} \left(y_{t+h} - \sum_{j=1}^{h} \widehat{b}_{h-j}^{GLS,full} \widehat{\varepsilon}^q_{t+j}\right) y_t = \widehat{\alpha}^h \frac{\sum_{t=1}^{T-h} y_t^2}{\sum_{t=1}^{T-h} y_t^2} = \widehat{\alpha}^h = \widehat{\delta}_h^{iter}. \tag{17}$$

Thus, a GLS transformation that fully utilizes the available $h = 1$ residuals is even numerically identical to $\widehat{\delta}_h^{iter}$, regardless of the sample size or whether the model is correctly specified. We have thus moved from a large sample equivalence as in eq.(15) to a finite sample identity $\widehat{b}_h^{GLS,full} \equiv \widehat{\delta}_h^{iter}$. This further confirms that the GLS transformation imposes the dynamic model on the full forecast horizon and that the robustness of LPs is lost in the process.

Iterative GLS LP with LP GLS residuals

We have argued that using estimated $h = 1$ residuals for GLS corrections imposes the assumed model's dynamics over the entire projection horizon, thereby eliminating the robustness of LPs to misspecification. This raises the question of whether Jordà's suggestion to use previous horizon LP GLS residuals for the correction can mitigate this issue. That is, Jordà (2005) suggests the following sequence of GLS transformations

$$y_{t+h} - \sum_{j=1}^{h-1} b_{h-j} v_{t+j}^{j} = b_h y_t + v_{t+h}^{h}, \tag{18}$$

with v_{t+h}^{h} as defined in eq.(7), and operationalized by replacing b_{h-j} and v_{t+j}^{j} (for $j = 1, \ldots, h-1$) by estimates obtained at the previous horizon. The suggested GLS estimator is therefore

$$\widehat{b}_h^{GLS,LPe} = \left(\sum_{t=1}^{T-h} y_t^2\right)^{-1} \sum_{t=1}^{T-h} y_t \left(y_{t+h} - \sum_{j=1}^{h-1} \widehat{b}_{h-j}^{GLS,LPe} \widehat{v}_{t+j}^{j}\right),$$

where $\widehat{b}_0^{GLS,LPe} = 1$, $\widehat{b}_1^{GLS,LPe} = \widehat{\alpha}$ and $\widehat{v}_{t+j}^{j} = y_{t+j} - \widehat{b}_j^{GLS,LPe} y_t$. To assess its behavior, we again consider the difference from $\widehat{b}_h^{OLS}$ as $T \to \infty$ (see the Appendix for proof):

$$\widehat{b}_h^{GLS,LPe} - \widehat{b}_h^{OLS} = -\sum_{j=1}^{h-1} \left(\widehat{b}_{h-j}^{GLS,LPe} \frac{\sum_{t=1}^{T-h} y_t \widehat{v}_{t+j}^{j}}{\sum_{t=1}^{T-h} y_t^2} \right) = O_p(T^{-1}). \tag{19}$$

The rate of convergence in eq.(19) implies that $\widehat{b}_h^{GLS,LPe}$ shares the same asymptotic distribution as $\widehat{b}_h^{OLS}$ for $T \to \infty$, meaning that the proposed GLS transformation does not harm robustness but also does not result in efficiency gains over OLS. Intuitively, this is because the transformation terms on the left-hand side of eq.(18) must be estimated, which increases the variance to the same level as that of the original LP OLS estimator.

3 Illustrative Example

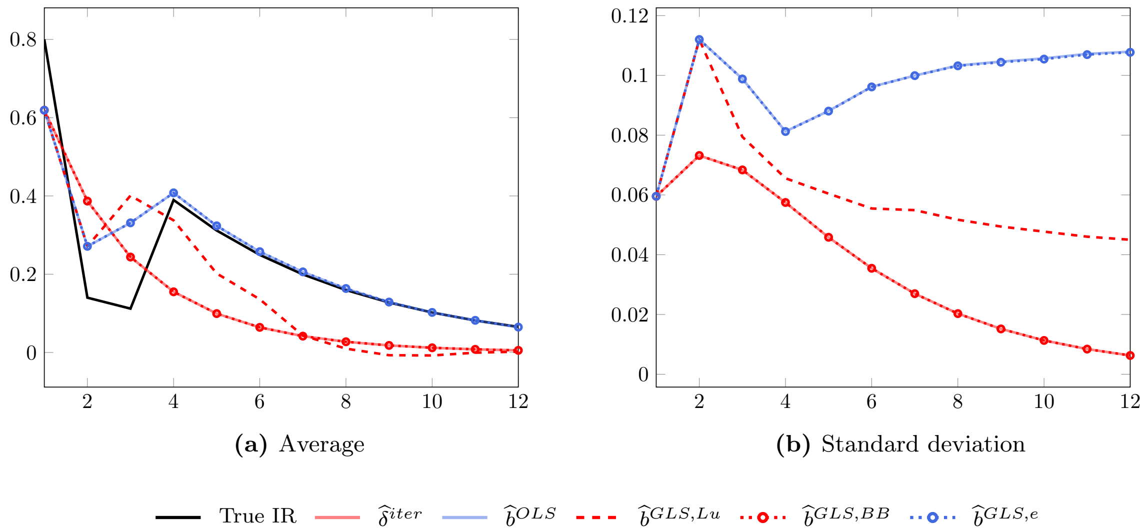
To illustrate the bias-variance trade-off between the estimators under consideration, we generate 5,000 samples using the following simple data-generating process:

$$y_{t+1} = \alpha y_t + \mu_{t+1} + \beta_1 \mu_{t-1} + \beta_2 \mu_{t-3}, \tag{20}$$

with $\mu_t \sim \mathcal{N}(0,1)$. The first 100 observations are discarded as burn-in. We set $\alpha = 0.8$ and consider two scenarios: one where the AR(1) model is correctly specified ($q = \beta_1 = \beta_2 = 0$) and another where it is misspecified ($q > 0$, $\beta_1 = -0.5$, $\beta_2 = 0.4$).

We estimate the impulse responses δ_h for $h = 1, \ldots, 12$ using the $\widehat{\delta}_h^{iter}$, $\widehat{b}_h^{OLS}$, $\widehat{b}_h^{GLS,Lu}$, $\widehat{b}_h^{GLS,BB}$ and $\widehat{b}_h^{GLS,LPe}$ estimators, all based on an AR(1) model. Thus, they are correctly specified when $\beta_1 = \beta_2 = 0$ but misspecified when $\beta_1 = -0.5$ and $\beta_2 = 0.4$. We do not report results for the $\widehat{b}_h^{GLS,full}$ estimator, as it is numerically equivalent to $\widehat{\delta}_h^{iter}$. Tables B-1-B-2 in Appendix B report the bias and the standard deviation of the considered estimators for $T \in \{25, 250\}$. Figure 1 illustrates these results for the misspecified model, estimated with a sample size of $T = 250$

Figure 1: Numerical illustration, trading robustness for efficiency



(a) Average **(b)** Standard deviation

Notes: Data samples of size $T = 250$ are generated from eq.(20) setting $\alpha = 0.8$, $\beta_1 = -0.5$ and $\beta_2 = 0.4$. Reported are averages and standard deviations across 5,000 Monte Carlo draws of the various considered IR estimators. The horizontal axis represents the projection horizon $h = 1, \ldots, 12$.

The results confirm the conclusions made above. All estimators are biased but consistent when the model is correctly specified, whereas they are biased and inconsistent in the misspecified case. In the latter setting, the LP OLS estimator $\widehat{b}^{OLS}$ is closer to the true impulse response than the iterated estimator $\widehat{\delta}^{iter}$, although not uniformly across the entire projection horizon. This illustrates the inherent higher robustness to misspecification by LP OLS, and the sensitivity of $\widehat{\delta}^{iter}$. Conversely, $\widehat{b}^{OLS}$ exhibits a much larger variance compared to $\widehat{\delta}^{iter}$, especially as the projection horizon increases. This is true irrespective of whether the model is correctly specified or not.

The LP GLS estimator $\widehat{b}^{GLS,Lu}$ of Lusompa (2023) has a smaller variance than the LP OLS estimator, but this comes at the cost of a higher bias, which, in some cases along the projection horizon, even surpasses the bias of $\widehat{\delta}^{iter}$. Overall, $\widehat{b}^{GLS,Lu}$ strikes a middle ground between $\widehat{b}^{OLS}$ and $\widehat{\delta}^{iter}$ in terms of both bias and variance. Clearly, the GLS transformation has indeed resulted in sacrificing some of the robustness of LP OLS for efficiency.

The LP GLS estimator $\widehat{b}^{GLS,BB}$ of Breitung and Brüggemann (2023) takes things a step further, and we confirm that the differences with $\widehat{\delta}^{iter}$ are negligibly small, as evidenced by the overlapping lines in Figure 1, regardless of whether the model is well-specified or misspecified. This is consistent with their asymptotic equivalence, as demonstrated in Section 2.

Lastly, the differences between the LP GLS estimator $\widehat{b}^{GLS,LPe}$, which uses LP residuals, and the LP OLS estimator are also minimal, with their lines similarly overlapping in Figure 1. This aligns with the asymptotic equivalence discussed in Section 2, showing that this transformation indeed does not lead to any improvements. For $T = 25$, $\widehat{b}^{GLS,LPe}$ tends to slightly inflate the variance.

4 Conclusion

This paper highlights the crucial trade-off between efficiency and robustness when estimating LPs using GLS. Depending on the specific implementation, GLS estimators either converge to the iterated LP estimator or the original LP OLS estimator.

GLS transformations that rely on dynamic model residuals – as in the approaches proposed by Lusompa (2023) and Breitung and Brüggemann (2023), and full GLS – achieve efficiency by imposing the dynamic model's structure over the entire projection horizon. While this results in estimators that are either close to or identical to forward-iterated IRs, it also makes them highly sensitive to misspecification. This trade-off should be a significant concern for researchers, as it undermines the very robustness that LPs are valued for.

On the other hand, iterative GLS LP methods, which use LP GLS residuals from previous horizons, fail to provide the expected efficiency improvements over LP OLS. The asymptotic distribution of these estimators remains equivalent to that of LP OLS, indicating that GLS corrections do not bring efficiency gains in this setting.

Although we have used a simple AR(1) setting for illustrative purposes, the conclusions naturally extend to higher-order, multivariate, and panel data models, where the same efficiency-robustness trade-off will arise when incorporating estimated forecast errors. It is crucial to emphasize that this trade-off disappears when LPs are augmented with ex-post observed forecast errors, as is for instance done by Faust and Wright (2013) and Teulings and Zubanov (2014). When these observed forecast errors are, in population, uncorrelated with the original regressors, they preserve the robustness of the LPs while simultaneously enhancing their efficiency. Further note that Teulings and Zubanov (2014) primarily use this LP augmentation to mitigate the incidental parameter bias of the Fixed-Effects LP estimator. In our analysis, we avoided such biases by excluding deterministic terms from the specifications, focusing solely on improving efficiency and leaving bias correction for future research.

In conclusion, this paper serves as a warning for researchers to carefully consider the use of GLS transformations in LPs, particularly in settings where robustness to misspecification is crucial.

References

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.

Breitung, J. and Brüggemann, R. (2023). Projection Estimators for Structural Impulse Responses. *Oxford Bulletin of Economics and Statistics*, 85(6):1320–1340.

Clark, T. E., Huber, F., Koop, G., Marcellino, M., and Pfarrhofer, M. (2024). Investigating Growth-at-Risk Using a Multicountry Nonparametric Quantile Factor Model. *Journal of Business & Economic Statistics*, 42(4):1302–1317.

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

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.

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

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

Lusompa, A. (2023). Local projections, autocorrelation, and efficiency. *Quantitative Economics*, 14:1199–1220.

Montiel Olea, J. L. and Plagborg-Møller, M. (2021). Local Projection Inference is Simpler and More Robust Than You Think. *Econometrica*, 89(4):1789–1823.

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.

Teulings, C. and Zubanov, N. (2014). Is Economic Recovery a Myth? Robust Estimation of Impulse Responses. *Journal of Applied Econometrics*, 29:497–514.

Appendices

Appendix A Proofs

Assumption 1. Let in eq.(1) $|\alpha| < 1$ and μ_t be an i.i.d. process with $\mathbb{E}(\mu_t) = 0$, $\mathbb{E}(\mu_t^2) = \sigma^2 > 0$ and $\mathbb{E}(\mu_t^4) < \infty$ for all t, and $\beta(q)$ is stable and invertible for all $q \geq 0$.

A.1 Lemmas

Lemma 1. $\widehat{\Gamma}_{T-h} = \frac{1}{T}\sum_{t=1}^{T-h} y_t^2 = \Gamma + O_p(T^{-1/2})$ and $\widehat{\Gamma}_{T-h}^{-1} = \Gamma^{-1} + O_p(T^{-1/2})$, with $\Gamma = \mathbb{E}(y_t^2) = O(1)$ as $T \to \infty$.

Proof: Assumption 1 implies that y_t is a mean 0, weakly dependent stationary series with finite 4th moments, from which follows that $\Gamma = \mathbb{E}(y_t^2) = \mathbb{V}ar(y_t) = O(1)$ for all t, with $\Gamma > 0$, and we have by standard results $\widehat{\Gamma}_{T-h} = \Gamma + O_p(T^{-1/2})$. Since $\Gamma > 0$ and $rk(\widehat{\Gamma}_{T-h}) - rk(\Gamma) \to 0$, we also have that $\widehat{\Gamma}_{T-h}^{-1} = \Gamma^{-1} + O_p(T^{-1/2})$ with $\Gamma^{-1} = O(1)$.

Lemma 2. $\widehat{b}_h^{OLS} - \delta_h^p = O_p(T^{-1/2})$ as $T \to \infty$ for $h \geq 1$ and $q \geq 0$.

Proof: Consider (4) and the pseudo-true IR parameter defined in (6): $\delta_h^p = \alpha^h + \sum_{j=1}^h \alpha^{h-j} c_j$, with $c_j = \mathbb{C}ov(y_t, \varepsilon_{t+j}^q)/\mathbb{V}ar(y_t)$. Then, the backward iterated LP is

$$y_{t+h} = \delta_h^p y_t + e_{t+h}^h, \qquad e_{t+h}^h = \sum_{j=1}^{h} \alpha^{h-j}(\varepsilon_{t+j}^q - c_j y_t), \tag{A-1}$$

where $\mathbb{C}ov(y_t, e_{t+h}^h) = 0$ by construction for all h and we note that e_{t+h}^h can also be written as in eq.(7). Accordingly, substituting (A-1) and $\widehat{\Gamma}_{T-h} = \frac{1}{T}\sum_{t=1}^{T-h} y_t^2$ into (4) yields

$$\widehat{b}_h^{OLS} = \left(\frac{1}{T}\sum_{t=1}^{T-h} y_t^2\right)^{-1} \frac{1}{T}\sum_{t=1}^{T-h} y_t(\delta_h^p y_t + e_{t+h}^h) = \delta_h^p + \widehat{\Gamma}_{T-h}^{-1}\frac{1}{T}\sum_{t=1}^{T-h} y_t e_{t+h}^h, \tag{A-2}$$

from which follows the required

$$\left|\widehat{b}_h^{OLS} - \delta_h^p\right| \leq \left|\widehat{\Gamma}_{T-h}^{-1}\right| \left|\frac{1}{T}\sum_{t=1}^{T-h} y_t e_{t+h}^h\right| = O_p(T^{-1/2}), \tag{A-3}$$

because $|\widehat{\Gamma}_{T-h}^{-1}| = O_p(1)$ and

$$\left|\frac{1}{T}\sum_{t=1}^{T-h} y_t e_{t+h}^h\right| = O_p(T^{-1/2}), \tag{A-4}$$

since y_t and e_{t+h}^h are uncorrelated and stationary variables with finite 4th moments by Assumption 1. Given that convergence is to the pseudo-true parameter, we note that the result holds irrespective of the misspecification parameter $q \geq 0$.

A.2 Proof of eq.(11)

The GLS estimator in eq.(9) can by substituting in (A-1) be written as

$$\begin{aligned}
\widehat{b}_h^{GLS,Lu} &= \left(\frac{1}{T}\sum_{t=1}^{T-h} y_t^2\right)^{-1} \frac{1}{T}\sum_{t=1}^{T-h} y_t \left(y_{t+h} - \sum_{j=1}^{h-1} \widehat{b}_{h-j}^{GLS,Lu} \widehat{\varepsilon}_{t+j}^q\right), \\
&= \widehat{\Gamma}_{T-h}^{-1} \frac{1}{T}\sum_{t=1}^{T-h} y_t \left(\delta_h^p y_t + e_{t+h}^h - \sum_{j=1}^{h-1} \widehat{b}_{h-j}^{GLS,Lu} \widehat{\varepsilon}_{t+j}^q\right), \\
&= \delta_h^p + \widehat{\Gamma}_{T-h}^{-1} \left[\frac{1}{T}\sum_{t=1}^{T-h} y_t e_{t+h}^h - \sum_{j=1}^{h-1} \widehat{b}_{h-j}^{GLS,Lu} \frac{1}{T}\sum_{t=1}^{T-h} y_t \widehat{\varepsilon}_{t+j}^q\right], \\
&= \delta_h^p - \widehat{\Gamma}_{T-h}^{-1} \sum_{j=1}^{h-1} \widehat{b}_{h-j}^{GLS,Lu} \left(\frac{1}{T}\sum_{t=1}^{T-h} y_t \widehat{\varepsilon}_{t+j}^q\right) + O_p(T^{-1/2}), \\
&= \delta_h^p - \widehat{\Gamma}_{T-h}^{-1} \sum_{j=1}^{h-1} \widehat{b}_{h-j}^{GLS,Lu} \widehat{\phi}_j + O_p(T^{-1/2}), \qquad \text{(A-5)}
\end{aligned}$$

where we defined $\widehat{\phi}_j = \frac{1}{T}\sum_{t=1}^{T-h} y_t \widehat{\varepsilon}_{t+j}^q$ and used $\widehat{\Gamma}_{T-h} = \frac{1}{T}\sum_{t=1}^{T-h} y_t^2 = O_p(1)$ by Lemma 1 together with $\left|\frac{1}{T}\sum_{t=1}^{T-h} y_t e_{t+h}^h\right| = O_p(T^{-1/2})$ obtained in (A-4).

Consider next that the estimated one-period ahead forecast errors are given by $\widehat{\varepsilon}_{t+j}^q = y_{t+j} - \widehat{\alpha} y_{t+j-1} = y_{t+j} - \widehat{b}_1^{OLS} y_{t+j-1}$ since by definition $\widehat{\alpha} = \widehat{b}_1^{OLS}$. In turn, setting $h = 1$ in (A-1) and shifting the time index reveals that $y_{t+j} = \delta_1^p y_{t+j-1} + e_{t+j}^1$ with $e_{t+j}^1 = \varepsilon_{t+j}^q - c_1 y_{t+j-1}$ and $c_1 = \mathbb{C}ov(y_t, \varepsilon_{t+1}^q)/\mathbb{V}ar(y_t)$. Accordingly, substituting the latter into the former yields

$$\widehat{\varepsilon}_{t+j}^q = e_{t+j}^1 - (\widehat{b}_1^{OLS} - \delta_1^p) y_{t+j-1}, \qquad \text{(A-6)}$$

which results in the following decomposition of $\widehat{\phi}_j$:

$$\begin{aligned}
\widehat{\phi}_j &= \frac{1}{T}\sum_{t=1}^{T-h} y_t (e_{t+j}^1 - (\widehat{b}_1^{OLS} - \delta_1^p) y_{t+j-1}) = \frac{1}{T}\sum_{t=1}^{T-h} y_t e_{t+j}^1 - (\widehat{b}_1^{OLS} - \delta_1^p) \frac{1}{T}\sum_{t=1}^{T-h} y_t y_{t+j-1}, \\
&= \frac{1}{T}\sum_{t=1}^{T-h} y_t e_{t+j}^1 + O_p(T^{-1/2}),
\end{aligned}$$

where the final line follows from $|T^{-1}\sum_{t=1}^{T-h} y_t y_{t+j-1}| = O_p(1)$ by the stationarity in Assumption 1 and because $|\widehat{b}_1^{OLS} - \delta_1^p| = O_p(T^{-1/2})$ by setting $h = 1$ in Lemma 2. Define next $\phi_j = \mathbb{E}(y_t e_{t+j}^1)$, and note that $\phi_1 = 0$ by definition for $j = 1$, and $\phi_j = O(1)$ in general by the stationarity in Assumption 1. Given that y_t and e_{t+j}^1 are covariance stationary variables with finite fourth moments, we have that $|T^{-1}\sum_{t=1}^{T-h} y_t e^1_{t+j} - \phi_j| = O_p(T^{-1/2})$, from which follows

$$\widehat{\phi}_j = \phi_j + O_p(T^{-1/2}). \tag{A-7}$$

Next, note that $\widehat{\Gamma}^{-1}_{T-h} = \Gamma^{-1} + O_p(T^{-1/2})$ (Lemma 1) and consider that for $h = 2$ we have $\widehat{b}_1^{GLS,Lu} = \widehat{b}_1^{OLS}$ in (A-5). Lemma 2 then implies $|\widehat{b}_1^{GLS,Lu}| = O_p(1)$ and $|\widehat{b}_1^{GLS,Lu} - \delta_1^p| = O_p(T^{-1/2})$, such that we obtain for $h = 2$

$$\widehat{b}_2^{GLS,Lu} = \delta_2^p - \widehat{\Gamma}^{-1}_{T-2}\widehat{b}_1^{GLS,Lu}\widehat{\phi}_j + O_p(T^{-1/2}) = \delta_2^p - \Gamma^{-1}\delta_1^p\phi_1 + O_p(T^{-1/2}), \tag{A-8}$$

which indicates in turn that $|\widehat{b}_2^{GLS,Lu}| = O_p(1)$. Accordingly, given the latter and also $|\widehat{b}_1^{GLS,Lu}| = O_p(1)$, we have for $h = 3$ by also substituting in (A-8)

$$\begin{aligned}
\widehat{b}_3^{GLS,Lu} &= \delta_3^p - \Gamma^{-1}\sum_{j=1}^{2}\widehat{b}_{3-j}^{GLS,Lu}\phi_j + O_p(T^{-1/2}),\\
&= \delta_3^p - \Gamma^{-1}\widehat{b}_2^{GLS,Lu}\phi_1 - \Gamma^{-1}\widehat{b}_1^{GLS,Lu}\phi_2 + O_p(T^{-1/2}),\\
&= \delta_3^p - \Gamma^{-1}(\delta_2^p - \Gamma^{-1}\delta_1^p\phi_1)\phi_1 - \Gamma^{-1}\delta_1^p\phi_2 + O_p(T^{-1/2}),
\end{aligned}$$

from which it is clear that the inconsistency is defined recursively, and we obtain by iteratively following the steps above, for general h,

$$\widehat{b}_h^{GLS,Lu} = \delta_h^p + B_h^{Lu} + O_p(T^{-1/2}), \tag{A-9}$$

where $B_h^{Lu} = -\Gamma^{-1}\sum_{j=1}^{h-1}\left(\delta_{h-j}^p + B_{h-j}^{Lu}\right)\phi_j$ and we have $B_1^{Lu} = 0$. The stated result then follows from subtracting $\widehat{b}_h^{OLS} = \delta_h^p + O_p(T^{-1/2})$ by Lemma 2:

$$\widehat{b}_h^{GLS,Lu} - \widehat{b}_h^{OLS} = B_h^{Lu} + O_p(T^{-1/2}), \tag{A-10}$$

as required.

A.3 Proof of eq.(12)

First note that backward iterating h periods yields

$$y_{t+h} = \widehat{\alpha}y_{t+h-1} + \widehat{\varepsilon}^q_{t+h} = \widehat{\alpha}^h y_t + \sum_{j=1}^{h-1}\widehat{\alpha}^{h-j}\widehat{\varepsilon}^q_{t+j} + \widehat{\varepsilon}^q_{t+h}.$$

Substituting this into the expression for $\widehat{b}_h^{GLS,Lu}$, we have for $h \geq 2$

$$\widehat{b}_h^{GLS,Lu} = \widehat{\Gamma}^{-1}_{T-h}\left[\frac{1}{T}\sum_{t=1}^{T-h} y_t\left(y_{t+h} - \sum_{j=1}^{h-1}\widehat{b}_{h-j}^{GLS,Lu}\widehat{\varepsilon}^q_{t+j}\right)\right],$$

$$= \widehat{\Gamma}_{T-h}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T-h} y_t\left(\widehat{\alpha}^h y_t + \sum_{j=1}^{h-1}(\widehat{\alpha}^{h-j} - \widehat{b}_{h-j}^{GLS,Lu})\widehat{\varepsilon}_{t+j}^q + \widehat{\varepsilon}_{t+h}^q\right)\right],$$
$$= \widehat{\alpha}^h + \widehat{\Gamma}_{T-h}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T-h} y_t\left(\widehat{\varepsilon}_{t+h}^q + \sum_{j=1}^{h-1}(\widehat{\alpha}^{h-j} - \widehat{b}_{h-j}^{GLS,Lu})\widehat{\varepsilon}_{t+j}^q\right)\right],$$
$$= \widehat{\alpha}^h + \widehat{\Gamma}_{T-h}^{-1}\left[\widehat{\phi}_h - \sum_{j=1}^{h-1}(\widehat{b}_{h-j}^{GLS,Lu} - \widehat{\alpha}^{h-j})\widehat{\phi}_j\right].$$

Consider then that for $h = 2$, by making use of $(\widehat{b}_1^{GLS,Lu} - \widehat{\alpha}) = 0$ and subsequently $\widehat{\Gamma}_{T-h}^{-1} = \Gamma^{-1} + O_p(T^{-1/2})$ (Lemma 1) and (A-7)

$$\widehat{b}_2^{GLS,Lu} - \widehat{\alpha}^2 = \widehat{\Gamma}_{T-2}^{-1}\left[\widehat{\phi}_2 - (\widehat{b}_1^{GLS,Lu} - \widehat{\alpha})\widehat{\phi}_j\right] = \widehat{\Gamma}_{T-2}^{-1}\widehat{\phi}_2 = C_2^{Lu} + O_p(T^{-1/2}), \tag{A-11}$$

where $C_2^{Lu} = \Gamma^{-1}\phi_2$.
Similarly, for $h = 3$, making use of (A-7) and substituting in (A-11) gives

$$\widehat{b}_3^{GLS,Lu} - \widehat{\alpha}^3 = \widehat{\Gamma}_{T-3}^{-1}\left[\widehat{\phi}_3 - \sum_{j=1}^{2}(\widehat{b}_{3-j}^{GLS,Lu} - \widehat{\alpha}^{3-j})\widehat{\phi}_j\right] = \widehat{\Gamma}_{T-3}^{-1}\left[\widehat{\phi}_3 - (\widehat{b}_2^{GLS,Lu} - \widehat{\alpha}^2)\widehat{\phi}_1\right],$$
$$= \Gamma^{-1}(\phi_3 - C_2^{Lu}\phi_1) + O_p(T^{-1/2}),$$
$$= C_3^{Lu} + O_p(T^{-1/2}), \tag{A-12}$$

where $C_3^{Lu} = \Gamma^{-1}(\phi_3 - C_2^{Lu}\phi_1)$.
Accordingly, iteratively substituting in expressions as above yields for general h

$$\widehat{b}_h^{GLS,Lu} - \widehat{\alpha}^h = C_h^{Lu} + O_p(T^{-1/2}), \qquad for \qquad h \geq 1,$$

where $C_h^{Lu} = \Gamma^{-1}\left(\phi_h - \sum_{j=1}^{h-1} C_{h-j}^{Lu}\phi_j\right)$, with $C_1^{Lu} = 0$, and where we used that h is finite. The stated expression then substitutes in the definition $\widehat{\delta}_h^{iter} = \widehat{\alpha}^h$.

A.4 Proof of eq.(15)

Let $\widehat{\alpha}_{T-a}$ be the OLS estimator in (1) estimated with $T - a$ observations. That is,

$$\widehat{\alpha}_{T-a} = \left(\frac{1}{T}\sum_{t=1}^{T-a} y_t^2\right)^{-1} \frac{1}{T}\sum_{t=1}^{T-a} y_t y_{t+1} = \widehat{\Gamma}_{T-a}^{-1}\widehat{\gamma}_{T-a}, \qquad for \qquad 1 \leq a \leq T-1, \tag{A-13}$$

and note that we can write, for any fixed and finite $2 \leq a \leq H$:

$$\widehat{\Gamma}_{T-a} = \frac{1}{T}\left[\sum_{t=1}^{T-1} y_t^2 - \sum_{\ell=0}^{a-1} y_{T-1-\ell}^2\right] = \widehat{\Gamma}_{T-1} - \frac{1}{T}\sum_{\ell=0}^{a-1} y_{T-1-\ell}^2 = \widehat{\Gamma}_{T-1} + O_p(T^{-1}),$$
$$\widehat{\gamma}_{T-a} = \frac{1}{T}\left[\sum_{t=1}^{T-1} y_t y_{t+1} - \sum_{\ell=0}^{a-1} y_{T-1-\ell}y_{T-\ell}\right] = \widehat{\gamma}_{T-1} - \frac{1}{T}\sum_{\ell=0}^{a-1} y_{T-1-\ell}y_{T-\ell} = \widehat{\gamma}_{T-1} + O_p(T^{-1}),$$

since we have by Assumption 1 that $|y_t y_{t+i}| = O_p(1)$ for all i, so that the fixed and finite a implies $|\sum_{\ell=0}^{a-1} y_{T-1-\ell}^2| = O_p(1)$ and $|\sum_{\ell=0}^{a-1} y_{T-1-\ell} y_{T-\ell}| = O_p(1)$ as $T \to \infty$. It thus follows for $2 \leq a \leq H$:

$$\widehat{\alpha}_{T-a} = \widehat{\Gamma}_{T-a}^{-1}\widehat{\gamma}_{T-a} = \widehat{\Gamma}_{T-1}^{-1}\widehat{\gamma}_{T-1} + O_p(T^{-1}) = \widehat{\alpha}_{T-1} + O_p(T^{-1}). \tag{A-14}$$

Consider then the estimator in eq.(14),

$$\widehat{b}_h^{GLS,BB} = \left(\sum_{t=1}^{T-h} y_t^2\right)^{-1} \sum_{t=1}^{T-h} y_t \left(y_{t+h} - \sum_{j=2}^{h} \widehat{b}_{h-j}^{GLS,BB} \widehat{\varepsilon}_{t+j}^{q}\right), \qquad \text{for} \qquad h = 2, \dots, H, \tag{A-15}$$

which is initiated with $\widehat{b}_1^{GLS,BB} = \widehat{\alpha}_{T-1}$ and $\widehat{b}_0^{GLS,BB} = 1$.
For $h = 2$, we have $\widehat{b}_2^{GLS,BB} = \left(\sum_{t=1}^{T-2} y_t^2\right)^{-1} \sum_{t=1}^{T-2} y_t \left(y_{t+2} - \widehat{\varepsilon}_{t+2}^{q}\right)$, where we note that

$$y_{t+h} = \widehat{\alpha}_{T-1} y_{t+h-1} + \widehat{\varepsilon}_{t+h}^{q}, \tag{A-16}$$

yields for $h = 2$ that $y_{t+2} - \widehat{\varepsilon}_{t+2}^{q} = \widehat{\alpha}_{T-1} y_{t+1}$, and therefore

$$\widehat{b}_2^{GLS,BB} = \frac{\sum_{t=1}^{T-2} y_t \left(y_{t+2} - \widehat{\varepsilon}_{t+2}^{q}\right)}{\sum_{t=1}^{T-2} y_t^2} = \widehat{\alpha}_{T-1} \frac{\sum_{t=1}^{T-2} y_t y_{t+1}}{\sum_{t=1}^{T-2} y_t^2} = \widehat{\alpha}_{T-1}\widehat{\alpha}_{T-2} = \widehat{\alpha}_{T-1}^2 + O_p(T^{-1}),$$

where we used $|\widehat{\alpha}_{T-1}| = O_p(1)$ (Lemma 2) and substituted in (A-14) with $a = 2$.

For $h = 3$, we similarly have by making use of (A-16)

$$y_{t+3} - \widehat{\varepsilon}_{t+3}^{q} - \widehat{b}_1^{GLS,BB}\widehat{\varepsilon}_{t+2}^{q} = \widehat{\alpha}_{T-1} y_{t+2} - \widehat{\alpha}_{T-1}\widehat{\varepsilon}_{t+2}^{q} = \widehat{\alpha}_{T-1}^2 y_{t+1},$$

such that, by substituting in (A-14) with $a = 3$,

$$\begin{aligned}\widehat{b}_3^{GLS,BB} &= \frac{\sum_{t=1}^{T-3} y_t \left(y_{t+3} - \widehat{\varepsilon}_{t+3}^{q} - \widehat{b}_1^{GLS,BB}\widehat{\varepsilon}_{t+2}^{q}\right)}{\sum_{t=1}^{T-3} y_t^2} = \widehat{\alpha}_{T-1}^2 \frac{\sum_{t=1}^{T-3} y_t y_{t+1}}{\sum_{t=1}^{T-3} y_t^2} = \widehat{\alpha}_{T-1}^2\widehat{\alpha}_{T-3},\\ &= \widehat{\alpha}_{T-1}^3 + O_p(T^{-1}).\end{aligned}$$

For $h = 4$, making use of $|\widehat{\varepsilon}_{t+\ell}^{q}| = O_p(1)$ and the previous results gives

$$\begin{aligned}y_{t+4} - \widehat{\varepsilon}_{t+4}^{q} - \widehat{b}_1^{GLS,BB}\widehat{\varepsilon}_{t+3}^{q} - \widehat{b}_2^{GLS,BB}\widehat{\varepsilon}_{t+2}^{q} &= \widehat{\alpha}_{T-1} y_{t+3} - \widehat{\alpha}_{T-1}\widehat{\varepsilon}_{t+3}^{q} - \widehat{\alpha}_{T-1}^2\widehat{\varepsilon}_{t+2}^{q} + O_p(T^{-1}),\\ &= \widehat{\alpha}_{T-1}^2 y_{t+2} - \widehat{\alpha}_{T-1}^2\widehat{\varepsilon}_{t+2}^{q} + O_p(T^{-1}),\\ &= \widehat{\alpha}_{T-1}^3 y_{t+1} + O_p(T^{-1}),\end{aligned}$$

where the second and third equalities substitute in (A-16). Accordingly, since averaging does not alter the stochastic orders

$$\widehat{b}_4^{GLS,BB} = \widehat{\Gamma}_{T-4}^{-1}\frac{1}{T}\sum_{t=1}^{T-4} y_t \left(y_{t+4} - \widehat{\varepsilon}_{t+4}^{q} - \widehat{b}_1^{GLS,BB}\widehat{\varepsilon}_{t+3}^{q} - \widehat{b}_2^{GLS,BB}\widehat{\varepsilon}_{t+2}^{q}\right),$$

$$
\begin{aligned}
&= \widehat{\Gamma}_{T-4}^{-1} \frac{1}{T} \sum_{t=1}^{T-4} y_t \left(\widehat{\alpha}_{T-1}^{3} y_{t+1} + O_p(T^{-1}) \right), \\
&= \widehat{\alpha}_{T-1}^{3} \widehat{\Gamma}_{T-4}^{-1} \widehat{\gamma}_{T-4} + O_p(T^{-1}), \\
&= \widehat{\alpha}_{T-1}^{3} \widehat{\alpha}_{T-4} + O_p(T^{-1}), \\
&= \widehat{\alpha}_{T-1}^{4} + O_p(T^{-1}),
\end{aligned}
$$

where the last step uses again (A-14).

Continuing in this fashion yields for general h

$$
\begin{aligned}
\widehat{b}_h^{GLS,BB} &= \frac{\frac{1}{T}\sum_{t=1}^{T-h} y_t \left(y_{t+h} - \sum_{j=2}^{h} \widehat{b}_{h-j}^{GLS,BB} \widehat{\varepsilon}_{t+j}^{q} \right)}{\frac{1}{T}\sum_{t=1}^{T-h} y_t^2} = \widehat{\alpha}_{T-1}^{h-1} \frac{\frac{1}{T}\sum_{t=1}^{T-h} y_t y_{t+1}}{\frac{1}{T}\sum_{t=1}^{T-h} y_t^2} + O_p(T^{-1}), \\
&= \widehat{\alpha}_{T-1}^{h-1} \widehat{\alpha}_{T-h} + O_p(T^{-1}) = \widehat{\alpha}_{T-1}^{h} + O_p(T^{-1}), \\
&= \widehat{\delta}_h^{iter} + O_p(T^{-1}),
\end{aligned}
$$

as was to be proved, where we used in the final line $\widehat{\delta}_h^{iter} = \widehat{\alpha}_{T-1}^{h}$ by definition.

A.5 Proof of eq.(19)

Consider the difference with the OLS estimator

$$
\widehat{b}_h^{GLS,LPe} - \widehat{b}_h^{OLS} = - \sum_{j=1}^{h-1} \left(\widehat{b}_{h-j}^{GLS,LPe} \frac{\sum_{t=1}^{T-h} y_t \widehat{v}_{t+j}^{j}}{\sum_{t=1}^{T-h} y_t^2} \right) = -\widehat{\Gamma}_{T-h}^{-1} \sum_{j=1}^{h-1} \widehat{b}_{h-j}^{GLS,LPe} \left(\frac{1}{T} \sum_{t=1}^{T-h} y_t \widehat{v}_{t+j}^{j} \right),
$$

and note that y_t and $\widehat{v}_{t+j}^{j}$ are orthogonal by construction over the sample period $1, \ldots, T-j$ by virtue of OLS estimation, whereas the summation runs over $t = 1, \ldots, T-h$, with $h > j$. Accordingly, we can write

$$
\begin{aligned}
\left| \frac{1}{T} \sum_{t=1}^{T-h} y_t \widehat{v}_{t+j}^{j} \right| &= \left| \frac{1}{T} \left[\sum_{t=1}^{T-j} y_t \widehat{v}_{t+j}^{j} - \sum_{\ell=1}^{h-j} y_{T-h+\ell} \widehat{v}_{T-h+\ell+j}^{j} \right] \right| = \frac{1}{T} \left| \sum_{\ell=1}^{h-j} y_{T-h+\ell} \widehat{v}_{T-h+\ell+j}^{j} \right|, \\
&= O_p(T^{-1}),
\end{aligned}
$$

where we used $|\sum_{\ell=1}^{h-j} y_{T-h+\ell} \widehat{v}_{T-h+\ell+j}^{j}| = O_p(1)$ because $|y_{T-h+\ell} \widehat{v}_{T-h+\ell+j}^{j}| = O_p(1)$ for all ℓ and $h-j$ is a fixed and finite integer. Noting then that $\widehat{b}_0^{GLS,LPe} = 1$, $\widehat{b}_1^{GLS,LPe} = \widehat{\alpha}$ and iteratively substituting in the result above shows that $|\widehat{b}_h^{GLS,LPe}| = O_p(1)$ for each h, and accordingly, that $|\widehat{b}_h^{GLS,LPe} - \widehat{b}_h^{OLS}| = O_p(T^{-1})$, as required.

Appendix B Simulation Results

Table B-1: MC results, correctly specified model

$T = 25$	Bias					Standard deviation				
h	$\widehat{\delta}^{iter}$	$\widehat{b}^{OLS}$	$\widehat{b}^{GLS,Lu}$	$\widehat{b}^{GLS,BB}$	$\widehat{b}^{GLS,LPe}$	$\widehat{\delta}^{iter}$	$\widehat{b}^{OLS}$	$\widehat{b}^{GLS,Lu}$	$\widehat{b}^{GLS,BB}$	$\widehat{b}^{GLS,LPe}$
1	-0.05	-0.05	-0.05	-0.05	-0.05	0.14	0.14	0.14	0.14	0.14
2	-0.06	-0.08	-0.08	-0.06	-0.08	0.19	0.22	0.22	0.19	0.22
3	-0.05	-0.09	-0.07	-0.06	-0.08	0.21	0.27	0.24	0.21	0.27
4	-0.04	-0.09	-0.07	-0.04	-0.08	0.21	0.31	0.26	0.21	0.31
5	-0.02	-0.08	-0.05	-0.03	-0.07	0.21	0.34	0.26	0.20	0.34
6	-0.01	-0.07	-0.05	-0.01	-0.06	0.20	0.36	0.26	0.19	0.38
7	0.01	-0.07	-0.03	-0.00	-0.05	0.19	0.39	0.26	0.18	0.41
8	0.02	-0.06	-0.03	0.01	-0.04	0.18	0.41	0.26	0.17	0.45
9	0.02	-0.04	-0.02	0.01	-0.02	0.17	0.44	0.26	0.16	0.50
10	0.03	-0.03	-0.02	0.02	0.00	0.16	0.47	0.26	0.15	0.56
11	0.03	-0.02	-0.01	0.02	0.02	0.16	0.51	0.27	0.14	0.63
12	0.04	-0.01	-0.02	0.02	0.04	0.15	0.54	0.27	0.13	0.72
$T = 250$	Bias					Standard deviation				
h	$\widehat{\delta}^{iter}$	$\widehat{b}^{OLS}$	$\widehat{b}^{GLS,Lu}$	$\widehat{b}^{GLS,BB}$	$\widehat{b}^{GLS,LPe}$	$\widehat{\delta}^{iter}$	$\widehat{b}^{OLS}$	$\widehat{b}^{GLS,Lu}$	$\widehat{b}^{GLS,BB}$	$\widehat{b}^{GLS,LPe}$
1	-0.01	-0.01	-0.01	-0.01	-0.01	0.04	0.04	0.04	0.04	0.04
2	-0.01	-0.01	-0.01	-0.01	-0.01	0.06	0.06	0.06	0.06	0.06
3	-0.01	-0.01	-0.01	-0.01	-0.01	0.07	0.08	0.08	0.07	0.08
4	-0.01	-0.01	-0.01	-0.01	-0.01	0.08	0.10	0.08	0.08	0.10
5	-0.00	-0.01	-0.01	-0.00	-0.01	0.08	0.11	0.08	0.07	0.11
6	-0.00	-0.01	-0.01	-0.00	-0.01	0.07	0.12	0.08	0.07	0.12
7	-0.00	-0.01	-0.00	-0.00	-0.01	0.07	0.12	0.08	0.07	0.12
8	0.00	-0.01	-0.00	0.00	-0.01	0.06	0.12	0.07	0.06	0.12
9	0.00	-0.01	-0.00	0.00	-0.01	0.06	0.13	0.07	0.06	0.13
10	0.00	-0.01	-0.00	0.00	-0.01	0.05	0.13	0.06	0.05	0.13
11	0.00	-0.01	-0.00	0.00	-0.01	0.04	0.13	0.06	0.04	0.13
12	0.00	-0.01	-0.00	0.00	-0.01	0.04	0.13	0.06	0.04	0.13

Notes: Data samples of size $T \in \{25, 250\}$ are generated from equation (20), with parameters set to $\alpha = 0.8$, $\beta_1 = \beta_2 = 0$. The reported values are the average bias and standard deviation across the projection horizon $h = 1, \ldots, 12$, based on 5,000 Monte Carlo simulations for the various impulse response estimators considered.

Table B-2: MC results, correctly specified model

$T=25$	Bias					Standard deviation				
h	$\widehat{\delta}^{iter}$	$\widehat{b}^{OLS}$	$\widehat{b}^{GLS,Lu}$	$\widehat{b}^{GLS,BB}$	$\widehat{b}^{GLS,LPe}$	$\widehat{\delta}^{iter}$	$\widehat{b}^{OLS}$	$\widehat{b}^{GLS,Lu}$	$\widehat{b}^{GLS,BB}$	$\widehat{b}^{GLS,LPe}$
1	-0.22	-0.22	-0.22	-0.22	-0.22	0.18	0.18	0.18	0.18	0.18
2	0.22	0.06	0.06	0.22	0.06	0.20	0.32	0.32	0.20	0.32
3	0.13	0.16	0.24	0.13	0.16	0.18	0.28	0.24	0.18	0.29
4	-0.22	-0.03	-0.10	-0.22	-0.03	0.16	0.26	0.23	0.16	0.26
5	-0.19	-0.04	-0.13	-0.19	-0.04	0.14	0.28	0.22	0.14	0.29
6	-0.16	-0.04	-0.11	-0.16	-0.03	0.12	0.30	0.21	0.12	0.32
7	-0.13	-0.03	-0.14	-0.13	-0.02	0.10	0.32	0.21	0.10	0.35
8	-0.11	-0.04	-0.13	-0.11	-0.02	0.09	0.34	0.20	0.09	0.38
9	-0.09	-0.04	-0.13	-0.09	-0.01	0.08	0.36	0.20	0.08	0.41
10	-0.07	-0.03	-0.11	-0.07	0.01	0.07	0.38	0.21	0.07	0.46
11	-0.05	-0.01	-0.09	-0.06	0.03	0.06	0.41	0.21	0.06	0.50
12	-0.04	-0.01	-0.07	-0.05	0.04	0.06	0.44	0.22	0.05	0.57
$T=250$	Bias					Standard deviation				
h	$\widehat{\delta}^{iter}$	$\widehat{b}^{OLS}$	$\widehat{b}^{GLS,Lu}$	$\widehat{b}^{GLS,BB}$	$\widehat{b}^{GLS,LPe}$	$\widehat{\delta}^{iter}$	$\widehat{b}^{OLS}$	$\widehat{b}^{GLS,Lu}$	$\widehat{b}^{GLS,BB}$	$\widehat{b}^{GLS,LPe}$
1	-0.18	-0.18	-0.18	-0.18	-0.18	0.06	0.06	0.06	0.06	0.06
2	0.25	0.13	0.13	0.25	0.13	0.07	0.11	0.11	0.07	0.11
3	0.13	0.22	0.29	0.13	0.22	0.07	0.10	0.08	0.07	0.10
4	-0.23	0.02	-0.05	-0.23	0.02	0.06	0.08	0.07	0.06	0.08
5	-0.21	0.01	-0.11	-0.21	0.01	0.05	0.09	0.06	0.05	0.09
6	-0.19	0.01	-0.11	-0.19	0.01	0.04	0.10	0.06	0.04	0.10
7	-0.16	0.01	-0.16	-0.16	0.01	0.03	0.10	0.05	0.03	0.10
8	-0.13	0.00	-0.15	-0.13	0.00	0.02	0.10	0.05	0.02	0.10
9	-0.11	0.00	-0.13	-0.11	0.00	0.02	0.10	0.05	0.02	0.10
10	-0.09	-0.00	-0.11	-0.09	0.00	0.01	0.11	0.05	0.01	0.11
11	-0.07	0.00	-0.08	-0.07	0.00	0.01	0.11	0.05	0.01	0.11
12	-0.06	-0.00	-0.06	-0.06	-0.00	0.01	0.11	0.05	0.01	0.11

Notes: Data samples of size $T \in \{25, 250\}$ are generated from equation (20), with parameters set to $\alpha = 0.8$, $\beta_1 = -0.5$, and $\beta_2 = 0.4$. The reported values are the average bias and standard deviation across the projection horizon $h = 1, \ldots, 12$, based on 5,000 Monte Carlo simulations for the various impulse response estimators considered.