ISSN 1936-5330

kcFED Research Working Papers

Local Projections, Autocorrelation, and Efficiency

Amaze Lusompa

March 2021
RWP 21-01
http://doi.org/10.18651/RWP2021-01

Federal Reserve Bank of Kansas City

Local Projections, Autocorrelation, and Efficiency

Amaze Lusompa*

Federal Reserve Bank of Kansas City

amaze.lusompa@kc.frb.org

March 3, 2021

Abstract

It is well known that Local Projections (LP) residuals are autocorrelated. Conventional wisdom says that LP have to be estimated by OLS with Newey-West (or some type of Heteroskedastic and Autocorrelation Consistent (HAC)) standard errors and that GLS is not possible because the autocorrelation process is unknown and/or because the GLS estimator would be inconsistent. I show that the autocorrelation process of LP is known and can be corrected for using a consistent GLS estimator. Estimating LP with GLS has three major implications: 1) LP GLS can be less biased, more efficient, and generally has better coverage properties than estimation by OLS with HAC standard errors. 2) Consistency of the LP GLS estimator gives a general counterexample showing that strict exogeneity is not a necessary condition for GLS. 3) Since the autocorrelation process can be modeled explicitly, it is now possible to estimate time-varying parameter LP.

*I thank Regis Barnichon, Bill Branch, Todd Clark, Ivan Jeliazkov, Òscar Jordà, Lutz Kilian, Daniel Lewis, Fabio Milani, Eric Swanson, Jonathan Wright, and seminar participants at several venues for helpful comments, discussions, and/or suggestions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1106401, the Federal Reserve Bank of San Francisco's Thomas J. Sargent Dissertation Fellowship, and the Federal Reserve Bank of Boston under the American Economic Association Summer Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation, the Federal Reserve Bank of San Francisco, the Federal Reserve Bank of Boston, the Federal Reserve Bank of Kansas City, or the Federal Reserve System.

1 Introduction

For decades, Vector Autoregressions (VARs) have been a major tool used in empirical macroeconomic analysis, primarily being used for causal analysis and forecasting through the estimation of impulse response functions. In a seminal paper, Jordà (2005) argued that impulse response functions could be estimated directly using linear regressions called Local Projections (LP) and that LP are more robust to model misspecification than VARs.[1,2] LP have been growing in popularity ever since, and the two methods often give different results when applied to the same problem (Ramey, 2016, Nakamura and Steinsson, 2018b). If the true model is a VAR, then a correctly specified VAR is more efficient than LP because VARs impose more structure than LP (Ramey, 2016). If the true model is not a finite order VAR or if the lag length of the VAR is not sufficiently long, then LP can outperform VARs (Plagborg-Møller and Wolf, 2019). Being that LP impulse responses nest VAR impulse responses, the choice of whether to use impulse responses from LP or VARs can be thought of as the bias-variance tradeoff problem with VARs and LP lying on a spectrum of small sample bias variance choices.

It is well known that LP residuals are autocorrelated. Practitioners exclusively estimate LP via OLS with Newey-West standard errors (or some type of Heteroskedastic and Autocorrelation Consistent (HAC) standard errors) (Ramey, 2016). Jordà (2005) argues that since the true data-generating process (DGP) is unknown, Generalized Least Squares (GLS) is not possible and HAC standard errors must be used. Hansen and Hodrick (1980) claim that direct forecast regressions (LP) cannot be estimated by GLS because estimates would be inconsistent.[3] I show that under standard time series assumptions, the autocorrelation process is known and can be corrected for using GLS. Moreover, I show the consistency and asymptotically normality of the LP GLS estimator.[4]

Being able to estimate LP with GLS has 3 major implications. First, LP GLS can be less biased, substantially more efficient, and generally has better coverage properties than LP estimated via OLS with HAC standard errors. Monte Carlo simulations for a wide range of models highlight the benefits of LP GLS. Moreover, under assumptions discussed in section 4, LP GLS impulse responses can be approximately as efficient as VAR impulse responses. Whether or not LP GLS impulse responses are approximately as efficient depends

[1]As noted in Stock and Watson (2018), LP are direct multistep forecasts. However, the goal of direct multistep forecast is an optimal multistep ahead forecast, whereas the goal of LP are accurate estimates of the corresponding impulse responses.

[2]In the case of stationary time series, Plagborg-Møller and Wolf (2019) show linear time-invariant VAR(∞) and LP(∞) estimate the same impulse responses.

[3]Hansen and Hodrick (1980) assume strict exogeneity (which neither LP or VARs satisfy) is a necessary condition for GLS.

[4]Montiel Olea and Plagborg-Møller (2020) have concurrently come up with an LP procedure that avoids HAC inference, but their procedure assumes the model can be written as a finite order VAR(p), at least $p+1$ lags are included in estimation, and that the error term satisfies mean independence. Many macroeconometric models cannot be written as finite order VARs (Nakamura and Steinsson, 2018b, Kilian and Lütkepohl, 2017). Mean independence is in general not satisfied by conditional heteroskedasticity models (see Brüggemann et al. (2016) Remark 2.5). It excludes many non-symmetric parametric models, models where the conditional error has a non-symmetric distribution, and stochastic volatility models (Goncalves and Kilian, 2004). The proposed GLS procedure is derived under more general conditions allowing for infinite lag representations and general conditional or unconditional heteroskedasticity.

on the persistence of the system, the horizon, and the dependence structure of the system. All else equal, the more persistent the system, the more likely LP GLS impulse responses will be approximately as efficient for horizons typically relevant in practice. It follows that LP can be much more efficient than previously believed.

Second, LP GLS shows that strict exogeneity is not a necessary condition for GLS estimation. It is often claimed in popular econometric textbooks that strict exogeneity is a necessary condition for GLS, which makes GLS more restrictive than OLS (Hayashi, 2000, Stock and Watson, 2007). Hansen and Hodrick (1980) is the earliest paper I find that claims strict exogeneity is a necessary condition, but they do not the provide a proof showing strict exogeneity is a necessary condition. Hamilton (1994) pg. 225 and Greene (2012) pg. 918 gives popular time series examples of what can go wrong when doing GLS with lagged endogenous variables, but simply showing an example where a GLS procedure is inconsistent is not sufficient to show that strict exogeneity is a necessary condition.[5] Since it was assumed that strict exogeneity is a necessary condition for GLS, GLS was in part abandoned for OLS with HAC estimation since OLS with HAC can be done under weaker conditions (Hayashi, 2000, Stock and Watson, 2007). Since consistency of LP GLS provides a general counterexample that strict exogeneity is not necessary condition for GLS, it follows that GLS estimation is not as restrictive as previously thought and that GLS may be extended to other situations where strict exogeneity is not satisfied.

Third, since autocorrelation is explicitly modeled, it is now possible to estimate time-varying parameter LP. This was not possible before because the Kalman filter and other popular techniques used to estimate time-varying parameter models require that the error term is uncorrelated or that the autocorrelation process is specified (Hamilton, 1994). Time-varying parameter models can be useful for several reasons. Researchers are often interested in whether there is parameter instability in regression models. As noted in Granger and Newbold (1977), macro data encountered in practice are unlikely to be stationary. Stock and Watson (1996) and Ang and Bekaert (2002) show many macroeconomic and financial time series exhibit parameter instability. It is also commonplace for regressions with macroeconomic time series to display heteroskedasticity of unknown form (Stock and Watson, 2007), and in order to do valid inference, the heteroskedasticity must be taken into account. Parameter instability can occur for many reasons such as policy changes, technological evolution, changing economic conditions, etc. If parameter instability is not appropriately taken into account, it can lead to invalid inference, poor out of sample forecasting, and incorrect policy evaluation. Moreover, as shown in Granger (2008), time-varying parameter models can approximate any non-linear model (non-linear in the variables and/or the parameters), which makes them more robust to model misspecification.

The paper is outlined as follows: Section 2 contains the core result showing that the autocorrelation

[5] LP GLS can consistently estimate both of them.

process of LP is known and illustrates why GLS is possible. Section 3 explains how to estimate LP GLS. Section 4 discusses the relative efficiency of LP estimated by OLS with HAC standard errors vs LP GLS. Section 5 contains Monte Carlo evidence of the small sample properties of LP GLS. Section 6 discusses time-varying parameter LP, Section 7 the empirical application to Gertler and Karadi (2015), and Section 8 concludes.

Some notation: $N(\cdot,\cdot)$ is the normal distribution. $plim$ is the probability limit, $\xrightarrow{p}$ is converges in probability, and $\xrightarrow{d}$ is converges in distribution. $\xrightarrow{p^*}$ is converges in probability, and $\xrightarrow{d^*}$ is converges in distribution with respect to the bootstrap probability measure. vec is the vector operator and $\otimes$ is the Kronecker product.

2 The Autocorrelation Process, OLS, and GLS

This section is broken up into 3 parts. Subsection 2.1 discusses how LP work, drawbacks of OLS estimation with LP, and how GLS estimation can improve upon it. Subsection 2.2 presents the core result: the autocorrelation process of LP is known and can be corrected for via GLS. Subsection 2.3 discusses how to do the FGLS correction and its basic properties.

2.1 LP and OLS

To illustrate how LP work, take the simple VAR(1) model

$$y_{t+1} = A_1 y_t + \varepsilon_{t+1},$$

where y_t is a demeaned $r \times 1$ vector of endogenous variables and ε_t is an $r \times 1$ vector white noise process with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \Sigma$. Assume that the eigenvalues of A_1 have moduli less than unity and $A_1 \neq 0$. Iterating forward leads to

$$y_{t+h} = A_1^h y_t + A_1^{h-1}\varepsilon_{t+1} + \ldots + A_1\varepsilon_{t+h-1} + \varepsilon_{t+h}.$$

To estimate the impulse responses of a VAR, one would estimate A_1 from equation (1) and then use the delta method, bootstrapping, or Monte Carlo integration to perform inference on the impulse responses: $\{A_1, \ldots, A_1^h\}$. To estimate impulse responses using LP, one would estimate the impulse responses directly at each horizon with separate regressions

$$y_{t+1} = B_1^{(1)} y_t + e_{t+1}^{(1)},$$

$$y_{t+2} = B_1^{(2)} y_t + e_{t+2}^{(2)},$$

$$\vdots$$

$$y_{t+h} = B_1^{(h)} y_t + e_{t+h}^{(h)},$$

where h is the horizon, and when the true DGP is a VAR(1), $\{B_1^{(1)}, \dots, B_1^{(h)}\}$ and $\{A_1, \dots, A_1^h\}$ are equivalent. Even if the true DGP is not a VAR(1), $B_1^{(1)} = A_1$ because the horizon 1 LP is a VAR. In practice, it is common for more than one lag to be used. A VAR(k) and the horizon h LP(k) can be expressed as

$$y_{t+1} = A_1 y_t + \dots + A_k y_{t-k+1} + \varepsilon_{t+1},$$

and

$$y_{t+h} = B_1^{(h)} y_t + \dots + B_k^{(h)} y_{t-k+1} + e_{t+h}^{(h)},$$

respectively. Bear in mind that any VAR(k) can be written as a VAR(1) (companion form), so results and examples involving the VAR(1) the can be generalized to higher order VARs.

LP have been advocated by Jordà (2005) as an alternative to VARs. There are several advantages of using LP as opposed to VARs. First, LP do not constrain the shape of the impulse response function like VARs, so it can be less sensitive to model misspecification (i.e. insufficient lag length) because misspecifications are not compounded in the impulse responses when iterating forward.[6] Second, LP can be estimated using simple linear regressions. Third, joint or point-wise analytic inference is simple. Fourth, LP can easily be adapted to handle non-linearities (in the variables or parameters).

As pointed out in Nakamura and Steinsson (2018b), Kilian and Lütkepohl (2017) among others, many macroeconometric models cannot be written as finite order VARs or LP. The finite order VARs and LP are just approximating infinite lag versions of themselves, so truncation bias (bias from omitted lags) can play a major role in inference. To illustrate, I conduct a simple Monte Carlo simulation where I generate 1,000 samples of length 250 from the following MA(35):

$$y_t = \varepsilon_t + \sum_{i=1}^{35} \theta_i \varepsilon_{t-i}, \; \varepsilon_t \sim N(0,1),$$

where

$$\theta_i = \frac{\theta_i^*}{\sum_{i=1}^{35} \theta_i^*}, \quad \theta_i^* = \alpha exp\Big\{-\Big(\frac{j-\beta}{\delta}\Big)^2\Big\} \quad for \; j = 1, \dots, 35 \,,$$

and

$$\alpha = 1, \qquad \beta = 6, \qquad \delta = 12.$$

The parameters are chosen so that the true impulse response is hump shaped, and the cumulative impulse

[6] In the case of the linear time-invariant estimators, VAR(∞) and LP(∞) estimate the same impulse responses asymptotically (Plagborg-Møller and Wolf, 2019). This result does not hold if the models are augmented with nonlinear terms.

response sums up to 1. I then estimate the impulse responses using a VAR and LP estimated with OLS, using lag lengths chosen by the AIC. Figure 1 plots the mean impulse response for both estimation methods along with the true impulse response. The VAR does an poor job approximating the shape of the impulse response, and it will be shown in the Monte Carlo section that when including uncertainty bands to do inference, the researcher will often come to the wrong conclusion. LP, though not perfect, can capture the shape of the true impulse response. The general qualitative results are not sensitive to using BIC, HQIC, or using the shortest lag length that yields white noise residuals for the VAR. Plagborg-Møller and Wolf (2019) prove that in finite samples it's possible to choose a large enough lag length such that the LP and VAR impulse responses are approximately the same, but there is currently no method or criteria for how to select such a lag length and obviously the impulse responses need not agree when using popular lag length selection procedures.

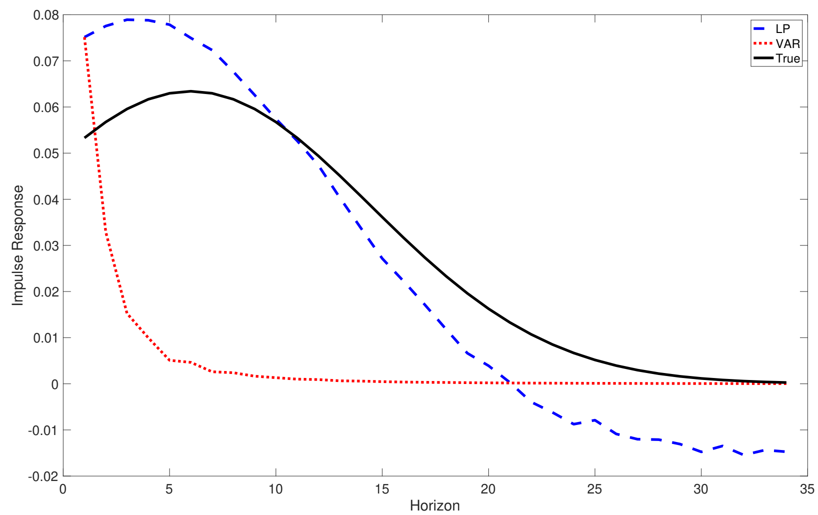

Figure 1: Mean Impulse Response of MA(35)

LP do have a couple of drawbacks. First, because the dependent variable is a lead, a total of h observations are lost from the original sample when estimating projections for horizon h. Second, the error terms in LP for horizons greater than 1 are inherently autocorrelated. Assuming the true model is a VAR(1), it is obvious that autocorrelation occurs because the LP residuals follow an VMA($h-1$) process of the residuals in equation (1). That is,

$$e_{t+h}^{(h)} = A_1^{h-1}\varepsilon_{t+1} + \ldots + A_1\varepsilon_{t+h-1} + \varepsilon_{t+h},$$

or written in terms of LP

$$e_{t+h}^{(h)} = B_1^{(h-1)}\varepsilon_{t+1} + \ldots + B_1^{(1)}\varepsilon_{t+h-1} + \varepsilon_{t+h}.$$

Frequentists account for the inherent autocorrelation using HAC standard errors, which will yield asymptotically correct standard errors in the presence of autocorrelation and heteroskedasticity of unknown forms.[7] Autocorrelation can be corrected for explicitly by including $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the conditioning set of the horizon h LP. Obviously $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ are unobserved and would have to be estimated, but this issue can be ignored for now and is addressed later.

There are two major advantages of correcting for autocorrelation explicitly. The first is that it fixes what I dub the "increasing variance problem". To my knowledge, the increasing variance problem has not been noticed in the literature. If the true model is a VAR(1), then $var(e^{(h)}_{t+h}) = \sum_{i=0}^{h-1} A_1^i \Sigma A_1^{i\prime}$, which is increasing in h.[8] HAC standard errors are valid in the presence of autocorrelation because they take into account autocorrelation is present when estimating the covariance matrix; they do not, however, eliminate autocorrelation.[9,10] To illustrate, let the true model be an AR(1) with

$$y_{t+1} = .99 y_t + \varepsilon_{t+1},$$

where $var(\varepsilon_t) = 1$. The $var(e^{(h)}_{t+h}) = \sum_{i=0}^{h-1} A_1^i \Sigma_\varepsilon A_1^{i\prime} = \sum_{i=0}^{h} .99^{2i}$. The table below presents the asymptotic variance of the residuals for different horizons when estimated by OLS with HAC standard errors vs LP estimated with GLS.

Table 1: Asymptotic Variance of Residuals for LP Horizons

Horizons	5	10	20	40
LP OLS	5.7093	9.9683	17.3036	28.2102
LP GLS	1	1	1	1

Even if HAC standard errors are used, the increasing variance problem persists. In terms of the MLE and OLS, correcting for autocorrelation explicitly is asymptotically more efficient because $var(\varepsilon_{t+h}) \leq var(e^{(h)}_{t+h})$, where the equality only binds when $A_1 = 0$. The increasing variance problem not only causes standard errors to be larger than they have to be, the larger variance is one of the reasons why LP impulse responses are sometimes erratic.[11]

The second major advantage of correcting for autocorrelation explicitly is that it helps remedy what I dub the "increased small sample bias problem". When LP are estimated with OLS and HAC standard errors, the small sample bias from estimating dynamic models increases relative to the model with no autocorrelation. To see why, let us first review the finite sample bias problem with VARs (see (Pope, 1990) for detailed

[7]There is a major line of research indicating that HAC standard errors perform poorly in small samples with persistent data (Müller, 2014).

[8]Since A_1 has moduli less than unity, geometric progression can be used to show that the sum is bounded asymptotically.

[9]This is a major reason why Kilian and Kim (2011) found that LP had excessive average length relative to the bias-adjusted bootstrap VAR interval in their Monte Carlo simulations. I provide Monte Carlo evidence of this in section 5.

[10]Macro variables tend to be persistent, so A_1^i may decay slowly leading to the increase in the variance to be pretty persistent as h increases.

[11]Obviously, eliminating the increasing variance problem would not prevent erratic behavior of LP impulse responses since they are not restricted.

derivations). Assume the true model is a VAR(1). The OLS estimate for the VAR is

$$\hat{A}_1 = A_1 + \sum_{t=1}^{T-1} \varepsilon_{t+1} y_t' (\sum_{t=1}^{T-1} y_t y_t')^{-1}.$$

This estimate is biased in finite samples because $E(\sum_{t=1}^{T-1} \varepsilon_{t+1} y_t' (\sum_{t=1}^{T-1} y_t y_t')^{-1}) \neq 0$ because ε_{t+1} and $y_t'(\sum_{t=1}^{T-1} y_t y_t')^{-1}$ are not uncorrelated, i.e. strict exogeneity is not satisfied. The stronger the correlation between ε_{t+1} and $y_t'(\sum_{t=1}^{T-1} y_t y_t')^{-1}$, the larger the bias. In macroeconomic applications, the bias is typically downward. The bias disappears asymptotically since ε_{t+1} would be correlated with an increasingly smaller share of $y_t'(\sum_{t=1}^{T-1} y_t y_t')^{-1}$.

If one were to estimate LP via OLS with HAC standard errors at horizon h, the OLS estimate would be

$$\hat{B}_1^{(h),OLS} = B_1^{(h)} + \sum_{t=1}^{T-h} e_{t+h}^{(h)} y_t' (\sum_{t=1}^{T-h} y_t y_t')^{-1}.$$

If one were to correct for autocorrelation by including $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$, the estimate would be

$$\hat{B}_1^{(h),GLS} = B_1^{(h)} + \sum_{t=1}^{T-h} \varepsilon_{t+h} y_t' (\sum_{t=1}^{T-h} y_t y_t')^{-1}.$$

The absolute value of the correlation between $e_{t+h}^{(h)}$ and $y_t'(\sum_{t=1}^{T-h} y_t y_t')^{-1}$ is larger than the absolute value of the correlation between ε_{t+h} and $y_t'(\sum_{t=1}^{T-h} y_t y_t')^{-1}$ because $e_{t+h}^{(h)} = A_1^{h-1}\varepsilon_{t+1} + \ldots + A_1 \varepsilon_{t+h-1} + \varepsilon_{t+h}$ is correlated with a larger share of $y_t'(\sum_{t=1}^{T-h} y_t y_t')^{-1}$.[12] To illustrate, I conduct a simple Monte Carlo simulation where I generate 1,000 samples of length 200 for the following AR(1):

$$y_{t+1} = .99 y_t + \varepsilon_{t+1},$$

where $var(\varepsilon_t) = 1$. I then estimate the impulse responses using a VAR, LP estimated with OLS, and LP estimated with GLS. To correct for autocorrelation using GLS, I include the estimated residuals. Below is the table of the mean impulse responses at different horizons for the different methods.

Table 2: Mean Impulse Response Estimates for T=200

Horizons	5	10	20	40
True	.951	.9044	.8179	.6690
VAR	.8355	.7072	.5231	.3148
LP OLS	.8259	.6713	.4223	.0787
LP GLS	.8347	.7045	.5160	.2965

[12] This is a major reason why Kilian and Kim (2011) found that LP impulse responses were more biased than the VAR impulse responses in their Monte Carlo simulations.

All of the estimates can be substantially biased, but not correcting for autocorrelation can make the bias considerably worse. Even if autocorrelation is corrected for in LP, there can still be a small sample bias due to the correlation between ε_{t+h} and $y_t'(\sum_{t=1}^{T-h} y_t y_t')^{-1}$ not being zero in finite samples, but additional bias due to not explicitly correcting for autocorrelation would be eliminated.[13] Intuitively, the bias in LP OLS is caused by the autocorrelation in the errors, and by strict exogeneity not being satisfied. The reason why LP GLS and VAR OLS bias are approximately the same in this example is because the only thing causing the bias in both cases is strict exogeneity not being satisfied, while in the OLS case it's additionally being caused by the autocorrelation in the errors.

2.2 The Autocorrelation Process of LP

First, I will show that even when the true DGP is not a VAR, including the horizon 1 LP residuals (or equivalently, VAR residuals), $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$, in the horizon h conditioning set will eliminate autocorrelation as long as the data are stationary and the horizon 1 LP residuals are uncorrelated. Second, I will show that the autocorrelation process of $e_{t+h}^{(h)}$ is known.

Assumption 1. *The data $\{y_t\}$ are covariance stationary and purely non-deterministic so there exists a Wold representation*

$$y_t = \varepsilon_t + \sum_{i=1}^{\infty} \Theta_i \varepsilon_{t-i}.$$

Assumption 1 implies that by the Wold representation theorem, there exists a linear and time-invariant Vector Moving Average (VMA) representation of the uncorrelated one-step ahead forecast errors $\{\varepsilon_t\}$. It follows from the Wold representation theorem that $\varepsilon_t = y_t - Proj(y_t|y_{t-1}, y_{t-2}, \ldots)$ where $Proj(y_t|y_{t-1}, y_{t-2}, \ldots)$ is the (population) orthogonal projection of y_t onto $\{y_{t-1}, y_{t-2}, \ldots\}$.

Consider for each horizon $h = 1, 2, \ldots$ the infinite lag linear LP

$$y_{t+h} = B_1^{(h)} y_t + B_2^{(h)} y_{t-1} + \ldots + e_{t+h}^{(h)}.$$

Proposition 1. *Under Assumption 1, including $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the conditioning set of the horizon h LP will eliminate autocorrelation in the horizon h LP residuals.*

Proof. I first show that

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_t, y_{t-1}, \ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+h-1}, y_{t+h-2}, \ldots).$$

From the Wold representation we know that $\varepsilon_{t+h-1} = y_{t+h-1} - Proj(y_{t+h-1}|y_{t+h-2}, y_{t+h-3}, \ldots)$, which implies

[13] LP GLS tends to be a little more biased than the VAR because LP estimated at horizon h lose h observations at the end of the sample.

that $\{\varepsilon_{t+h-1}, y_{t+h-1}, y_{t+h-2}, y_{t+h-3}, \ldots\}$ are linearly dependent. This implies that y_{t+h-1} can be dropped from $Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+h-1}, y_{t+h-2}, \ldots)$ since it contains redundant information. Therefore,

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+h-1}, y_{t+h-2}, \ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+h-2}, y_{t+h-3}, \ldots).$$

Similarly, $\varepsilon_{t+h-2} = y_{t+h-2} - Proj(y_{t+h-2}|y_{t+h-3}, y_{t+h-4}, \ldots)$, which implies that $\{\varepsilon_{t+h-2}, y_{t+h-2}, y_{t+h-3}, y_{t+h-4}, \ldots\}$ are linearly dependent. This implies that y_{t+h-2} can be dropped from $Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+h-2}, y_{t+h-3}, \ldots)$ since it contains redundant information. Therefore,

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+h-2}, y_{t+h-3}, \ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+h-3}, y_{t+h-4}, \ldots).$$

This process is repeated until y_{t+1} is being dropped due to linear dependence yielding

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+1}, y_t, \ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_t, y_{t-1}, \ldots).$$

Therefore, if the data are stationary and the horizon 1 LP residuals are uncorrelated,

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_t, y_{t-1}, \ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_{t+h-1}, y_{t+h-2}, \ldots).$$

It follows that

$$[y_{t+h} - Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_t, y_{t-1}, \ldots)] \perp [y_{t+h-i} - Proj(y_{t+h-i}|\varepsilon_{t+h-i-1}, \ldots, \varepsilon_{t-i+1}, y_{t-i}, y_{t-i-1}, \ldots)] \forall i \geq 1,$$

where $\perp$ is the orthogonal symbol. □

Therefore, if the data are stationary and the residuals $\{\varepsilon_t\}$ are uncorrelated, autocorrelation can be eliminated in the horizon h LP by including $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the conditioning set. Of course, if the true model requires only finitely many lags in the LP specification, then the proof above applies to that case as well, since the extraneous lags will all have coefficients of zero in population.

Theorem 1. *Under Assumption 1, the autocorrelation process of the horizon h LP residuals ($e_{t+h}^{(h)}$) is known.*

Proof. We know from the Wold representation that $\varepsilon_t \perp y_{t-1}, y_{t-2}, \ldots$, hence $\varepsilon_t \perp \varepsilon_s \; for \; t \neq s$. Recall that the infinite lag horizon h LP is

$$y_{t+h} = B_1^{(h)} y_t + B_2^{(h)} y_{t-1} + \ldots + e_{t+h}^{(h)} = Proj(y_{t+h}|y_t, y_{t-1}, \ldots) + e_{t+h}^{(h)}.$$

By Proposition 1, including $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the conditioning set eliminates autocorrelation, so the horizon h LP can be rewritten as

$$y_{t+h} = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_t, y_{t-1}, \ldots) + u_{t+h}^{(h)},$$

where $u_{t+h}^{(h)} = e_{t+h}^{(h)} - Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}) = e_{t+h}^{(h)} - Proj(y_{t+h}|\varepsilon_{t+h-1}) - \ldots - Proj(y_{t+h}|\varepsilon_{t+1})$. The $Proj$ can be broken up additively because $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ are orthogonal to each other and to $\{y_t, y_{t-1}, \ldots\}$. By Proposition 1, $u_{t+h}^{(h)}$ is not autocorrelated. By the Wold representation we know that

$$Proj(y_{t+h}|\varepsilon_t) = \Theta_h \varepsilon_t.$$

This implies, the horizon h LP can be written as

$$y_{t+h} = B_1^{(h+1)} y_t + B_2^{(h+1)} y_{t-1} + \ldots + \Theta_{h-1}\varepsilon_{t+1} + \ldots + \Theta_1 \varepsilon_{t+h-1} + u_{t+h}^{(h)},$$

which implies

$$e_{t+h}^{(h)} = \Theta_{h-1}\varepsilon_{t+1} + \ldots + \Theta_1 \varepsilon_{t+h-1} + u_{t+h}^{(h)}.$$

As a result, the autocorrelation process of $e_{t+h}^{(h)}$ is known. Using the same linear dependence arguments as in Proposition 1, it can be shown that

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, y_t, y_{t-1}, \ldots) = Proj(y_{t+h}|y_{t+h-1}, y_{t+h-2}, \ldots),$$

which implies that

$$u_{t+h}^{(h)} = \varepsilon_{t+h},$$

in population. □

Thus in population, the error process is a VMA($h-1$) even if the true model is not a VAR. In population

$$B_1^{(h)} = \Theta_h,$$

which implies

$$e_{t+h}^{(h)} = B_1^{(h-1)} \varepsilon_{t+1} + \ldots + B_1^{(1)} \varepsilon_{t+h-1} + \varepsilon_{t+h}.$$

2.3 LP GLS and Its Properties

Since $e_{t+h}^{(h)}$ can be written as

$$e_{t+h}^{(h)} = B_1^{(h-1)} \varepsilon_{t+1} + \ldots + B_1^{(1)} \varepsilon_{t+h-1} + u_{t+h}^{(h)},$$

GLS can be used to eliminate autocorrelation in LP while avoiding increasing the number of parameters by including $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the horizon h conditioning set. To understand how, I'll first explain what happens when $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ is included in the conditioning set. Just like it is impossible to estimate a VAR(∞) in practice, one cannot estimate LP with infinite lags since there is insufficient data. In practice truncated LP are used where the lags are truncated at k. The proofs of consistency and asymptotic normality discuss the rate at which k needs to grow with the sample size to ensure consistent estimation of the impulse responses. In practice, k, needs to be large enough that the estimated residuals from the horizon 1 LP are uncorrelated, which is what will be assumed for now. From Theorem 1 we know the horizon h LP is

$$y_{t+h} = B_1^{(h)} y_t + \ldots + B_k^{(h)} y_{t-k+1} + B_1^{(h-1)} \varepsilon_{t+1} + \ldots + B_1^{(1)} \varepsilon_{t+h-1} + u_{t+h,k}^{(h)},$$

where $u_{t+h,k}^{(h)}$ is the lag k analogue of $u_{t+h}^{(h)}$. Due to $\{\varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ being unobserved, the estimates $\{\hat{\varepsilon}_{t+1,k}, \ldots, \hat{\varepsilon}_{t+h-1,k}\}$ from the horizon 1 LP/VAR with k lags must be used instead. Estimates of the impulse responses are still consistent (will be shown in Theorem 2), however, even if the sample size is large, textbook formulas for GLS standard errors underrepresent uncertainty because $\{\hat{\varepsilon}_{t+1,k}, \ldots, \hat{\varepsilon}_{t+h-1,k}\}$ are generated regressors (Pagan, 1984) and because the textbook formulas for GLS assume strict exogeneity is satisfied. In order to do valid inference, one must use formulas that take into account that the generated regressors were estimated, which the FGLS estimator does.[14]

Including $\{\hat{\varepsilon}_{t+1,k}, \ldots, \hat{\varepsilon}_{t+h-1,k}\}$ in the conditioning set increases the number of parameters in each equation in the system by $(h-1) \times r$. If consistent estimates of $\{B_1^{(h-1)}, \ldots, B_1^{(1)}\}$ are obtained in previous horizons, one can do a Feasible GLS (FGLS) transformation. Let $\widetilde{y}_{t+h}^{(h)} = y_{t+h} - \hat{B}_1^{(h-1),GLS} \hat{\varepsilon}_{t+1,k} - \ldots - \hat{B}_1^{(1),OLS} \hat{\varepsilon}_{t+h-1,k}$. Then one can estimate horizon h via the following equation:

$$\widetilde{y}_{t+h}^{(h)} = B_1^{(h)} y_t + \ldots + B_k^{(h)} y_{t-k+1} + \tilde{u}_{t+h,k}^{(h)}.$$

$\widetilde{y}_{t+h}^{(h)}$ is just a GLS transformation that eliminates the autocorrelation problem in LP without having to sacrifice degrees of freedom and $\tilde{u}_{t+h,k}^{(h)}$ is the error term corresponding to this FGLS transformation. If the impulse responses are estimated consistently, then by the continuous mapping theorem, $\widetilde{y}_{t+h}^{(h)}$ converges in probability to the true GLS transformation $y_{t+h}^{(h)} = y_{t+h} - B_1^{(h-1)} \varepsilon_{t+1} - \ldots - B_1^{(1)} \varepsilon_{t+h-1}$ asymptotically. For clarification LP can be estimated sequentially horizon by horizon as follows. First estimate the horizon 1 LP/VAR

$$y_{t+1} = B_1^{(1)} y_t + \ldots + B_k^{(1)} y_{t-k+1} + \varepsilon_{t+1,k}.$$

[14]In the proof of asymptotic normality of the limiting distribution, it can be seen that the impact of the generated regressors does not disappear asymptotically.

$\hat{B}_1^{(1),OLS}$ and $\hat{\varepsilon}_{t,k}$ are estimates of $B_1^{(1)}$ and $\varepsilon_{t,k}$ respectively . Horizon 2 can be estimated as

$$\widetilde{y}_{t+2}^{(2)} = B_1^{(2)} y_t + \ldots + B_k^{(2)} y_{t-k+1} + \tilde{u}_{t+2,k}^{(2)},$$

where $\tilde{y}_{t+2}^{(2)} = y_{t+2} - \hat{B}_1^{(1),OLS} \hat{\varepsilon}_{t+1,k}$, and $\hat{B}_1^{(2),GLS}$ is the GLS estimate of $B_1^{(2)}$. Horizon 3 can be estimated as

$$\tilde{y}_{t+3}^{(3)} = B_1^{(3)} y_t + \ldots + B_k^{(3)} y_{t-k+1} + \tilde{u}_{t+3,k}^{(3)},$$

where $\widetilde{y}_{t+3}^{(3)} = y_{t+3} - \hat{B}_1^{(2),GLS} \hat{\varepsilon}_{t+1,k} - \hat{B}_1^{(1),OLS} \hat{\varepsilon}_{t+2,k}$, and $\hat{B}_1^{(3),GLS}$ is the GLS estimate of $B_1^{(3)}$. Horizon 4 can be estimated as

$$\tilde{y}_{t+4}^{(4)} = B_1^{(4)} y_t + \ldots + B_k^{(4)} y_{t-k+1} + \tilde{u}_{t+4,k}^{(4)},$$

where $\widetilde{y}_{t+4}^{(4)} = y_{t+4} - \hat{B}_1^{(3),GLS} \hat{\varepsilon}_{t+1,k} - \hat{B}_1^{(2),GLS} \hat{\varepsilon}_{t+2,k} - \hat{B}_1^{(1),OLS} \hat{\varepsilon}_{t+3,k}$, $\hat{B}_1^{(4),GLS}$ is the GLS estimate of $B_1^{(4)}$, and so on.

The LP GLS estimator has several desirable properties. But first, some assumptions need to be introduced.

Assumption 2. *Let y_t satisfy the Wold representation as presented in Assumption 1. Assume that in addition*
(i) ε_t is strictly stationary and ergodic such that $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ $a.s$, where $\mathcal{F}_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ is the sigma field generated by $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$.
(ii) $E(\varepsilon_t \varepsilon_t') = \Sigma$ is positive definite.
(iii) Θ_h satisfy $\sum_{h=0}^{\infty} \parallel \Theta_h \parallel < \infty$ where $\parallel \Theta_h \parallel^2 = tr(\Theta_h' \Theta_h)$ with $\Theta_0 = I_r$.

Note that for any Wold representation $det\{\Theta(z)\} \neq 0$ for $|z| \leq 1$ where $\Theta(z) = \sum_{h=0}^{\infty} \Theta_h z^h$. It follows from Assumption 2 that the Wold representation can be written as an infinite order VAR representation

$$y_t = \sum_{j=1}^{\infty} A_j y_{t-j} + \varepsilon_t,$$

with $\sum_{j=1}^{\infty} \parallel A_j \parallel < \infty$ and $A(z) = \Theta(z)^{-1}$. By recursive substitution

$$y_{t+h} = B_1^{(h)} y_t + B_2^{(h)} y_{t-1} + \ldots + \varepsilon_{t+h} + \Theta_1 \varepsilon_{t+h-1} + \ldots + \Theta_{h-1} \varepsilon_{t+1},$$

where $B_1^{(h)} = \Theta_h$, $B_j^{(h)} = \Theta_{h-1} A_j + B_{j+1}^{(h-1)}$ for $h \geq 1$ and with $B_{j+1}^{(0)} = 0$; $\Theta_0 = I_r$ for $j \geq 1$. The standard horizon h LP consists of estimating Θ_h from a least squares estimate of A_1^h with truncated regression

$$y_{t+h} = B_1^{(h)} y_t + \ldots + B_k^{(h)} y_{t-k+1} + e_{t+h,k}^{(h)},$$

where

$$e_{t+h,k}^{(h)} = \sum_{j=k+1}^{\infty} B_j^{(h)} y_{t-j+1} + \varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}.$$

Assumption 3. *Let y_t satisfy Assumption 2. Assume that in addition,*

(i) The r-dimensional ε_t has absolutely summable cumulants up to order 8. That is,

$$\sum_{i_2=-\infty}^{\infty} \cdots \sum_{i_h=-\infty}^{\infty} \mid \kappa_{\boldsymbol{a}}(0, i_2, \dots, i_j) \mid < \infty \; for \; j = 2, \dots, 8,$$

$a_1, \dots, a_j \in \{1, \dots, r\}$, $\boldsymbol{a} = \{a_1, \dots, a_j\}$, *and* $\kappa_{\boldsymbol{a}}(0, i_2, \dots, i_j)$ *denotes the jth joint cumulant of* $\varepsilon_{0,a_1}, \varepsilon_{i_2,a_2}, \dots, \varepsilon_{i_j,a_j}$. *In particular, this condition includes the existence of the eight moments of* ε.

(ii) $L_r E(vec(\varepsilon_t \varepsilon_{t-j}') vec(\varepsilon_t \varepsilon_{t-j}')') L_r'$ *is positive definite for all* j*, and* L_r is a finite $r(r+1)/2 \times r^2$ elimination matrix as defined $L_r vec(A) = vech(A)$.

(iii) k *satisfies*

$$\frac{k^4}{T} \to 0; \; T, k \to \infty.$$

(iv) k *satisfies*

$$(T - k - H)^{1/2} \sum_{j=k+1}^{\infty} \parallel A_j \parallel \to 0 \; T, k \to \infty.$$

Theorem 2. *Under Assumption 3, the LP GLS estimator is consistent. In particular*

$$\hat{B}_1^{(h),GLS} \xrightarrow{p} \Theta_h.$$

More generally

$$\parallel \hat{B}(k, h, GLS) - B(k, h) \parallel \xrightarrow{p} 0,$$

where

$$\underbrace{\hat{B}(k, h, GLS)}_{r \times kr} = (\hat{B}_1^{(h),GLS}, \dots, \hat{B}_k^{(h),GLS}), \qquad \underbrace{B(k, h)}_{r \times kr} = (B_1^{(h)}, \dots, B_k^{(h)}).$$

Theorem 3. *(Asymptotic Normality of Limiting Distribution) Under Assumption 3,*

$$\sqrt{T - k - H} l(k)' vec[\hat{B}(k, h, GLS) - B(k, h)] \xrightarrow{d} N(0, \Omega(k, h, GLS)),$$

where $l(k)$ *is a sequence of* $kr^2 \times 1$ *vectors such that*

$$0 < M_1 \leq \parallel l(k) \parallel^2 \leq M_2 < \infty.$$

and the explicit formula for $\Omega(k, h, GLS)$ *is defined in the next section.*

Remark. $l(k, H)$ is simply a Cramer-Wold device, which is used to show that any linear combinations of the parameters that satisfy the condition have asymptotically normal limiting distributions. Goncalves and

Kilian (2007) use these assumptions to show consistency and asymptotic normality of the VAR(∞) when there is conditional heteroskedasticity. These assumptions are more general versions of the ones used by Lewis and Reinsel (1985) and Jordà and Kozicki (2011) who show consistency and asymptotic normality of the VAR(∞) and the LP(∞) respectively in the i.i.d. case. These are sufficient conditions. Some of the proofs can be written under weaker conditions (e.g. Theorem 2 with $\frac{k^2}{T} \to 0$ and Theorem 3 with $\frac{k^3}{T} \to 0$), but for sake of brevity these will suffice.

Proof. See appendix. □

As noted earlier, the parameters used in the GLS correction are not known, and their uncertainty must be taken into account in order to do valid inference. To take into account the uncertainty in the generated regressors, frequentist can use bootstrapping, multi-step estimation (Murphy and Topel, 1985), or joint estimation (Greene, 2012). Bootstrapping and multistep estimation for reduced form and structural inference will be discussed in the next section.

3 LP GLS Estimation

Section 3 is broken up into 5 subsections. Subsections 3.1 and 3.2 discuss how to do reduced form inference using multistep estimation and bootstrapping respectively. Subsections 3.3 and 3.4 discuss how to do structural inference using multistep estimation and bootstrapping respectively. Section 5 discusses how structural identification is handled.

3.1 Multistep Estimation for Reduced Form Inference

The limiting distributions of the multistep estimators are derived in the appendix, so multistep estimation is straightforward. Researchers can simply apply the FGLS method of section 2.3 and use the multistep standard errors that are adjusted for uncertainty in the FGLS transformation. Define:

$$\underbrace{\hat{B}(k,h,GLS)}_{r\times kr} = (\hat{B}_1^{(h),GLS}, \ldots, \hat{B}_k^{(h),GLS}) = (T-k-H)^{-1}\sum_{t=k}^{T-h} \widetilde{y}_{t+h}^{(h)} X_{t,k}' \hat{\Gamma}_k^{-1},$$

$$\underbrace{\Gamma_k}_{kr\times kr} = E(X_{t,k}X'_{t,k}), \qquad \underbrace{\Gamma_{(m-n),k}}_{kr\times kr} = E(X_{m,k}X'_{n,k}), \qquad \underbrace{X_{t,k}}_{kr\times 1} = (y'_t,\ldots,y'_{t-k+1})'.$$

From Theorem 3,

$$(T-k-H)^{1/2}l(k)'vec[\hat{B}(k,h,GLS)-B(k,h)] \xrightarrow{d} N(0,\Omega(k,h,GLS)),$$

with the multistep covariance matrix

$$\Omega(k,h,GLS) = l(k)'\{E[(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+h}\varepsilon'_{t+h}(\Gamma_k^{-1}X_{t,k}\otimes I_r)'] + E[s_{k,h}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1}\varepsilon'_{t+1}(\Gamma_k^{-1}X_{t,k}\otimes I_r)'s'_{k,h}]$$

$$+E[(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+h}\varepsilon'_{t+h}(\Gamma_k^{-1}X_{t+h-1,k}\otimes I_r)'s'_{k,h}] + E[s_{k,h}(\Gamma_k^{-1}X_{t+h-1,k}\otimes I_r)\varepsilon_{t+h}\varepsilon'_{t+h}(\Gamma_k^{-1}X_{t,k}\otimes I_r)']\}l(k),$$

where

$$s_{k,h} = \Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big),$$

and $l(k)$ is the Cramer-Wold device defined earlier. Anywhere a parameter is not know, the estimated sample analogue would be use, i.e. $\hat{\Gamma}_k^{-1}$ for Γ_k^{-1}. For joint inference on reduced form impulse responses from multiple horizons, it is straightforward to deduce the following corollary from Theorem 3.

Corollary 1. *Under Assumption 3,*

$$l(k,H)'\begin{bmatrix}(T-k-H)^{1/2}vec[\hat{B}(k,H,GLS)-B(k,H)]\\ \vdots \\ (T-k-H)^{1/2}vec[\hat{B}(k,2,GLS)-B(k,2)]\\ (T-k-H)^{1/2}vec[\hat{B}(k,1,OLS)-B(k,1)]\end{bmatrix} \xrightarrow{d} N(0,V_{11}(k,H)),$$

where $l(k,H)$ *is a sequence of* $kr^2H\times 1$ *vectors such that* $0 < M_1 \leq \| l(k,H)\|^2 \leq M_2 < \infty$,

$$V_{11}(k,H) = \sum_{p=-\infty}^{\infty} cov(Score_t^{(H)}, Score_{t-p}^{(H)}),$$

$$Score_{t+H}^{(H)} = l(k,H)'\begin{bmatrix}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+H} + s_{k,H}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1}\\ \vdots \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+2} + s_{k,2}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1}\\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1}\end{bmatrix}.$$

Proof. See appendix. □

3.2 Bootstrapping Estimation for Reduced Form Inference

LP are typically bootstrapped using a block bootstrap since the errors are autocorrelated and it was assumed that the autocorrelation process is unknown. If one is only interested in functions of the reduced form estimates, block bootstrapping is not necessary. LP GLS can be implemented using a score wild bootstrap (Kline and Santos, 2012). More specifically, the wild bootstrap is applied to a "rearranged" scaled regression score. It follows from Corollary 1 that for finite H

$$plim\left(l(k,H)' \begin{bmatrix} \sqrt{T-k-H}vec[\hat{B}(k,H,GLS)-B(k,H)] \\ \vdots \\ \sqrt{T-k-H}vec[\hat{B}(k,2,GLS)-B(k,2)] \\ \sqrt{T-k-H}vec[\hat{B}(k,1,OLS)-B(k,1)] \end{bmatrix} \right)$$

$$= plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} Score_{t+H}^{(H)}\} = plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} j(k,H)' \begin{bmatrix} (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+H} \\ s_{k,H}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ \vdots \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+2} \\ s_{k,2}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \end{bmatrix}\}$$

$$= plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} Rscore_{t+1}^{(H)}\} = plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} j(k,H)' \begin{bmatrix} (\Gamma_k^{-1}X_{t-H+1,k}\otimes I_r)\varepsilon_{t+1} \\ s_{k,H}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ \vdots \\ (\Gamma_k^{-1}X_{t-1,k}\otimes I_r)\varepsilon_{t+1} \\ s_{k,2}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \end{bmatrix}\},$$

where $Score_{t+H}^{(H)}$ is the original scaled score. $Rscore_{t+1}^{(H)}$ is the rearranged scaled score where the time subscripts of ε line up, and $j(k,H)'$ is a $(2H-1)kr^2\times 1$ Cramer-Wold device such that

$$Score_{t+H}^{(H)} = l(k,H)' \begin{bmatrix} (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+H} + s_{k,H}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ \vdots \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+2} + s_{k,2}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \end{bmatrix} = j(k,H)' \begin{bmatrix} (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+H} \\ s_{k,H}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ \vdots \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+2} \\ s_{k,2}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \end{bmatrix}.$$

The score wild bootstrap is implemented by applying the wild bootstrap directly to the score (in our case the rearranged scaled score).[15] Since we're interested in doing inference on the limiting distribution of the scaled score or a function of it, the score wild bootstrap obviates the need to recompute the estimator in each bootstrap iteration. Let η be a zero mean unit variance random variable with finite fourth moments, then the sample analogue of the bootstrap rearranged scaled score

$$R\hat{Score}_t^{(H),*} = \hat{Rscore}_t^{(H)}\eta_t,$$

can be used to back out into bootstrap LP GLS estimates

$$l(k,H)'\begin{bmatrix} vec[\hat{B}^*(k,H,GLS) - \hat{B}(k,H,GLS)] \\ \vdots \\ vec[\hat{B}^*(k,2,GLS) - \hat{B}(k,2,GLS)] \\ vec[\hat{B}^*(k,1,OLS) - \hat{B}(k,1,OLS)] \end{bmatrix} = (T-k-H)^{-1}\sum_{t=k}^{T-H} R\hat{Score}_{t+1}^{(H),*}.$$

Conveniently, this can be done by just applying the wild bootstrap to $\{\hat{\varepsilon}_t\}_{t=k+1}^T$. To summarize,

1. Decide on the number of bootstrap draws, J, and the maximum number of impulse response horizons to be estimated, H.

2. Use the FGLS procedure described in section 2.3 to obtain estimates of $\{B_1^{(h)}, \ldots, B_k^{(h)}\}$ for each of the H horizons. The horizon 1 LP yields estimates of $\{\hat{\varepsilon}_{t,k}\}_{t=k+1}^T$.

3. For each bootstrap draw, J, generate zero mean unit variance random normal variables η to generate $\{\hat{\varepsilon}_{t,k}^*\}_{t=k+1}^T$, where

$$\hat{\varepsilon}_{t,k}^* = \hat{\varepsilon}_{t,k}\eta_t.$$

Then bootstrap draws can be created for each horizon by

$$\hat{B}^*(k,1,OLS) = \hat{B}(k,1,OLS) + (T-k-H)^{-1}\{\sum_{t=k}^{T-H} \hat{\varepsilon}_{t+1,k}^* X_{t,k}'\}\hat{\Gamma}_k^{-1},$$

$$\hat{B}^*(k,2,GLS) = \hat{B}(k,2,GLS)+(T-k-H)^{-1}\Big(\{\sum_{t=k}^{T-H} \hat{\varepsilon}_{t+2,k}^* X_{t,k}'\}\hat{\Gamma}_k^{-1}+\sum_{l=1}^{1}\hat{\Theta}_l^*\{\sum_{t=k}^{T-H}\hat{\varepsilon}_{t+1,k}^* X_{t,k}'\}\hat{\Gamma}_k^{-1}\hat{\Gamma}_{(h-l-1),k}\hat{\Gamma}_k^{-1}\Big),$$

$$\vdots$$

$$\hat{B}^*(k,h,GLS) = \hat{B}(k,h,GLS)+(T-k-H)^{-1}\Big(\{\sum_{t=k}^{T-H} \hat{\varepsilon}_{t+h,k}^* X_{t,k}'\}\hat{\Gamma}_k^{-1}+\sum_{l=1}^{h-1}\hat{\Theta}_l^*\{\sum_{t=k}^{T-H}\hat{\varepsilon}_{t+1,k}^* X_{t,k}'\}\hat{\Gamma}_k^{-1}\hat{\Gamma}_{(h-l-1),k}\hat{\Gamma}_k^{-1}\Big).$$

[15] Since LP are projecting more than one period forward, it necessitates that the scaled score be rearranged before applying the wild bootstrap in order for the bootstrap to be valid.

Denoting $\{\hat{B}^{*,j}(k,h,GLS)\}$ as jth bootstrap replication for the impulse responses, 95% confidence intervals can then be constructed by taking the 2.5% and 97.5% quantiles of the parameter(s) of interest. When implemented, η is Gaussian, but it need not be.[16] The bootstrap can also be implemented with bias adjustment if desired. The bias of the LP parameters can be calculated by applying the bias correction of West and Zhao (2019) to the FGLS LP models.[17]

Theorem 4. *(Asymptotic Validity of the Bootstrap) Under Assumptions 3,*

$$l(k,H)'\begin{bmatrix}(T-k-H)^{1/2}vec[\hat{B}^*(k,H,GLS)-\hat{B}(k,H,GLS)]\\ \vdots \\ (T-k-H)^{1/2}vec[\hat{B}^*(k,2,GLS)-\hat{B}(k,2,GLS)]\\ (T-k-H)^{1/2}vec[\hat{B}^*(k,1,OLS)-\hat{B}(k,1,OLS)]\end{bmatrix}\overset{d^*}{\to}N(0,V_{11}(k,H)).$$

Proof. See appendix. □

3.3 Multistep Estimation for Structural Inference

Following Brüggemann et al. (2016), a mixing condition is imposed for structural inference.

Assumption 4. *Let y_t satisfy Assumption 3. Assume that in addition, ε_t is strong (α) mixing, with $\alpha(m)$* of size $-4(\nu+1)/\nu$ *for some $\nu>0$, where $\alpha(m)=sup_{A\in\mathcal{F}^0_{-\infty},B\in\mathcal{F}^\infty_m}|P(A\cap B)-P(A)P(B)|$ for $m=1,2,\ldots$ denote the α mixing process of ε_t where $\mathcal{F}^0_{-\infty}=\sigma(\ldots,\varepsilon_{-2},\varepsilon_{-1},\varepsilon_0)$ and $\mathcal{F}^\infty_m=\sigma(\varepsilon_m,\varepsilon_{m+1},\ldots)$.*

Remark. Assumption 4 is slightly to somewhat stronger than the condition used in Brüggemann et al. (2016), and is probably a stronger condition than necessary. However, it allows for the use of mixingale inequalities which in turn allows for a straightforward proof using the mixingale CLT. The proof can probably be written using weaker conditions and following a proof similar to Theorem 3.1 in Brüggemann et al. (2016) .

[16] See Kline and Santos (2012) for a list of other variables that can be used.

[17] Whether or not one should bias adjust in practice is debatable. Bias adjustment can push sample estimates further away from the true values, and bias adjustment can increase the variance (Efron and Tibshirani, 1993).

Theorem 5. *Under Assumption 4,*

$$\begin{pmatrix} l(k,H)' \begin{bmatrix} \sqrt{T-k-H}vec[\hat{B}(k,H,GLS)-B(k,H)] \\ \vdots \\ \sqrt{T-k-H}vec[\hat{B}(k,2,GLS)-B(k,2)] \\ \sqrt{T-k-H}vec[\hat{B}(k,1,OLS)-B(k,1)] \end{bmatrix} \\ \sqrt{T-k-H}vech[\hat{\Sigma}-\Sigma] \end{pmatrix} \xrightarrow{d} N(0, V(k,H)),$$

where

$$V(k,H) = \begin{bmatrix} V_{11}(k,H) & V_{12}(k,H) \\ V_{21}(k,H) & V_{22} \end{bmatrix},$$

$$V_{22} = L_r'\{\sum_{p=-\infty}^{\infty} E(vec(\varepsilon_t\varepsilon_t'), vec(\varepsilon_{t-p}\varepsilon'_{t-p})') - vec(\Sigma)vec(\Sigma)'\}L_r,$$

$$V_{12}(k,H) = V_{21}(k,H)' = \sum_{p=-\infty}^{\infty} cov(Score_t^{(H)}, vec(\varepsilon_{t-p}\varepsilon'_{t-p} - \Sigma)'L_r).$$

Proof. See appendix. □

Even though the GLS correction eliminates the need for a HAC estimator for $V_{11}(k,H)$, HAC estimators are still needed for $V_{12}(k,H) = V_{21}(k,H)'$ and V_{22}. This is also true in the VAR case as noted in Brüggemann et al. (2016).[18] A HAC estimator combined with the delta method would lead to asymptotically valid joint inference. I do not explore which type of HAC estimators will perform best. I instead propose a block wild bootstrap sampling scheme due to its simplicity.

Notwithstanding, the Delta Method can be applied directly following Theorem 5.

Corollary 2. *(Delta Method for Structural Inference) It follows from Theorem 5 that for continuous and differentiable* $g(\cdot)$ *and if H is fixed,*

$$\sqrt{T-k-H}\left(g(vec[\hat{B}_1^{(H),GLS}],\ldots,vec[\hat{B}_1^{(1),GLS}],vech[\hat{\Sigma}]) - g(vec[\Theta_H],\ldots,vec[\Theta_1],vech[\Sigma])\right)$$

$$\xrightarrow{d} N(0, V(k,H)_{delta}),$$

[18]This is due to the more general assumption of conditional heteroskedasticity for the errors. In the i.i.d. case a HAC estimator would not be needed.

where $\triangledown$ is the gradient,

$$V(k,H)_{delta} = \triangledown g(vec[B_1^{(H)}],\ldots,vec[B_1^{(1)}],vech[\Sigma])'V_{marginal} \triangledown g(vec[B_1^{(H)}],\ldots,vec[B_1^{(1)}],vech[\Sigma]),$$

and where $V(k,H)_{marginal}$ is covariance matrix of

$$\sqrt{T-k-H}\begin{bmatrix} vec[\hat{B}_1^{(H),GLS}-\Theta_H] \\ \vdots \\ vec[\hat{B}_1^{(2),GLS}-\Theta_2] \\ vec\hat{B}_1^{(1),GLS}-\Theta_1] \\ vech[\hat{\Sigma}-\Sigma] \end{bmatrix},$$

and

$$V(k,H)_{marginal} = \sum_{m=-\infty}^{\infty} cov(r_t^{(H),marginal}, r_{t-m}^{(H),marginal}{}'),$$

$$r_{t+H}^{(H),marginal} = \begin{bmatrix} \underbrace{\begin{bmatrix} I_{r^2} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ I_{r^2} & 0 & \ldots & 0 \end{bmatrix}}_{r^2H\times r^2kH} \begin{bmatrix} (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+H} + s_{k,H}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ \vdots \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+2} + s_{k,2}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \end{bmatrix} \\ vech\left[(\varepsilon_{t+1}\varepsilon_{t+1}' - \Sigma)\right] \end{bmatrix}.$$

Proof. Simply a direct application of the Delta method. □

3.4 Bootstrap Estimation for Structural Inference

As noted in Brüggemann et al. (2016), structural inference using the standard wild bootstrap is invalid. The intuition behind their result is that if you apply a wild bootstrap to the errors, it cannot properly mimic the fourth order moments, and since fourth order moments are needed to calculate $V_{12}(k,H) = V_{21}(k,H)'$ and V_{22}, structural inference based on the wild bootstrap would be invalid. A standard block bootstrap on y could be used, but for LP if one wants to calculate a statistic which is a function of parameters from multiple horizons, i.e. a cumulative multiplier, it has the drawback that the blocks would need to be of length $H+k+\ell$. To appreciate this point, note that up to this point, when a researcher wants to conduct joint inference using LP OLS via a block bootstrap, they would first need to construct all possible $\{y_{t+H}, ..., y_{t-k+1}\}$ tuples to preserve the joint dependence. Then blocks of ℓ consecutive tuples are concatenated together to create bootstrap samples of the data which are then used to construct LP estimates. This is equivalent to sampling random blocks of size $H + k + \ell$ and concatenating them. To highlight why this is relevant in practice, take Ramey's (2016) application of Gertler and Karadi (2015). Impulse responses were estimated 48 horizons out and the regressions included 2 lags. If one wanted to calculate the cumulative impact of a monetary policy shock for the 48 horizons, $H = 48$ and $k = 2$, and the block length would be $50 + \ell$. If one were to instead estimate impulse horizons 16 horizons out, the block length would be $18+\ell$. Considering the bias variance tradeoff involved in choosing a block length, having the block length also depend on H and k is clearly an undesirable feature.

To overcome these issues, I propose a hybrid score block wild bootstrap. This bootstrap combines the score wild bootstrap, with the block wild bootstraps of Shao (2011), Yeh (1998). Brüggemann et al. (2016) argue that the block wild bootstrap leads to invalid inference, but that result is due to the way they implemented the bootstrap. The key is to recognize since we're not doing inference on the error terms, we don't need to bootstrap the error terms and generating the dependent variable like you would in a traditional wild bootstrap. Just like the score wild bootstrap, one can just bootstrap the rearranged scaled score. Note that

$$plim\begin{pmatrix} l(k,H)' \begin{bmatrix} \sqrt{T-k-H}vec[\hat{B}(k,H,GLS)-B(k,H)] \\ \vdots \\ \sqrt{T-k-H}vec[\hat{B}(k,2,GLS)-B(k,2)] \\ \sqrt{T-k-H}vec[\hat{B}(k,1,OLS)-B(k,1)] \end{bmatrix} \\ \sqrt{T-k-H}vech[\hat{\Sigma}-\Sigma] \end{pmatrix}$$

$$= plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} StrucScore_{t+H}^{(H)}\} = plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} \begin{bmatrix} j(k,H)' \begin{bmatrix} (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+H} \\ s_{k,H}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ \vdots \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+2} \\ s_{k,2}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1} \end{bmatrix} \\ vech(\varepsilon_{t+1}\varepsilon_{t+1}'-\Sigma) \end{bmatrix}\}$$

$$= plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} RStrucScore_{t+1}^{(H)}\} = plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H}\begin{bmatrix} j(k,H)'\begin{bmatrix}(\Gamma_k^{-1}X_{t-H+1,k}\otimes I_r)\varepsilon_{t+1}\\ s_{k,H}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1}\\ \vdots\\ (\Gamma_k^{-1}X_{t-1,k}\otimes I_r)\varepsilon_{t+1}\\ s_{k,2}(\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1}\\ (\Gamma_k^{-1}X_{t,k}\otimes I_r)\varepsilon_{t+1}\end{bmatrix}\\ vech(\varepsilon_{t+1}\varepsilon_{t+1}'-\Sigma)\end{bmatrix}\},$$

where $StrucScore_{t+H}^{(H)}$ is the original scaled "structural"score and the $RStrucScore_{t+1}^{(H)}$ is the rearranged scaled "structural" score where the $\varepsilon's$ line up. Applying the block wild bootstrap to the sample analogue of the rearranged scaled score leads to a valid bootstrap. For simplicity assume $T-k-H = N\ell$ where N is the number of blocks of length ℓ is the length of each block. Instead of multiplying the rearranged scaled scores by the i.i.d. $\{\eta_{k+1},\ldots,\eta_{T-H+1}\}$, yielding

$$RStru\hat{c}Score_t^{(H),*} = RStruc\hat{S}core_t^{(H)}\eta_t,$$

one would create

$$RStru\hat{c}Score_t^{(H),*} = RStruc\hat{S}core_t^{(H)}\eta_{[t/\ell]}.$$

That is, cut $\{RStruc\hat{S}core_{k+1}^{(H)},\ldots,RStruc\hat{S}core_{T-H+1}^{(H)}\}$ into N blocks of length ℓ and multiply the jth block by η_j to get the bootstrap sample $\{RStru\hat{c}Score_{k+1}^{(H),*},\ldots,RStru\hat{c}Score_{T-H+1}^{(H),*}\}$. This can be implemented simply by applying the block wild bootstrap to

$$\begin{bmatrix}\hat{\varepsilon}_{t,k}^{*}\\ (\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}^{'}-\hat{\Sigma})^{*}\end{bmatrix} = \begin{bmatrix}\hat{\varepsilon}_{t,k}\\ (\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}^{'}-\hat{\Sigma})\end{bmatrix}\eta_{[t/\ell]},$$

and replace the replacing the corresponding sample analogues with their bootstrap quantities. To summarize,

1. Decide on the number of bootstrap draws, J, and the maximum number of impulse response horizons to be estimated, H.

2. Use the FGLS procedure described in section 2.3 is used to obtain estimates of $\{B_1^{(h)},\ldots,B_k^{(h)}\}$ for each horizon the H horizons. The horizon 1 LP yields estimates of $\{\hat{\varepsilon}_{t,k},(\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}^{'}-\hat{\Sigma})\}_{t=k+1}^{T}$.

3. Divide $\{\hat{\varepsilon}_{t,k},(\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}^{'}-\hat{\Sigma})\}_{t=k+1}^{T}$, into N blocks of length ℓ. For each bootstrap draw, J, generate N zero mean unit variance random normal variables η, and multiply the jth block by η_j where

$$\begin{bmatrix}\hat{\varepsilon}_{t,k}^{*}\\ (\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}^{'}-\hat{\Sigma})^{*}\end{bmatrix} = \begin{bmatrix}\hat{\varepsilon}_{t,k}\\ (\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}^{'}-\hat{\Sigma})\end{bmatrix}\eta_{[t/\ell]}.$$

Then bootstrap draws can be created for each horizon by

$$\hat{B}^*(k,1,OLS) = \hat{B}(k,1,OLS) + (T-k-H)^{-1}\{\sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+1,k} X'_{t,k}\}\hat{\Gamma}_k^{-1},$$

$$\hat{B}^*(k,2,GLS) = \hat{B}(k,2,GLS) + (T-k-H)^{-1}\Big(\{\sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+2,k} X'_{t,k}\}\hat{\Gamma}_k^{-1} + \sum_{l=1}^{1} \hat{\Theta}^*_l \{\sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+1,k} X'_{t,k}\}\hat{\Gamma}_k^{-1}\hat{\Gamma}_{(h-l-1),k}\hat{\Gamma}_k^{-1}\Big),$$

$$\vdots$$

$$\hat{B}^*(k,h,GLS) = \hat{B}(k,h,GLS) + (T-k-H)^{-1}\Big(\{\sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+h,k} X'_{t,k}\}\hat{\Gamma}_k^{-1} + \sum_{l=1}^{h-1} \hat{\Theta}^*_l \{\sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+1,k} X'_{t,k}\}\hat{\Gamma}_k^{-1}\hat{\Gamma}_{(h-l-1),k}\hat{\Gamma}_k^{-1}\Big),$$

and

$$\hat{\Sigma}^* = \hat{\Sigma} + (T-k-H)^{-1} \sum_{t=k}^{T-H} (\hat{\varepsilon}_{t,k}\hat{\varepsilon}'_{t,k} - \hat{\Sigma})^*.$$

The draws of $\hat{\Sigma}^*$ are not guaranteed to be positive semi-definite. Whenever $\hat{\Sigma}^*$ is not positive semi-definite, the entire iteration is redone with new draws of η.[19] With the exception of using a block wild bootstrap scheme and calculating bootstrap estimates of Σ, these steps are essentially identical to the ones used section 3.2.

Since the $X's$ are fixed and the joint autocovariances of $\begin{bmatrix} \varepsilon^*_{t,k} \\ (\varepsilon_{t,k}\varepsilon'_{t,k} - \Sigma)^* \end{bmatrix}$ are preserved for ℓ lags, the score block wild bootstrap properly mimics the fourth order moments of ε needed to yield consistent estimates of $V_{12}(k,H) = V_{21}(k,H)'$ and V_{22} if $\ell \to \infty$ at a suitable rate. The bootstrap would yield consistent estimates of $V_{11}(k,H)$, whether or not ℓ grows.

Assumption 5. *Let y_t satisfy Assumption 4. Assume that in addition,*

$$\frac{k^8}{T} \to 0;\ T,k \to \infty.$$

$$\frac{\ell^8}{T} \to 0;\ T,\ell \to \infty.$$

Theorem 6. *(Validity of Bootstrap for Structural Inference) Under Assumption 6*

$$\left(l(k,H)' \begin{bmatrix} \sqrt{T-k-H}vec[\hat{B}^*(k,H,GLS) - \hat{B}(k,H,GLS)] \\ \vdots \\ \sqrt{T-k-H}vec[\hat{B}^*(k,2,GLS) - \hat{B}(k,2,GLS)] \\ \sqrt{T-k-H}vec[\hat{B}^*(k,1,OLS) - \hat{B}(k,1,OLS)] \\ \sqrt{T-k-H}vech[\hat{\Sigma}^* - \hat{\Sigma}] \end{bmatrix} \right) \xrightarrow{d^*} N(0,V(k,H)).$$

[19]In the empirical application, only a handful of iterations had to be redone.

Proof. See appendix. □

Since structural inference only involved the first and second moments of the rearranged scaled score, and since the rearranged scaled score has a mean of 0, applying the block wild bootstrap to the rearranged scaled score is valid since it preserves the first and second moments of the scaled score, which is all we need to in order to do structural inference. By bootstrapping the rearranged scaled score, the structural inference problems discussed in Brüggemann et al. (2016) are avoided entirely. Theorem 6 includes the sieve VAR as a special case, thus the bootstrap also provides a sieve extension of Brüggemann et al. (2016).[20]

There are no great rules of thumb for choosing ℓ in general. Since the block length involves a bias variance tradeoff with longer block lengths yielding less biased test statistics with larger variances and shorter block lengths yielding the opposite, data dependent rules such as those listed in Ch 7. of Lahiri (2003), but which optimize coverage, should to be developed in future research.

It follow directly from Theorem 6 that a bootstrap version of Corollary 2 exists.

Corollary 3. *(Delta Method for Bootstrap Structural Inference) It follows from Theorem 6, that for continuous and differentiable $g(\cdot)$ and if H is fixed,*

$$\sqrt{T-k-H}\left(g(vec[\hat{B}_1^{(H),GLS,*},\ldots,\hat{B}_1^{(1),GLS,*}],vech[\hat{\Sigma}^*]) - g(vec[\hat{B}_1^{(H),GLS},\ldots,\hat{B}_1^{(1),GLS}],vech[\hat{\Sigma}])\right)$$

$$\xrightarrow{d^*} N(0,V(k,H)_{delta}).$$

Proof. Simply a direct application of the Delta method. □

3.5 Structural Identification

This subsection briefly discusses structural identification in LP GLS. These techniques can be applied to both the bootstrapped LP and the analytical LP. In the analytical case, uncertainty bands can be constructed using the delta method. For an extensive review of structural identification in VARs and LP see Ramey (2016), and for an extensive treatment of identification in VARs and LP using external instruments see Stock and Watson (2018). For simplicity of exposition, assume the Wold representation can be written as a finite order VAR(k). Going back to the horizon 1 LP

$$y_{t+1} = B_1^{(1)}y_t + B_2^{(1)}y_{t-1} + \ldots + B_k^{(1)}y_{t-k+1} + \varepsilon_{t+1},$$

[20] It should be noted that Brüggemann et al. (2016) use a moving block bootstrap in a traditional recursive design VAR bootstrap setup, while here I apply a block wild bootstrap to the scaled score.

and let $\varepsilon_t = Rs_t$ where s_t is a vector of structural shocks and R is an invertible matrix. If R is known, after estimating $\{B_1^{(1)}, \ldots, B_1^{(h)}\}$, one can construct the structural impulse responses, $\{G^{(1)}, \ldots, G^{(h)}\}$, via Monte Carlo integration where $G^{(h)} = B_1^{(h)}R$. Typically R is not known but can be estimated, so Monte Carlo integration can still be applied. An example of R being estimated would be a triangular (recursive) ordering.[21] One would estimated horizon 1 LP, and then apply a recursive ordering to bootstrap draws of Σ to obtain draws of R, and then draws of $G^{(h)}$ can be constructed via $G^{(h)} = B_1^{(h)}R$.

It is often the case that the researcher may not know all of the identifying restrictions in R or may believe that R is not invertible, but the researcher has an instrument that they believe can trace out impulse responses of interest. The impulse responses of interest can instead be estimated by LP instrumental variable regressions (LP-IV). Stock and Watson (2018) show that in order for LP-IV to be valid, 3 conditions need to be satisfied. Decompose s_t into $s_{1,t}$ and $s_{2,t}$ where $s_{1,t}$ is the structural shock of interest at time t and $s_{2,t}$ represents all other structural shocks at time t. Let z_t be an instrument that the researcher believes can trace out the impulse responses of $s_{1,t}$. The instrument must satisfy the following three conditions

$$(i)\; E[s_{1,t}z_t] \neq 0,$$

$$(ii)\; E[s_{2,t}z_t] = 0,$$

$$(iii)\; E[s_{t+j}z_t] = 0\; for\; j \neq 0.$$

The first two conditions are just the standard relevance and exogeneity conditions for instrumental variable regression. The third condition is a lead-lag exogeneity condition, which guarantees that the instrument, z_t, is only identifying the impulse response of the shock $s_{1,t}$. If the third condition is not satisfied, then z_t will amalgamate the impulse responses at different horizons. It may be the case that these conditions are only satisfied after conditioning on suitable control variables (e.g. the lags of a VAR/horizon 1 LP).

Researchers typically estimate LP-IV via two-stage least squares (2SLS). For example, say I want to estimate the impulse response, $g^{(h)}$, the impact a shock to monetary policy has on output at horizon h. Let output be denoted as $output_t$ and the monetary policy variable mp_t. One can estimate LP-IV by estimating

$$output_{t+h} = g^{(h)}mp_t + control\, variables + error_{t+h}^{(h)},$$

via 2SLS and using z_t as an instrument for mp_t. Newey and West (1987) standard errors are typically used to account for autocorrelation, but as shown section 2, this ignores the increasing variance problem. The increasing variance problem can be particularly problematic with LP-IV, because the increasing variance can

[21] In the literature a triangular (recursive) ordering is often called a cholesky ordering because people often apply a cholesky decomposition to impose the ordering. It should be noted that the cholesky normalizes the variances of the structural shocks to unity. If one does not want to normalize the structural shocks, one can instead use the LDL decomposition to impose recursive the ordering.

weaken the strength of instrument for $h \geq 1$ if one is estimating a cumulative multiplier directly. Alternatively, the impulse responses of shocks to $s_{1,t}$ can be recovered if z_t is included as an endogenous variable in the system, and ordering it first in a recursive identification scheme (Plagborg-Møller and Wolf, 2019). Let $\mathring{y}_t = \begin{bmatrix} z_t \\ y_t \end{bmatrix}$ where y_t contains mp_t, $output_t$, and the control variables at time t, then the horizon 1 LP/VAR is

$$\mathring{y}_{t+1} = \mathring{B}_1^{(1)} \mathring{y}_t + \mathring{B}_2^{(1)} \mathring{y}_{t-1} + \ldots + \mathring{B}_k^{(1)} \mathring{y}_{t-k+1} + \mathring{\varepsilon}_{t+1}.$$

Since z_t is ordered first due to its exogeneity, the residual for the z_t equation, $\mathring{\varepsilon}_{1,t}$, will be able to trace out the structural impulse responses of interest.[22] Going back to the monetary policy example, the impulse response $g^{(h)}$ can be constructed as the ratio of the impulse response of $output_{t+h}$ to $\mathring{\varepsilon}_{1,t}$ divided the impulse response of mp_t to $\mathring{\varepsilon}_{1,t}$. Hence by imbedding z_t as an endogenous variable in the system and ordering it first in a recursive identification scheme, one can just estimate equation (2) via their preferred LP GLS method and construct the impulse responses of interest.

4 LP GLS and Relative Efficiency

To give a sense of potential efficiency gains of estimating LP via GLS, I will compare the asymptotic relative efficiency of the LP GLS estimator and the LP OLS estimator when the true model is an AR(1). Take the simple AR(1) model

$$y_{t+1} = a y_t + \varepsilon_{t+1},$$

where $|a| < 1$, $a \neq 0$, and ε_t is an i.i.d. error process with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma^2$. Define $\{b^{(1)}, \ldots, b^{(h)}\}$ as the LP impulse responses for the AR(1) model. For simplicity, assume the lag length is known. By Proposition 6 in appendix, the limiting distribution of the LP GLS impulse response at horizon h is

$$\sqrt{T}(\hat{b}^{(h),GLS} - a^h) \xrightarrow{d} N(0, [1 + (h^2 - 1)a^{2h-2}](1 - a^2)).$$

The limiting distribution of the LP OLS impulse response at horizon h is

$$\sqrt{T}(\hat{b}^{(h),OLS} - a^h) \xrightarrow{d} N(0, (1 - a^2)^{-1}[1 + a^2 - \{2h + 1\}a^{2h} + \{2h - 1\}a^{2h+2}]),$$

(Bhansali, 1997).

[22] Even if the control variables are exogenous to the system, any VARX can be written as a VAR with the exogenous variables ordered first in a block recursive scheme.

Theorem 7. *(FGLS Efficiency) Assume the true model is an AR(1) as specified above with* $|a| < 1$*,* $a \neq 0$*, and* ε_t *is an i.i.d. error process with* $E(\varepsilon_t) = 0$ *and* $var(\varepsilon_t) = \sigma^2$*. Then,*

$$plim(var(\sqrt{T}(\hat{b}^{(h),GLS} - a^h))) \leq plim(var(\sqrt{T}(\hat{b}^{(h),OLS} - a^h))).$$

Proof. See Appendix. □

The relative efficiency between the LP GLS and LP impulse responses (given by the ratio of the variances),

$$\frac{[1 + (h^2 - 1)a^{2h-2}](1 - a^2)^2}{[1 + a^2 - \{2h + 1\}a^{2h} + \{2h - 1\}a^{2h+2}]},$$

determines how much more efficient one specification is relative to another. Note that the relative efficiency not only depends on the persistence, a, but on the horizon as well. Table 3 presents the relative efficiency between the LP GLS and LP OLS impulse responses for different values of a. The gains from LP GLS can be large but they are not necessarily monotonic. This is because if the persistence is not that high, the impulse responses decay to zero quickly making the variance of the impulse responses small, and the gains from correcting for autocorrelation are not as large.

Table 3: Relative Efficiency of LP (GLS) to LP (OLS)

Autocorrelation	Horizons					
Coefficient	3	5	10	20	30	40
$a = .99$	.993	.979	.945	.88	.818	.759
$a = .975$	.983	.948	.864	.713	.580	.464
$a = .95$	.966	.896	.735	.475	.288	.165
$a = .9$	.931	.792	.508	.179	.061	.029
$a = .75$	.827	.53	.195	.123	.123	.123
$a = .5$	.727	.496	.45	.45	.45	.45
$a = .25$	.854	.828	.827	.827	.827	.827
$a = .1$	.971	.97	.97	.97	.97	.97
$a = .01$	1	1	1	1	1	1

The efficiency gains of estimating LP GLS do not stop there. It turns out that when the true model is an AR(1) and the system is persistent enough, LP GLS can be approximately as efficient as the AR(1). Let $\hat{a}$ be the OLS estimate, the OLS estimate of a has the limiting distribution

$$\sqrt{T}(\hat{a} - a) \xrightarrow{d} N(0, 1 - a^2).$$

By the delta method, the horizon h impulse response has the limiting distribution

$$\sqrt{T}(\hat{a}^h - a^h) \xrightarrow{d} N(0, h^2 a^{2h-2}(1 - a^2)).$$

The asymptotic relative efficiency between the AR and LP GLS impulse responses

$$\frac{h^2 a^{2h-2}}{h^2 a^{2h-2} + (1 - a^{2h-2})},$$

determines which specification is more efficient. Since the true model is an AR(1), if the errors are normal, the AR(1) model will be asymptotically more efficient due to the Cramer-Rao lower bound (Bhansali, 1997). Table 4 presents the relative efficiency between the AR and LP GLS impulse responses for different values of a.

Table 4: Asymptotic Relative Efficiency of AR to LP (GLS)

Horizons	5	10	20	30	40
$a = .99$	.997	.998	.999	.999	.999
$a = .975$	.991	.994	.996	.996	.996
$a = .95$	.980	.985	.985	.980	.968
$a = .9$	.95	.946	.881	.667	.302
$a = .75$	.736	.362	.007	0	0
$a = .5$	0	0	0	0	0

If the data is persistent enough, the LP GLS impulse responses have approximately the same variance for horizons relevant in macro. For example, the economics profession has still not determined if GDP has a unit root or not. Assume that GDP is stationary but highly persistent with an AR(1) coefficient of .99. In this case, the AR(1) impulse responses has approximately the same variance for at least the first 40 horizons. Müller (2014) estimates the AR(1) coefficient for unemployment to be approximately .973. This would lead to the AR(1) impulse responses having approximately the same variance for at least the first 40 horizons. Other important macroeconomic variables such as inflation and the 3 month interest rate and most macro aggregates are also highly persistent and would display similar results. It is not until the AR(1) coefficient is .9 that you can see a notable difference over the first 40 horizons, and even then it is not until about 20 or so horizons out. For less persistent values of a the AR(1) dominates.

When the true model is a multivariate VAR things become more complicated. Efficiency still depends on the horizon and persistence, but because persistence can vary across the equations in the system, LP GLS could be approximately as efficient for some impulse responses and much less efficient for others. To see why, let us return to the VAR(1) model

$$y_{t+1} = A_1 y_t + \varepsilon_{t+1}.$$

Take the eigenvalue decomposition of $A_1 = E\Lambda_1 E^{-1}$, where $\Lambda_1 = diag(\lambda_1, \ldots, \lambda_r)$ is the diagonal matrix of eigenvalues which are assumed to be distinct and nonzero. E is the corresponding eigenmatrix. As a result $A_1^h = E\Lambda_1^h E^{-1}$. Define $w_t = E^{-1}y_t$ and $\epsilon_t = E^{-1}\varepsilon_t$. For simplicity assume E is known. This implies the VAR can be transformed into

$$w_{t+1} = \Lambda_1 w_t + \epsilon_{t+1},$$

which will be called the transformed model. Consequently,

$$w_{t+h} = \Lambda_1^h w_t + \Lambda_1^{h-1}\eta_{t+1} + \ldots + \Lambda_1 \eta_{t+h-1} + \epsilon_{t+h}.$$

Since Λ_1 is diagonal, each equation in the transformed VAR(1) is an AR(1) model. Therefore the results derived earlier in this subsection for the AR(1) model apply.

More generally, it should be noted that: 1) all of the variation in $\hat{A}_1^h$ is emanating from $\hat{\Lambda}_1^h$ (it was assumed E is known);

2)

$$var(\sqrt{T}vec[\hat{A}_1^h - A_1^h]) = var(\sqrt{T}vec[E(\hat{\Lambda}_1^h - \Lambda_1^h)E^{-1}]) = [E'^{-1} \otimes E]var(\sqrt{T}vec[\hat{\Lambda}_1^h - \Lambda_1^h])[E'^{-1} \otimes E]'.$$

Hence, the efficiency gains of impulse responses estimated via LP GLS impulse for a particular horizon depends on the relative efficiency of the eigenvalues, and how much an eigenvalue contributes to the variance of an impulse response. So if A_1 contains different eigenvalues, the eigenmatrices would determine how much the variance of an eigenvalue contributes to the variance of an impulse responses in the untransformed model and hence determine the relative efficiency of LP GLS impulse response to the VAR impulse responses.[23] Essentially, the efficiency gains of the VAR come from the less persistent components. Depending on the persistent eigenvalues and how much they contribute to the variance of the impulse responses, it is possible for LP GLS to be approximately as efficient as the VAR. Whether LP GLS impulse responses would be approximately as efficient would depend on the true DGP, the persistence of the system, the dependence structure of the variables, and the horizon. In other words, it would be specific to the situation. It should be noted that persistent eigenvalues would not necessarily get the most weight.[24] It follows that LP can be

[23]Note that if the correlation matrix of $var(\sqrt{T}vec[\hat{\Lambda}_1^h - \Lambda_1^h])$ differs across estimation methods, then the correlation matrix also determines relative efficiency for the untransformed model impulse responses.

[24]In order to apply that argument, A would have to be positive definite.

much more efficient than previously believed.

5 Monte Carlo Evidence

In this section, I present a battery of Monte Carlo evidence of the finite sample properties of the LP GLS bootstrap and the multistep (analytical) LP GLS estimator. I compare the following estimators:

- LP GLS bootstrap (LP GLS Boot),
- Bias-adjusted LP GLS bootstrap (LP GLS Boot BA),
- Analytical LP GLS estimator (LP GLS),
- Analytical VAR estimator (VAR),
- Bias-adjusted VAR bootstrap (Kilian, 1998) (VAR Boot BA),
- LP OLS with equal-weighted cosine HAC standard errors (Lazarus et al., 2018) (LP OLS).

The abbreviations in the parentheses are what the estimators are referred to in the figures. In summary, I find that the LP GLS bootstrap estimators minimizes downside risk. The VAR had the shortest confidence interval on average, but coverage can vary widely depending on the DGP. LP OLS typically had at least decent coverage, but coverage typically did not exceed that of its GLS counterparts, and it was relatively inefficient compared to the GLS estimators. LP GLS bootstrap coverage doesn't drop below approximately 86% and is often in the low to mid 90's. The analytical LP GLS estimator does not perform as well as the LP GLS bootstraps.

Unless stated otherwise, all simulations use a sample size of 250, which is representative of a quarterly data set dating back to 1960. Even though the most prominent macro variables such as GDP, inflation, and unemployment date back to at least 1948, many do not date back that far. The comprehensive McCracken and Ng (2016) data set goes back to 1959 for quarterly and monthly data of, so a sample size of 250 would be representative lower bound for practically most quarterly macroeconomic variables of interest.[25]

All of the methods use the same lag length for each simulation. The LP GLS methods requires that the VAR residuals are white noise. For the simulations, lag lengths are chosen using a lag length criteria (e.g. AIC, BIC, HQIC) and then the VAR residuals are tested for autocorrelation using the Ljung-Box Q-test. The baseline lag length used is AIC, but results are not sensitive to other choices. If the null of white noise is rejected, a lag is added, the model is reestimated, and the new residuals are tested for autocorrelation. This

[25]If one is doing LP-IV and the instrument is not available for the same sample period as the other variables in the system, then estimation can only be done for the sample period that the instrument is available (Stock and Watson, 2018). Fortunately, the major instruments used in macroeconomics are available for at least 200 observations (Ramey, 2016).

process is repeated until the null of white noise is not rejected for the VAR residuals. This lag length is then used for all estimation procedures (including the non LP GLS procedures).

Simulations were conducted 1,000 times, and bootstraps have 5,000 replications each. Coverage and average length for 95% confidence intervals of the reduced form impulse responses are calculated. That is, for each simulation, I estimated the model for each desired horizon using all of the estimation methods and then check if the 95% confidence intervals contain the true impulse response. I then calculate the probability that the 95% confidence interval contains the true impulse response over the Monte Carlo simulations which gives me the coverage for each method and horizon. For each simulation draw, I also save the length of the 95% interval for the the different methods for each horizon. The lengths are then averaged over each Monte Carlo simulation for each method and horizon to get the respective average length of the 95% confidence intervals for each method and horizon. Unless stated otherwise, 15 horizons are analyzed, which would be representative of analyzing four years of impulse responses for quarterly data. The selected results for the Monte Carlos are presented in figures 4-9 in the appendix.

I redo several of the Monte Carlos in Kilian and Kim (2011). I start with the following VAR (1):

$$y_{t+1} = A_1 y_t + \varepsilon_{t+1},$$

where

$$A_1 = \begin{bmatrix} A_{11} & 0 \\ .5 & .5 \end{bmatrix}, A_{11} \in \{.5, .9, .97\}, and\, \varepsilon_t \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}).$$

Despite the model being simplistic, it has been a benchmark in the literature. For this DGP, the bias-adjusted VAR bootstrap performs the best overall. The LP GLS bootstraps also perform well, but they're not as efficient, and for the persistent eigenvalues, the coverage is slightly worse than the bias-adjusted VAR. The analytical LP GLS, LP OLS, and the analytical VAR performance deteriorates the most when the eigenvalues are more persistent. Despite this, all of the estimators have coverage of at least 80% for all horizons. Select results are presented in Figure 4.

The second Monte Carlo I replicate from Kilian and Kim (2011) is the following ARMA(1,1),

$$y_{t+1} = .9y_t + \varepsilon_{t+1} + m\varepsilon_t,$$

where $m \in \{0, .25, .5, .75\}$, and $\varepsilon_t \sim N(0, 1)$. The bias-adjusted VAR bootstrap performs the best, but LP GLS bootstraps perform well, with coverage of at least approx. 90% at all horizons. Again the GLS estimators are more efficient than the LP OLS estimator. LP OLS has slightly better coverage than the analytical LP GLS estimators, but not the bootstraps. In terms of coverage, the analytical VAR performs the worst out of them all, and appears to have the shortest average length because it's underestimating uncertainty. Select results can be found in Figure 5.

The final DGP I replicate from Kilian and Kim (2011) is their empirically calibrated VARMA(1,1) based on quarterly investment growth, inflation, and the commercial paper rate.

$$y_{t+1} = A_1 y_t + \varepsilon_{t+1} + M_1 \varepsilon_t,$$

where

$$A_1 = \begin{bmatrix} .5417 & -.1971 & -.9395 \\ .04 & .9677 & .0323 \\ -.0015 & .0829 & .808 \end{bmatrix}, M_1 = \begin{bmatrix} -.1428 & -1.5133 & -.7053 \\ -.0202 & .0309 & .1561 \\ .0227 & .1178 & -.0153 \end{bmatrix}, P = \begin{bmatrix} 9.2352 & 0 & 0 \\ -1.4343 & 3.607 & 0 \\ -.7756 & 1.2296 & 2.7555 \end{bmatrix},$$

and $\varepsilon_t \sim N(0, PP')$. For the VARMA(1,1) only the analytic VAR doesn't have coverage of at least 90% for all parameters. Coverage drops as low as 79%, for one response, but 5 of the 9 responses are at least 90% for all horizons. The average length for the LP GLS estimators are quite a bit shorter than the LP OLS estimator. Select results are presented in Figure 6.

In summary so far, the LP GLS bootstraps performed well, but not as well as the bias-adjusted VAR. The LP analytical estimator tended to perform better than the analytical VAR and LP OLS, but all three of these estimators had coverage that tended to fall off more when estimators had more persistent eigenvalues. Relative to Kilian and Kim (2011), the LP estimators used in these Monte Carlos performed much better. The poor performance of LP estimators in Kilian and Kim (2011) was due to two reasons. First, Kilian and Kim (2011) were limited to using the LP estimators of the time. They used a block bootstrap LP estimator and an LP OLS estimator with Newey-West standard errors. The drawbacks of using a standard block bootstrap for LP was discussed in Section 3.4. Newey-West standard errors have well known coverage distortions (Müller, 2014). The equal-weighted cosine HAC standard errors of Lazarus et al. (2018) is a much better alternative. Monte Carlos with Newey-West standard errors are not included, but preliminary Monte Carlo evidence corroborates the evidence that equal-weighted cosine HAC standard errors are a better alternative relative to standard Newey-West. Second, not explicitly modeling for autocorrelation and doing a GLS correction appears to have negatively affected LP performance.

The Kilian and Kim (2011) Monte Carlos provide evidence that if the VAR can do a good job approximating the DGP, it will tend to have better performance than LP. If VARs cannot do a good job approximating the DGP, it can drastically impact inference. To illustrate, I return to the MA(35) from section 2.1. The results are presented in Figure 7. Here, the LP estimators have approximately 95% coverage for all horizons. The VAR estimators, on the other hand, have about 90% coverage for the first 1 or 2 horizons before coverage drops precipitously. Even though the MA(35) is not empirically calibrated, it may not be too dissimilar from what can occur in practice.

Next I'll present evidence for two empirically calibrated Monte Carlos for a fiscal VAR and technology shock VAR. But before I get into the details, it's important that I first discuss a potential shortcoming in the way we calibrate empirical models in the impulse response literature. Jordà (2005) and Kilian and Kim (2011) each empirically calibrated a VAR(12) in the spirit of Christiano et al. (1999) to give an empirically relevant Monte Carlo that would give a gauge of what is happening in practice. The problem with empirically calibrated VARs is that if the the model used to generate the data suffers from truncation bias, the Monte Carlo simulation may mask the actual amount of truncation bias that could occur in practice. For example, take the MA(35) from section 2.1. Say the VAR that corresponds to the mean impulse response for the VAR is used to generate data in a Monte Carlo. If the coverage in that Monte Carlo is great, one might conclude that the truncation bias is not an issue and in practice truncation bias is unlikely to be a problem. Obviously we know truncation bias can be an issue because we know what the true impulse responses are. In practice when empirically calibrating a Monte Carlo, the results of the Monte Carlo are only as good as the calibration, and in practice we don't know how good the calibration is.[26] To protect against this problem, I estimate empirically calibrated VARs with lag lengths longer than what is typically used in practice in order to protect against truncation bias. [27]

I empirically calibrate 2 different quarterly VARs. For the quarterly data sets, researchers will typically not include more than one or two years worth of lags, so I estimate a VAR(16). I assume the errors are normally distributed. The first empirically calibrated VAR is a fiscal VAR that includes growth rates of real GDP per capita and real spending per capita, which are the baseline variables used in fiscal multiplier analysis (Ramey and Zubairy, 2018). The data runs from 1947Q2-2019Q4. The results are presented in Figure 8. For the bias-adjusted VAR, coverage drops below 80% after approximately 7 horizons out for at least one response, and by 15 horizons out, 2 responses have coverage below 20%, and 3 below 60%. For the analytical VAR, coverage rates respond in a similar manner, with 2 of the four responses having coverage below 15% at 15 horizons out, and 3 of the four responses having coverage below 50%. All of the LP estimators, on the other hand, have coverage of approximately 90% or higher at all horizons. The results are only somewhat sensitive to using the conservative lag length suggestion of Kilian and Lütkepohl (2017) and including one or two years worth of lags. Though not included in the figures, including 2 years worth of lags slightly improves the coverage of the VAR estimators, but the truncation bias still causes massive coverage deterioration at higher horizons, with coverage rates still dropping below 12% for 2 of the 4 responses and below 71% for 3 of the 4 responses.

The second empirical Monte Carlo is a technology VAR that includes growth rates of labor productivity, real GDP per capita, real stock prices per capita, and total factor productivity. These are the baseline vari-

[26] This argument also applies to the calibrated VARMA(1,1) presented earlier.

[27] Alternatively, I could also generate the model using LP to estimate the Wold coefficients up to q horizons out and generate data based off of that MA(q). Unfortunately, it would be more cumbersome since it would require more choices in the setup and hence more robustness checks.

ables used in Ramey (2016). The data runs from 1947Q2-2015Q4. Again, I estimate a VAR(16) and use that to generate the data. Select results are presented in Figure 9. With the exception of the first horizon, the LP estimators have coverage of at least 90% for essentially all response at all horizons, with the occasional response dipping into the mid to high 80s. Since the horizon 1 LP is a VAR, the horizon 1 impulse responses had a sub par performance for all of the estimators. Like for the fiscal VAR, the VAR estimators have serious coverage distortions across most horizons. Excluding the horizon 1 impulse responses, responses had coverage fall below 80% as early as 5 horizons out. By horizon 15, most responses had coverage rates drop below 50%, with some dropping as low 11% for the bias-adjusted VAR bootstrap and 9% for the analytical VAR. Including 2 years worth of lags did not substantially improve the performance of the VAR estimators.

The coverage distortion results for the VAR should be alarming. For empirically calibrated VAR DGP, the results can be thought of as an upper bound, in the sense that the true model is probably not a VAR(16), so information gleamed from these Monte Carlo's are limited to how good the approximations are, and the truncation bias problem with the VARs can be much worse. Even though LPs have good coverage, the truncation bias can be worse for them as well, but probably not as bad as the VAR.

In summary, I find that LP GLS does the best at balancing efficiency while still have proper nominal coverage. LP GLS bootstrap estimators in general had coverage of at least 89%. They were generally more efficient than the LP OLS estimator (unless the LP OLS estimator was underestimating uncertainty).[28] The analytical LP GLS estimator did not perform as well as its bootstrap counterparts, but it tended to perform better than the LP OLS. LP OLS typically had decent coverage, but it was clearly the most inefficient out of all the estimators (unless it was underestimating uncertainty). The VAR is the most efficient out of the estimators. When the VAR does a good job of approximating the DGP, coverage is at or near nominal level, but coverage can vary widely depending on the DGP. As highlighted in the empirically calibrated Monte Carlos, it can easily be the case that the VAR has truncation bias issues that leads to poor coverage rates.

6 Structural Breaks and Time-Varying Parameter LP

It is worth reiterating that the GLS procedure presented in Section 2 and the consistency and asymptotic normality of the procedure assumes stationarity.[29] Nonstationarity can be caused by unit roots or structural

[28]There were cases where the OLS estimator had shorter confidence intervals, but the coverage was below the nominal level.

[29]If unit roots are the cause, consistency can still hold if the errors have enough moments (Jordà, 2009), so the procedure would still eliminate autocorrelation, but asymptotic normality of the results could break down in general. That being said, inference would be valid in the presence of unit roots in certain cases (see Jordà (2009) Proposition 4 for details). Montiel Olea and Plagborg-Møller (2020) show that lag augmentation with LP can handle unit roots more generally.

breaks. When nonstationarity is caused by structural breaks, all methods will break down if they do not properly take into account change(s) in the parameters. Stationarity guarantees that the model has a linear time-invariant VMA representation. If the data are not stationary and structural breaks are the cause, then the procedure may not eliminate autocorrelation. To understand why it matters if structural breaks are present, note that if the data are not stationary, it is possible for the estimated horizon 1 LP residuals to be uncorrelated since the VAR can still produce reasonable one-step ahead forecasts when the model is misspecified (Jordà, 2005). A "Wold representation" exists for nonstationary data, but the impulse responses for this VMA representation are allowed to be time dependent (Granger and Newbold, 1977, Priestley, 1988).[30] Assuming there is no deterministic component, any time series process can be written as

$$y_t = \varepsilon_t + \sum_{i=1}^{\infty} \Theta_{i,t}\varepsilon_{t-i},$$

where $\Theta_{i,t}$ is now indexed by the horizon and time period and $var(\varepsilon_t) = \Sigma_t$. Using recursive substitution, the time dependent Wold representation can be written as a time dependent VAR or a time dependent LP. It can be shown that a time dependent version of Theorem 1 exists. The horizon h time dependent LP is

$$y_{t+h} = B_{1,t}^{(h)} y_t + B_{2,t}^{(h)} y_{t-1} + \ldots + e_{t+h}^{(h)},$$

where

$$e_{t+h}^{(h)} = \Theta_{h-1,t}\varepsilon_{t+1} + \ldots + \Theta_{1,t}\varepsilon_{t+h-1} + \varepsilon_{t+h},$$

$$B_{1,t}^{(h)} = \Theta_{h,t}.$$

If impulse responses are time dependent at higher horizons, but a time invariant version of LP GLS is applied, autocorrelation may not be eliminated at these horizons because the time-invariant LP are misspecified. In other words, if the data are nonstationary and the nonstationarity is caused by structural breaks, the time invariant version of LP GLS may not eliminate autocorrelation in the residuals since the estimates of the impulse responses may not be consistent. In this sense, LP GLS is a type of general misspecification test, because if one had estimated LP using OLS and HAC standard errors, the autocorrelation in the residuals would not hint toward potential misspecification since the residuals are inherently autocorrelated.

Just like the time invariant case, k can be infinite in population but will be truncated to a finite value in finite samples. Similarly to the time-invariant transformation, one can do a GLS transformation $\widetilde{y}_{t+h}^{(h)} =$

[30] Nonstationarity in economics typically refers to explosive behavior (e.g. unit roots), but nonstationarity is more general and refers to a distribution that does not have a constant mean and/or variance over time. Depending on the true model, differencing may not make the data stationarity (Leybourne et al., 1996, Priestley, 1988).

$y_{t+h} - \hat{B}_{1,t}^{(h-1)}\hat{\varepsilon}_{t+1,k} - \ldots - \hat{B}_{1,t}^{(1)}\hat{\varepsilon}_{t+h-1,k}$. Then one can estimate horizon h via the following equation:

$$\widetilde{y}_{t+h}^{(h)} = B_{1,t}^{(h)} y_t + B_{2,t}^{(h)} y_{t-1} + \ldots + B_{k,t}^{(h)} y_{t-k+1} + \tilde{u}_{t+h,k}^{(h)}.$$

Estimation is carried out in the same way as in the time-invariant case, except the models are being estimated with time-varying parameters.

Just like a static LP model can be less sensitive to model misspecification than a static VAR, a time-varying parameter LP model can be less sensitive to model misspecification than a time-varying parameter VAR. If the true model is time varying, then the misspecification of the VAR can extend to the time variation as well. Due to the iterative nature of the VAR, misspecification in time variation would be compounded in the construction of the impulse responses alongside other misspecifications in the VAR. Time-varying parameter LP, however, allow for the amount and nature of time variation to change across horizons. Since time-varying parameter models can also approximate any non-linear model, time-varying parameter LP can do a to better job capture the time variation in the impulse responses at each horizon.

As noted in Granger and Newbold (1977), macro data encountered in practice are unlikely to be stationary, implying that the Wold representation may be time dependent. If the impulse responses of the Wold representation are time dependent, since time-varying parameter models can approximate any form of non-linearity (Granger, 2008), a time varying version of LP GLS may be applied. The time-varying parameter version of the above GLS procedure presented in section 2 will be able to eliminate autocorrelation as long as the parameter changes are not so violent that a time-varying parameter model cannot track them. All else equal, the more adaptive the time-varying parameter model, the better the time-varying parameter model will be able to track changes and the better the approximation.[31] If the nature of the time dependence is known, that is, the researcher knows when the structural breaks occur or the nature of the time variation, then that specific time dependent model can be applied to the LP GLS procedure. The conditions under which this procedure is consistent and asymptotically normal, as well as the proofs for consistency and asymptotic normality could vary depending on the type of time-dependent model being used and the estimation procedure and is therefore left for future research.

7 Application to Gertler and Karadi (2015)

To illustrate how results can change when using LP GLS, I redo the analysis Gertler and Karadi (2015). One of the reasons why the Gertler and Karadi (2015) analysis is so interesting is because there has been

[31] Baumeister and Peersman (2012) show via Monte Carlo simulations that time-varying parameter models are able to capture discrete breaks in a satisfactory manner should they occur.

tension in the literature about the results. Using a Proxy Structural SVAR, Gertler and Karadi (2015) find an increase in the one-year Treasury rate leads to a decrease in both industrial production and CPI. Several papers have challenged different aspects of Gertler and Karadi's (2015) methodology and implementation (Ramey, 2016, Brüggemann et al., 2016, Stock and Watson, 2018, Jentsch and Lunsford, 2019a).[32] Using LP-IV with Newey-West standard errors, Ramey (2016) finds that an increase in the one-year Treasury leads to a significant decrease in CPI but a significant increase in industrial production. Output does not respond for at least a year, and inflation does not respond for at least 30 months. Both output and CPI respond more slowly relative to the Gertler and Karadi (2015) results. The Gertler and Karadi (2015) and the Ramey (2016) results are presented in Figure 2.

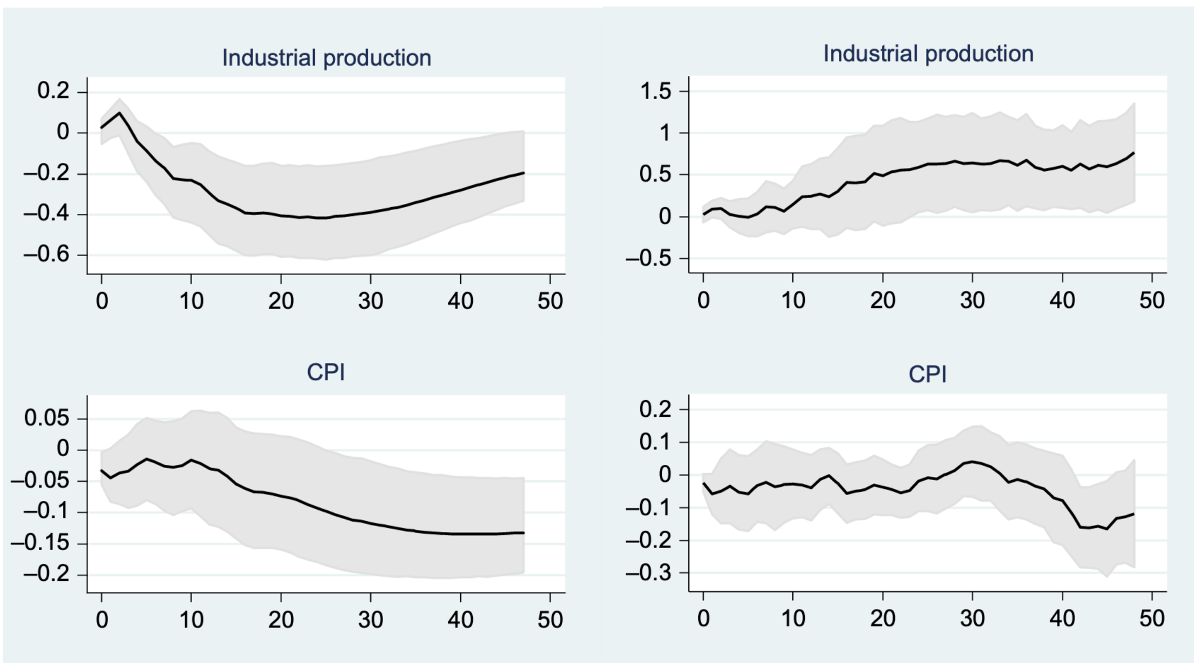

Figure 2: Left: Proxy SVAR results with 90% bootstrap confidence intervals. Right: LP-IV results with 90% Newey-West confidence intervals.

The system of macroeconomic variables includes output (log of industrial production), inflation (log of CPI), the one-year Treasury yield, the excess bond premium spread, and a high frequency identification instrument. I do baseline analyses using 2 different instruments: the surprise to the one-month ahead fed funds futures and the surprise to the three-month ahead fed funds futures. When using the the three-month surprise the data spans 1990M1-2012M6 and spans from 1988M10-2012M6 when using the one-month surprise. I estimate the systems using the LP GLS bootstrap (without bias adjustment) based on 20,000

[32] Jentsch and Lunsford (2019a,b) prove the invalidity of Gertler and Karadi's (2015) Proxy SVAR bootstrap and show that it can dramatically underestimate uncertainty. Ramey (2016), Stock and Watson (2018) point out that the high frequency identification instruments are correlated, thus violating the lead-lag exogeneity condition discussed in section 3.5.

replications. I use 12 lags in estimation as in Gertler and Karadi (2015) and a block length of 20, but my results are qualitatively similar across alternative choices for these parameters. The structural impulse responses are constructed as discussed in Subsection 3.5.

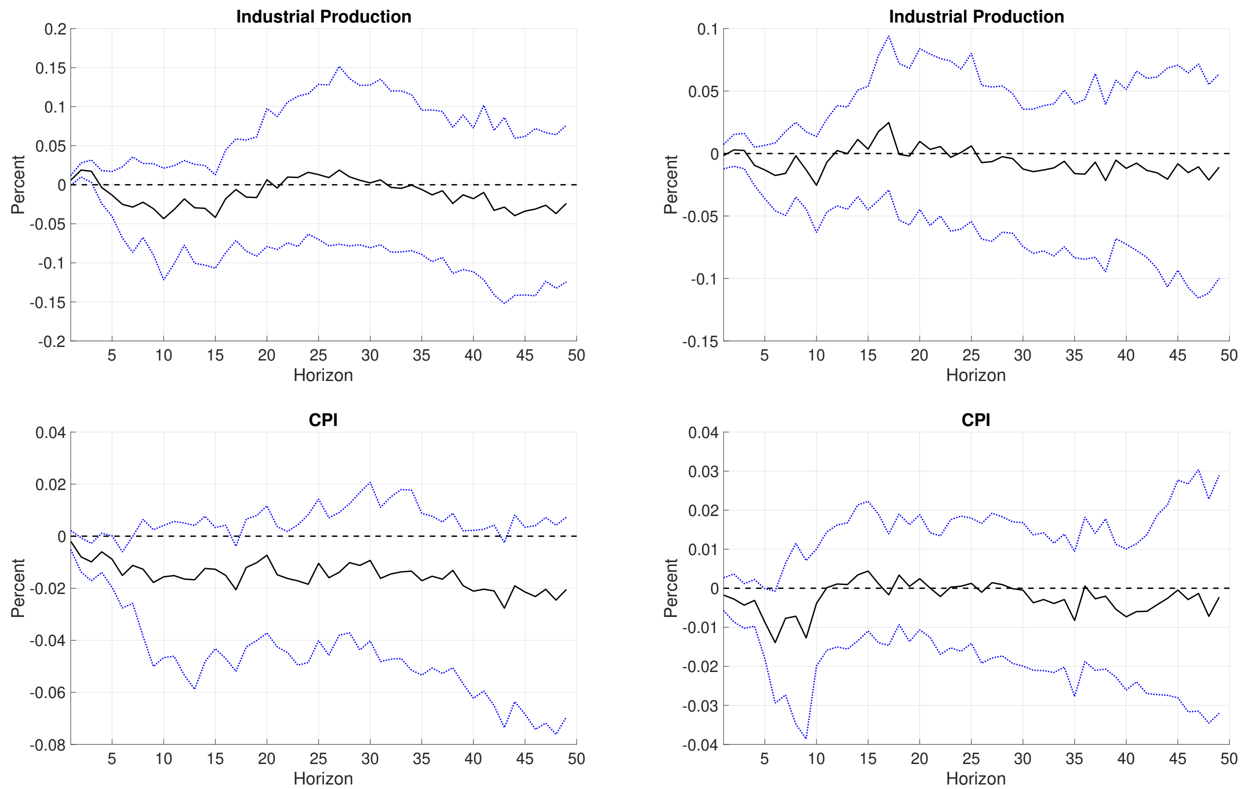

Figure 3: LP GLS bootstrap results with 90% confidence intervals. Left: Three-month fed funds futures surprise as instrument. Right: One-month fed funds futures surprise as instrument.

The LP GLS bootstrap results for output and inflation are presented in Figures 3. In general, I cannot reject the null hypothesis that a change in the one-year Treasury has no impact on output or inflation during the first four years. There is, however, evidence of an output puzzle at horizon 1 when using the 3 month surprise as the instrument.[33] Relative to the VAR, the LP GLS results are more uncertain and less pronounced. Relative to LP OLS, the puzzling output result mostly go away for the LP GLS. The LP GLS results are generally not significant while the LP OLS results are at higher horizons.[34] The results are also not sensitive to using growth rates of industrial production and CPI instead of log levels.[35] F-test indicates the instrument is relevant, with the bootstrap LP FGLS having a mean F-statistic of approximately 25 and 32 for the three-

[33]Puzzles are not uncommon when using high frequency monetary policy instruments (Ramey, 2016).

[34]The significance of the OLS results is likely due to Newey-West standard errors underestimating uncertainty. See Müller (2014) for a summary of papers studying the size distortions of HAC estimators. An earlier version of this paper used Newey-West standard errors in the Monte Carlo analysis and found that coverage could be drastically short of the nominal level.

[35]The results are noticeably more erratic when growth rates are used.

month and one-month future surprises respectively.[36]

The results are consistent with what Nakamura and Steinsson (2018a) refer to as the "power problem". That is, the signal to noise ratio may be too small to estimate the impact of monetary policy on lower frequency macroeconomic variables such as output and inflation with any precision. The high frequency identification shocks are changes in the federal funds futures in a tight window (e.g. 30 minutes) around an FOMC meeting. Even if the identification scheme is valid, the shocks may be too small to determine changes in output and inflation, which are monthly variables that are impacted by a host of structural shocks. The horizon h structural impulse response of output to the one-year Treasury yield, for example, is the horizon h response of output to the high frequency instrument divided by the contemporaneous response of the one-year Treasury yield to the instrument. Even if the instrument is relevant (the contemporaneous response of the one-year Treasury yield to the instrument is nonzero and is estimated with precision), if the response of output to the instrument cannot be estimated with any precision, no meaningful inference can be done. The high frequency instruments have an insignificant impact on output and inflation, despite being relevant. The LP GLS results indicate that maybe the high frequency identification shocks cannot be used to determine the impact that monetary policy has on lower frequency aggregate variables like output and inflation.[37]

8 Concluding Remarks

I show that LP can be estimated with GLS. Estimating LP with GLS has three major implications. First, LP GLS can be substantially more efficient and less biased than estimation by OLS with HAC standard errors. Monte Carlo simulations for a wide range of models highlight the benefits of LP GLS. Under assumptions discussed in section 4, it can be shown that impulse responses from LP GLS can be approximately as efficient as impulse responses from VARs. Whether or not the LP GLS is approximately as efficient depends on the persistence of the system, the horizon, and the dependence structure of the system. All else equal, the more persistent the system, the more likely LP GLS impulse responses will be approximately as efficient for horizons typically relevant in practice. It follows that LP can be much more efficient than previously believed.

Second, LP GLS shows that strict exogeneity is not a necessary condition for GLS estimation. Conventional wisdom says that strict exogeneity is a necessary condition for GLS, which makes GLS is more restrictive than OLS (Hayashi, 2000, Stock and Watson, 2007). Since strict exogeneity is considered a necessary condition for GLS, GLS was in part abandoned for OLS with HAC estimation, since OLS with HAC estimation was considered to be more general. Since LP GLS provides a general counterexample that strict exogeneity

[36]Instrument strength varied with block length, but was always above the rule of thumb threshold of 10.

[37]Using similar data and an asymptotically valid Proxy SVAR block bootstrap of Jentsch and Lunsford (2019a), Paul (2020) could not reject the null that monetary policy has no effect on output and inflation at the 95% level (see Paul (2020) for more details). Note that Proxy SVARs assume invertibility (Plagborg-Møller and Wolf, 2019).

is not necessary condition for GLS, it follows that GLS estimation is not as restrictive as previously thought and that GLS may be extended to other situations where strict exogeneity is not satisfied.

Third, since the autocorrelation process can be modeled explicitly, it is now possible to estimate time-varying parameter LP. This was not possible before because the Kalman filter and other popular techniques used to estimate time-varying parameter models require that the error term is uncorrelated or that the autocorrelation process is specified. Time-varying parameter LP can take into account structural instability in the regression coefficients and/or the covariance matrix, and since time-varying parameter models can approximate any form of non-linearity, makes them more robust to model misspecification (Granger, 2008).[38]

The results in this paper have many potential extensions. It would be useful to derive a data dependent rule or cross validation method for the optimal block length when using block bootstrapping for LP. It may be interesting to derive conditions for consistency and asymptotically normality for time-varying parameter LP estimators. It may also be useful to extend LP GLS to a non-linear (in the variables) or non-parametric setting. One potential solution would be to extend polynomial LP, which are motivated by a non-linear version of the Wold representation (see Jordà (2005) section 3 for more details). If one does not want to make assumptions about the functional form or the model, the second potential solution would be to extend nonparametric LP. Lastly, since LP are direct multistep forecasts, the results in this paper have the potential to improve the forecast accuracy of direct multistep forecasts.

[38]Even though time-varying parameter models can approximate any non-linear model (non-linear in the variables and/or the parameters), the approximation is for the conditional mean. If the true model is non-linear in the variables, estimation of the linear (in the variables) time-invariant or time-varying parameter LP GLS would lead to inconsistent estimates of the true impulse responses.

References

Ang, A. and G. Bekaert (2002). Regime switches in interest rates. *Journal of Business and Economic Statistics 20*(2), 163–182.

Baumeister, C. and G. Peersman (2012). The role of time-varying price elasticities in accounting for volatility changes in the crude oil market. *Journal of Applied Econometrics 28*(7), 1087–1109.

Bhansali, R. J. (1997). Direct autoregressive predictors for multistep prediction: Order selection and performance relative to the plug in predictors. *Statistica Sinica 7*, 425–449.

Billingsley, P. (1995). *Probability and Measure* (3rd ed.). Wiley.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (Second ed.). Springer-Verlag New York.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). Inference in vars with conditional heteroskedasticity of unknown form. *Journal of Econometrics 191*(1), 69–85.

Christiano, L., M. Eichenbaum, and C. Evans (1999). *Handbook of Macroeconomics*, Volume 1, Chapter Chapter 2 Monetary policy shocks: What have we learned and to what end?, pp. 65–148. North Holland.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. CHAPMAN HALL/CRC.

Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Goncalves, S. and L. Kilian (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics 123*, 89–120.

Goncalves, S. and L. Kilian (2007). Asymptotic and bootstrap inference for ar(â) processes with conditional heteroskedasticity. *Econometric Reviews 26*(6), 609–641.

Granger, C. and P. Newbold (1977). *Forecasting Economic Time Series*.

Granger, C. W. (2008). Non-linear models: Where do we go next - time varying parameter models? *Studies in Nonlinear Dynamics Econometrics 12*(3), Online.

Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Prentice Hall.

Hamilton, J. (1994). *Time Series Analysis*.

Hannan, E. J. (1970). *Multiple Time Series*. 1970 John Wiley Sons, Inc.

Hansen, L. P. and R. J. Hodrick (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy 88*(5), 829–853.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Jentsch, C. and K. G. Lunsford (2019a). Asymptotically valid bootstrap inference for proxy svars. *Working Paper*.

Jentsch, C. and K. G. Lunsford (2019b). The dynamic effects of personal and corporate income tax changes in the united states: Comment. *American Economic Review 109*(7), 2655–2678.

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

Jordà, Ò. (2009). Simultaneous confidence regions for impulse responses. *Review of Economics and Statistics 91*(3), 629–647.

Jordà, Ò. and S. Kozicki (2011). Estimation and inference by the method of projection minimum distance: An application to the new keynesian hybrid phillips curve. *International Economic Review 52*(2), 461–487.

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

Kilian, L. and Y. J. Kim (2011). How reliable are local projection estimators of impulse responses? *Review of Economics and Statistics 93*(4), 1460–1466.

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

Kline, P. and A. Santos (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods 1*(1), 23–41.

Lahiri, S. (2003). *Resampling Methods for Dependent Data*. New York: Springer.

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

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

Leybourne, S. J., B. P. M. McCabe, and A. R. Tremayne (1996). Can economic time series be differenced to stationarity? *Journal of Business and Economic Statistics 14*(4), 435–446.

McCracken, M. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics 34*(4), 574–589.

Montiel Olea, J. L. and M. Plagborg-Møller (2020). Local projection inference is simpler and more robust than you think. *Working Paper*.

Müller, U. K. (2014). Hac corrections for strongly autocorrelated time series. *Journal of Business and Economic Statistics 32*(3), 311–322.

Murphy, K. M. and R. H. Topel (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics 3*(4), 370–379.

Nakamura, E. and J. Steinsson (2018a). High-frequency identification of monetary non-neutrality: The information effect. *Quarterly Journal of Economics 133*(3), 1283–1330.

Nakamura, E. and J. Steinsson (2018b). Identification in macroeconomics. *Journal of Economic Perspectives 32*(3), 59–86.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review 25*(1), 221–247.

Paul, P. (2020). The time-varying effect of monetary policy on asset prices. *Review of Economics and Statistics 102*(4), 690–704.

Plagborg-Møller, M. and C. K. Wolf (2019). Local projections and vars estimate the same impulse responses. *Working Paper*.

Pope, A. L. (1990). Biases of estimators in multivariate non-gaussian autoregressions. *Jour-nal of Time Series Analysis 11*(3), 249–258.

Priestley, M. B. (1988). *Non-linear and Non-stationary Time Series Analysis*. Academic Press.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, Chapter 2, pp. 71–162.

Ramey, V. A. and S. Zubairy (2018). Government spending multipliers in good times and in bad: Evidence from u.s. historical data. *Journal of Political Economy 126*(2), 850–901.

Shao, X. (2011). A bootstrap-assisted spectral test of white noise under unknown dependence. *Journal of Econometrics 162*(2), 213–224.

Stock, J. and M. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics 14*(1), 11–30.

Stock, J. and M. Watson (2007). *Introduction to Econometrics*. Addison Wesley Longman.

Stock, J. and M. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal 128*(610), 917–948.

West, K. D. and Z. Zhao (2019). *Handbook of Statistics*, Volume 41, Chapter Adjusting for bias in longhorizon regressions using R, pp. 65–80.

White, H. (2001). *Asymptotic Theory for Econometricians*. Emerald Group Publishing Limited.

Yeh, A. B. (1998). A bootstrap procedure in linear regression with nonstationary errors. *The Canadian Journal of Statistics 26*(1), 149–160.

Online Appendix

Preliminaries

Some preliminaries that will be used in the proofs. The proofs rely on several results from Goncalves and Kilian (2007) who focus on univariate autoregressions. As noted in Goncalves and Kilian (2007), multivariate generalizations are possible for all of their results but at the cost of more complicated notation. Define the matrix norm $\| C \|_1^2 = sup_{l \neq 0} l'Cl/l'l$, that is, the largest eigenvalue of C. When C is symmetric, this is the square of the largest eigenvalue of C. A couple of useful inequalities are $\| AB \|^2 \leq \| A \|_1^2 \| B \|^2$ and $\| AB \|_1^2 \leq \| A \|^2 \| B \|_1^2$. Let $E^*(\cdot)$ and $var^*(\cdot)$ denote the expectation and variance with respect to the bootstrap data conditional on the original data.

$$\hat{B}(k,h,OLS) - B(k,h) = U_{1T}\hat{\Gamma}_k^{-1} + U_{2T}\hat{\Gamma}_k^{-1} + U_{3T}\hat{\Gamma}_k^{-1},$$

$$\hat{B}(k,h,GLS) - B(k,h) = U_{1T}\hat{\Gamma}_k^{-1} + U_{2T}\hat{\Gamma}_k^{-1} + U_{3T}\hat{\Gamma}_k^{-1} - U_{4T}\hat{\Gamma}_k^{-1},$$

where

$$U_{1T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} B_j^{(h)} y_{t-j+1}) X'_{t,k}\},$$

$$U_{2T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h} X'_{t,k}\},$$

$$U_{3T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}\},$$

$$U_{4T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \hat{\Theta}_l \hat{\varepsilon}_{t+h-l,k}) X'_{t,k}\}.$$

The Mixingale Central Limit Theorem will be useful in proving several results (c.f. White (2001) pages 124-25).

Definition. Let $\{r_t, \mathcal{F}_t\}$ be an adapted stochastic sequence with $E(r_t^2) < \infty$. Then $\{r_t, \mathcal{F}_t\}$ is an adapted mixingale if there exists finite nonnegative sequences $\{c_t\}$ and $\{\gamma_i\}$ such that $\gamma_i \to 0$ as $i \to \infty$ and

$$(E(E(r_t \big| \mathcal{F}_{t-i})^2))^{1/2} \leq c_t \gamma_i.$$

Theorem. *Mixingale CLT. Let $\{r_t, \mathcal{F}_t\}$ be a stationary ergodic adapted mixingale with $\gamma_i = O_p(i^{-1-\delta})$ for some*

$\delta > 0$. *Then* $var(\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} r_t\}) \xrightarrow{p} \sum_{p=-\infty}^{\infty} cov(r_t, r_{t-p}) < \infty$, *and if* $\sum_{p=-\infty}^{\infty} cov(r_t, r_{t-p}) > 0$,

$$\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H} r_t\} \xrightarrow{d} N(0, \sum_{p=-\infty}^{\infty} cov(r_t, r_{t-p})).$$

A.1 Auxiliary Propositions and Lemmas

A.1.1 Propositions

Proposition 2. *Under Assumption 3,*

$$\| \hat{B}(k,h,OLS) - B(k,h) \| \xrightarrow{p} 0.$$

Proof.

$$\| \hat{B}(k,h,OLS) - B(k,h) \| \leq \{\| U_{1T} \| + \| U_{2T} \| + \| U_{3T} \|\} \| \hat{\Gamma}_k^{-1} \|_1 .$$

Lemma A.1 in Goncalves and Kilian (2007) establishes that $\| \hat{\Gamma}_k^{-1} \|_1$ is bounded in probability , so consistency in LP OLS consists of showing that $\| U_{1T} \|$, $\| U_{2T} \|$, and $\| U_{3T} \|$ converge in probability to 0. This was shown in Jordà and Kozicki (2011), but assuming the errors are i.i.d. However, their proof showing $\| U_{3T} \| \xrightarrow{p} 0$ is incorrect. It is incorrect because $(\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l})X'_{t,k}$ is assumed to be independent across time. It is not. Here I will present a correct proof under the more general conditions stated in Assumption 3 (which include Jordà and Kozicki (2011) as a special case). A correct proof is

$$\| U_{3T} \|^2 = (T-k-H)^{-2} trace\{\sum_{m=k}^{T-H}\sum_{n=k}^{T-H}(\sum_{l=1}^{h-1}\Theta_l \varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l \varepsilon_{m+h-l})X'_{m,k}X_{n,k}\},$$

by the cyclic property of traces.

$$E \| U_{3T} \|^2 = (T-k-H)^{-2} trace \sum_{m=k}^{T-H}\sum_{n=k}^{T-H} E\{(\sum_{l=1}^{h-1}\Theta_l \varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l \varepsilon_{m+h-l})X'_{m,k}X_{n,k}\}.$$

For $|n-m| > h-1$

$$E\{(\sum_{l=1}^{h-1}\Theta_l \varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l \varepsilon_{m+h-l})X'_{m,k}X_{n,k}\} = 0,$$

by the martingale difference assumption. So

$$E \| U_{3T} \|^2 = (T-k-H)^{-2} trace \sum_{m=k}^{T-H}\sum_{|n-m|<h} E\{(\sum_{l=1}^{h-1}\Theta_l \varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l \varepsilon_{m+h-l})X'_{m,k}X_{n,k}\}.$$

Note that

$$| E\{(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})X'_{m,k}X_{n,k}\} |\leq (E[\{(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})\}^2])^{1/2}(E[\{X'_{m,k}X_{n,k}\}^2])^{1/2}$$

by Cauchy-Schwarz inequality. $E[(X'_{m,k}X_{n,k})^2] = O_p(k^2)$ and $| E([(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})]^2) |< \infty$ due to the finite fourth moments of ε and $\sum_{h=0}^{\infty} \| \Theta_h \|< \infty$. Consequently for $|n-m| \leq h-1$,

$$trace\{(E[\{(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})\}^2])^{1/2}(E[\{X'_{m,k}X_{n,k}\}^2])^{1/2}\} = O_p(k).$$

Since h is finite it follows that

$$E \| U_{3T} \|^2 = (T-k-H)^{-2} trace\sum_{m=k}^{T-H}\sum_{|n-m|<h} E\{(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})X'_{m,k}X_{n,k}\} \leq \frac{k \times constant}{T-k-H}.$$

Therefore $\| U_{3T} \| = O_p\Big(\frac{k^{1/2}}{(T-k-H)^{1/2}}\Big) \xrightarrow{p} 0$. To complete the proof of consistency it just needs to be shown that $\| U_{1T} \| \xrightarrow{p} 0$ and $\| U_{2T} \| \xrightarrow{p} 0$. The proof that $\| U_{1T} \| \xrightarrow{p} 0$ is unaffected by allowing for conditional heteroskedasticity, so the proof of convergence in Jordà and Kozicki (2011) (their Proposition 1) can be used. The proof that $\| U_{2T} \| \xrightarrow{p} 0$ follows from Lemma A.2 part C in Goncalves and Kilian (2007). □

Proposition 3. *Under Assumptions 3, for LP OLS*

$$(T-k-H)^{1/2} l(k)' vec[\hat{B}(k,h,OLS) - B(k,h)] \xrightarrow{d} N(0, \Omega(k,h,OLS)),$$

where

$$\Omega(k,h,OLS) = \sum_{p=-h+1}^{h-1} cov(r_t^{(h),OLS}, r_{t-p}^{(h),OLS}).$$

Proof. Under the assumptions Lewis and Reinsel (1985) used to show asymptotic normality of the limiting distribution of the VAR(∞), Jordà and Kozicki (2011) showed the asymptotic normality of the limiting distribution of the LP(∞). It turns out Jordà and Kozicki (2011) use the incorrect Central Limit Theorem. Jordà and Kozicki (2011) proof follows the same argument as Lewis and Reinsel (1985). Lewis and Reinsel (1985) use a martingale CLT to prove asymptotic normality. This is not possible with LP because

$$r_{t+h}^{(h),OLS} = l(k)' vec\{(\varepsilon_{t+h} + \sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l})X'_{t,k}\Gamma_k^{-1}\},$$

is not a martingale difference sequence. Since ε_t is stationary and ergodic, $r_{t+h}^{(h),OLS}$ is stationary and ergodic by Theorem 3.35 in White (2001). Here I will show the corrected proof of LP OLS under the more general conditions of Assumption 3 using the mixingale CLT. The proof will proceed in 2 steps.

1. I'll show $\{r_t^{(h),OLS}, \mathcal{F}_t\}$ is an adapted mixingale with $\gamma_i = O_p(i^{-1-\delta})$ for some $\delta > 0$,

2. I'll show $\sum_{p=-\infty}^{\infty} cov(r_t^{(h),OLS}, r_{t-p}^{(h),OLS}) > 0$.

Step 1. Note that when $i > h-1$, $E[r_t^{(h),OLS} \big| \mathcal{F}_{t-i}] = 0$ by the martingale difference sequence assumption on the errors. Let $c_t = (E(E(r_t^{(h),OLS} \big| \mathcal{F}_{t-i})^2))^{1/2}\Delta$, where $\Delta = h^{\nu/(\nu+1)}$ for any $\nu > 0$, and $\gamma_i = i^{-(\nu+1)/\nu}$. Note that $-(\nu+1)/\nu < -1$ for any $\nu > 0$ and $\delta = 1/\nu$.

Step 2. $r_t^{(h),OLS}$ can only be correlated up to $h-1$ horizons out. Since $r_t^{(h),OLS}$ is a zero mean stationary process with absolutely summable autocovariances, $r_t^{(h),OLS}$ can be written in terms of its the Wold representation $r_t^{(h),OLS} = f_t + \sum_{j=1}^{\infty} \mu_j f_{t-j}$, with $det\{\mu(z)\} \neq 0$ for $|z| \leq 1$, since the Wold representation is invertible by construction. Since the autocovariances are absolutely summable $\sum_{p=-\infty}^{\infty} cov(r_t^{(h),OLS}, r_{t-p}^{(h),OLS}) = s_{r^{(h),OLS}}(0)$, where $s_{r^{(h),OLS}}(\omega) = |\mu(e^{-i\omega})|^2(2\pi)^{-1}var(f_t)$ is the spectral density of $r_t^{(h),OLS}$ at frequency ω. By equation 5.7.9 in Brockwell and Davis (1991) $|\mu(e^0)|^2 = |\mu(1)|^2 = |\sum_{j=1}^{\infty} \mu_j|^2 > 0$, therefore

$$\sum_{p=-\infty}^{\infty} cov(r_t^{(h),OLS}, r_{t-p}^{(h),OLS}) = |\sum_{j=1}^{\infty} \mu_j|^2(2\pi)^{-1}var(f_t) > 0.$$

□

Proposition 4. *Under Assumption 3, for LP GLS*

$$(T-k-H)^{1/2}l(k)'vec[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} (T-k-H)^{1/2}l(k)'vec[U_{2T}\Gamma_k^{-1} + U_{3T}\Gamma_k^{-1} - U_{4T}\Gamma_k^{-1}].$$

Proof. To show

$$(T-k-H)^{1/2}l(k)'vec[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} (T-k-H)^{1/2}l(k)'vec[U_{2T}\Gamma_k^{-1} + U_{3T}\Gamma_k^{-1} - U_{4T}\Gamma_k^{-1}],$$

we need to show that

$$\| (T-k-H)^{1/2}l(k)'vec[U_{1T} + U_{2T} + U_{3T} - U_{4T}](\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \| \xrightarrow{p} 0,$$

and

$$\| (T-k-H)^{1/2}l(k)'vec[U_{1T}\Gamma_k^{-1}] \| \xrightarrow{p} 0.$$

Jordà and Kozicki (2011) already showed that

$$\| (T-k-H)^{1/2}l(k)'vec\{[U_{1T} + U_{2T} + U_{3T}](\hat{\Gamma}_k^{-1} - \Gamma_k^{-1})\} \| \xrightarrow{p} 0,$$

under the assumption that the errors are iid (see their Proposition 2). Under Assumption 3, their proof still holds since Goncalves and Kilian (2007) showed $k^{1/2} \parallel \hat{\Gamma}_k^{-1} - \Gamma_k^{-1} \parallel_1 = o_p(1)$ (see their Lemma A.1). From Proposition 2 in Jordà and Kozicki (2011), we know that $\parallel \sqrt{T-k-H} U_{1T} \Gamma_k^{-1} \parallel \xrightarrow{p} 0$. So to complete the proof, I just need to show

$$\parallel (T-k-H)^{1/2} l(k)' vec[U_{4T}(\hat{\Gamma}_k^{-1} - \Gamma_k^{-1})] \parallel \xrightarrow{p} 0.$$

Since $0 < M_1 \leq \parallel l(k) \parallel^2 \leq M_2 < \infty$, it suffices to show that $\parallel (T-k-H)^{1/2} U_{4T}(\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \parallel \xrightarrow{p} 0$. Note that

$$\sqrt{T-k-H} U_{4T}(\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) = \{(T-k-H)^{-1/2} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \hat{\Theta}_l \hat{\varepsilon}_{t+h-l,k}) X'_{t,k}\} (\hat{\Gamma}_k^{-1} - \Gamma_k^{-1})$$

$$= \{(T-k-H)^{-1/2} \sum_{l=1}^{h-1} \hat{\Theta}_l \sum_{t=k}^{T-H} (\varepsilon_{t+h-l} + (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k})) X'_{t,k}\} (\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}),$$

since $\hat{\varepsilon}_{t,k} = \varepsilon_t + (\sum_{j=k+1}^{\infty} A_j y_{t-j}) - (\hat{B}(k,1) - B(k,1)) X_{t-1,k})$. So

$$\parallel \sqrt{T-k-H} U_{4T}(\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \parallel$$

$$\leq \sum_{l=1}^{h-1} \parallel \hat{\Theta}_l \parallel \Big(\parallel \{[k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\varepsilon_{t+h-l} + (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k})) X'_{t,k}\} \parallel \Big)$$
$$\times \{k^{1/2} \parallel (\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \parallel_1\}.$$

By Theorem 2, $\parallel \hat{\Theta}_l \parallel \xrightarrow{p} \parallel \Theta_l \parallel < \infty$ for each $1 \leq l \leq h-1$. We know from Goncalves and Kilian (2007) that $k^{1/2} \parallel (\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \parallel_1 \xrightarrow{p} 0$. Since $h-1$ is finite, I just need to show that

$$\Big(\parallel \{[k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\varepsilon_{t+h-l} + (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k})) X'_{t,k}\} \parallel \Big),$$

is bounded for each $1 \leq l \leq h-1$.

$$\Big(\parallel \{[k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\varepsilon_{t+h-l} + (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k})) X'_{t,k}\} \parallel \Big)$$

$$\leq \parallel [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k} \parallel +$$

$$\parallel [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) X'_{t,k} \parallel + \parallel -[k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k} X'_{t,k} \parallel.$$

$\parallel [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k} \parallel$ is bounded since it was shown in Theorem 2 that

$$\parallel (T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k} \parallel = O_p\Big(\Big(\frac{k}{T-k-H}\Big)^{1/2}\Big).$$

Jordà and Kozicki (2011) show that $\| \; [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) X'_{t,k} \; \| \xrightarrow{p} 0$ (see their Proposition 2). For the final term note that

$$\| \; [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k} X'_{t,k} \; \|$$

$$\leq \underbrace{\left(\frac{T-k-H}{k}\right)^{1/2} \| \; (\hat{B}(k,1) - B(k,1)) \; \|}_{bounded} \underbrace{\| \; (T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t+h-l-1,k} X'_{t,k} \; \|_1}_{bounded} .$$

□

Proposition 5. *Under assumption 4,*

$$(T-k-H)^{1/2} vech[\hat{\Sigma} - \Sigma] \xrightarrow{d} N(0, V_{22}).$$

Proof. Substituting out $\hat{\varepsilon}_{t,k} = \varepsilon_t + (\sum_{j=k+1}^{\infty} A_j y_{t-j}) - (\hat{B}(k,1) - B(k,1)) X_{t-1,k})$,

$$\sqrt{T-k-H}\hat{\Sigma} = \sqrt{T-k-H} \frac{\sum_{t=k}^{T-H} \hat{\varepsilon}_{t,k} \hat{\varepsilon}'_{t,k}}{T-k-H}$$

$$= \sqrt{T-k-H} \frac{\sum_{t=k}^{T-H} \varepsilon_t \varepsilon'_t}{T-k-H} + \underbrace{\sqrt{T-k-H} \frac{\sum_{t=k}^{T-H} \varepsilon_t (\sum_{j=k+1}^{\infty} A_j y_{t-j})'}{T-k-H}}_{O_p(\sqrt{T-k-H} \sum_{j=k+1}^{\infty} \| A_j \|)} - \underbrace{\frac{\sum_{t=k}^{T-H} \varepsilon_t X'_{t-1,k}}{T-k-H}}_{O_p(\left(\frac{k}{T-k-H}\right)^{1/2})} \underbrace{\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))'}_{\xrightarrow{d}}$$

$$+ \underbrace{\sqrt{T-k-H} \frac{\sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} A_j y_{t-j}) \varepsilon'_t}{T-k-H}}_{O_p(\sqrt{T-k-H} \sum_{j=k+1}^{\infty} \| A_j \|)} + \underbrace{\sqrt{T-k-H} \frac{\sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} A_j y_{t-j})(\sum_{j=k+1}^{\infty} A_j y_{t-j})'}{T-k-H}}_{O_p(\sqrt{T-k-H} (\sum_{j=k+1}^{\infty} \| A_j \|)^2)}$$

$$- \underbrace{\frac{\sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} A_j y_{t-j}) X'_{t-1,k}}{T-k-H}}_{O_p(k^{1/2} \sum_{j=k+1}^{\infty} \| A_j \|)} \underbrace{\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))'}_{\xrightarrow{d}} - \underbrace{\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))}_{\xrightarrow{d}} \underbrace{\frac{\sum_{t=k}^{T-H} X_{t-1,k} \varepsilon'_t}{T-k-H}}_{O_p(\left(\frac{k}{T-k-H}\right)^{1/2})}$$

$$- \underbrace{\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))}_{\xrightarrow{d}} \underbrace{\frac{\sum_{t=k}^{T-H} X_{t-1,k} (\sum_{j=k+1}^{\infty} A_j y_{t-j})'}{T-k-H}}_{O_p(k^{1/2} \sum_{j=k+1}^{\infty} \| A_j \|)}$$

$$+ \underbrace{\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))}_{\xrightarrow{d}} \underbrace{\frac{\sum_{t=k}^{T-H} X_{t-1,k} X'_{t-1,k}}{T-k-H}}_{bounded} \underbrace{(\hat{B}(k,1) - B(k,1))'}_{O_p(\left(\frac{k}{T-k-H}\right)^{1/2})} .$$

It follows that

$$\sqrt{T-k-H}[\hat{\Sigma}-\Sigma] \xrightarrow{p} \sqrt{T-k-H}\left[\frac{\sum_{t=k}^{T-H}\varepsilon_t\varepsilon_t' - \Sigma}{T-k-H}\right].$$

Since ε_t is mixing, by Theorem 3.49 in White (2001), $\varepsilon_t\varepsilon_t'$ is mixing of the same order. Assuming V_{22} is finite and positive definite, by the strong mixing Central Limit Theorem (Theorem A.8 in Lahiri (2003)), $(T-k-H)^{1/2}vech[\hat{\Sigma}-\Sigma] \xrightarrow{d} N(0,V_{22})$. To show that V_{22} is finite and positive definite, note that absolute summability of fourth order cumulants implies absolute summability of fourth order moments (it follows from Hannan (1970) equation 5.1 on pg. 23). Absolute summability of the fourth order moments of ε implies $V_{22} < \infty$, and since the autocovariances of $\varepsilon_t\varepsilon_t'$ are absolutely summable, positive definiteness of V_{22} follows the same argument used in Step 2 of the mixingale CLT proof in Proposition 3. □

Proposition 6. *Assume that $y_{t+1} = ay_t + \varepsilon_{t+1}$, where $|a| < 1$ and ε_t is an i.i.d. process with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma^2$. If the true lag order is known, then*

$$\sqrt{T}(\hat{b}^{(h),GLS} - a^h) \xrightarrow{d} N(0, [\{1-a^{2h-2}\} + h^2a^{2h-2}](1-a^2)).$$

Proof. Define $\hat{\Gamma} = \frac{1}{T-H}(\sum_{t=1}^{T-H} y_t^2)$,

$$\hat{b}^{(h),GLS} = (\sum_{t=1}^{T-H} y_t^2)^{-1}(\sum_{t=1}^{T-H} y_t(y_{t+h} - \hat{b}^{(h-1),GLS}\hat{\varepsilon}_{t+1} - \ldots - \hat{b}^{(1),GLS}\hat{\varepsilon}_{t+h-1})).$$

Substituting out $y_{t+h} = a^h y_t + a^{h-1}\varepsilon_{t+1} + \ldots + a\varepsilon_{t+h-1} + \varepsilon_{t+h}$ yields

$$\hat{b}^{(h),GLS} = (\sum_{t=1}^{T-H} y_t^2)^{-1}(\sum_{t=1}^{T-H} y_t(a^h y_t + [a^{h-1}\varepsilon_{t+1} - \hat{b}^{(h-1),GLS}\hat{\varepsilon}_{t+1}] + \ldots + [a\varepsilon_{t+h-1} - \hat{b}^{(1),OLS}\hat{\varepsilon}_{t+h-1}] + \varepsilon_{t+h})),$$

$$\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h) = [\sum_{p=1}^{h-1} \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t[a^p\varepsilon_{t+h-p} - \hat{b}^{(p),GLS}\hat{\varepsilon}_{t+h-p}]}{\hat{\Gamma}}] + \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+h}}{\hat{\Gamma}}.$$

It follows from Lemma 5 that

$$\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h) = \underbrace{(\sum_{p=1}^{h-1}\hat{b}^{(p),GLS}a^{h-p-1})}_{plim=(h-1)a^{h-1}} \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+1}}{\hat{\Gamma}} + \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+h}}{\hat{\Gamma}}.$$

$\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+1}}{\hat{\Gamma}}$ and $\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+h}}{\hat{\Gamma}}$ jointly convergence to a normal distribution due to the Mixingale CLT (see proof of Proposition 3 for setup). Therefore

$$\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h) \xrightarrow{d} N(0, [\{1-a^{2h-2}\} + h^2a^{2h-2}](1-a^2)).$$

$\square$

A.1.2 Lemmas

Lemma 1. *If Assumption 3 holds,*

$$(T-k-H)^{1/2} l(k)' vec[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p}$$

$$(T-k-H)^{-1/2} l(k)' vec[(\sum_{t=k}^{T-H} \varepsilon_{t+h} X'_{t,k})\Gamma_k^{-1}] + l(k)' \Big(\sum_{l=1}^{h-1} \{\Gamma_k^{-1}\Gamma'_{(h-l-1),k} \otimes \Theta_l\}\Big) vec[\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))].$$

Proof. From Proposition 4 we know that

$$(T-k-H)^{1/2} l(k)' vec[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} (T-k-H)^{1/2} l(k)' vec[U_{2T}\Gamma_k^{-1} + U_{3T}\Gamma_k^{-1} - U_{4T}\Gamma_k^{-1}]$$

$$= (T-k-H)^{-1/2} l(k)' vec\Big\{(\sum_{t=k}^{T-H} \varepsilon_{t+h} X'_{t,k})\Gamma_k^{-1} + \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}\Gamma_k^{-1} - \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \hat{\Theta}_l \hat{\varepsilon}_{t+h-l}) X'_{t,k}\Gamma_k^{-1}\Big\}.$$

Note that

$$(T-k-H)^{-1/2} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \hat{\Theta}_l \hat{\varepsilon}_{t+h-l,k}) X'_{t,k}\Gamma_k^{-1} = \underbrace{(T-k-H)^{-1/2} \sum_{t=k}^{T-H} \sum_{l=1}^{h-1} \hat{\Theta}_l (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) X'_{t,k}\Gamma_k^{-1}}_{o_p(1)}$$

$$+ (T-k-H)^{-1/2} \sum_{t=k}^{T-H} \sum_{l=1}^{h-1} \hat{\Theta}_l \varepsilon_{t+h-l} X'_{t,k}\Gamma_k^{-1} - (T-k-H)^{-1/2} \sum_{t=k}^{T-H} \sum_{l=1}^{h-1} \hat{\Theta}_l (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k} X'_{t,k}\Gamma_k^{-1},$$

where the first term converges to zero since $h-1$ is finite, $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$, $\| \Gamma_k^{-1} \|_1 < \infty$, and Theorem 1 in Lewis and Reinsel (1985). Since $\| \hat{\Gamma}_{(h-l-1),k} \|$, $\| \hat{\Gamma}_k \|$, and $\| \hat{\Theta}_l \|$ are consistent and bounded in probability

$$\| \Big(\sum_{l=1}^{h-1} \{\hat{\Gamma}_k^{-1}\hat{\Gamma}'_{(h-l-1),k} \otimes \hat{\Theta}_l\}\Big) \| \xrightarrow{p} \| \Big(\sum_{l=1}^{h-1} \{\Gamma_k^{-1}\Gamma'_{(h-l-1),k} \otimes \Theta_l\}\Big) \| < \infty.$$

Therefore

$$(T-k-H)^{-1/2} l(k)' vec\Big\{\sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}\Gamma_k^{-1} - \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \hat{\Theta}_l \hat{\varepsilon}_{t+h-l}) X'_{t,k}\Gamma_k^{-1}\Big\} \xrightarrow{p}$$

$$l(k)' \Big(\sum_{l=1}^{h-1} \{\Gamma_k^{-1}\Gamma'_{(h-l-1),k} \otimes \Theta_l\}\Big) vec[\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))],$$

and

$$(T-k-H)^{1/2} l(k)' vec[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} (T-k-H)^{-1/2} l(k)' vec[(\sum_{t=k}^{T-H} \varepsilon_{t+h} X'_{t,k}) \Gamma_k^{-1}]$$

$$+ l(k)' \Big(\sum_{l=1}^{h-1} \{\Gamma_k^{-1} \Gamma'_{(h-l-1),k} \otimes \Theta_l\} \Big) vec[\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))].$$

□

Lemma 2. *Under Assumption 3, for the reduced form wild bootstrap*

$$\| \hat{B}^*(k,h,GLS) - \hat{B}(k,h) \| \xrightarrow{p^*} 0.$$

Proof. This will be a proof by induction. Assume the consistency for the previous $h-1$ horizons has been proven. Hence $\| \hat{\Theta}_l^* \| \xrightarrow{p^*} \| \hat{\Theta}_l \| < \infty$ for $1 \leq l \leq h-1$.

$$\| \hat{B}^*(k,h,GLS) - \hat{B}(k,h) \|$$

$$\leq \| (T-k-H)^{-1} \sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+h,k} X'_{t,k} \| \| \hat{\Gamma}_k^{-1} \|_1 + \sum_{l=1}^{h-1} \| \hat{\Theta}_l^* \| \| (T-k-H)^{-1} \sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+1,k} X'_{t,k} \| \| \hat{\Gamma}_k^{-1} \hat{\Gamma}_{(h-l-1),k} \hat{\Gamma}_k^{-1} \|_1 .$$

$\| \hat{\Gamma}_k^{-1} \|_1$, $\| \hat{\Theta}_l^* \|$, and $\| \hat{\Gamma}_k^{-1} \hat{\Gamma}_{(h-l-1),k} \hat{\Gamma}_k^{-1} \|_1$ are bounded in probability so it's sufficient to show that

$$\| (T-k-H)^{-1} \sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+h,k} X'_{t,k} \| \xrightarrow{p^*} 0 \quad and \quad \| (T-k-H)^{-1} \sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+1,k} X'_{t,k} \| \xrightarrow{p^*} 0.$$

The proofs are the same, so I'll just show $E^*[\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+h,k} X'_{t,k}\} \|^2] \xrightarrow{p^*} 0$. Note that

$$(T-k-H)^{-2} trace\{[\sum_{n=k}^{T-H} \hat{\varepsilon}^*_{n+h,k} X'_{n,k}]' [\sum_{m=k}^{T-H} \hat{\varepsilon}^*_{m+h,k} X'_{m,k}]\} = (T-k-H)^{-2} trace\{\sum_{m=k}^{T-H} \sum_{n=k}^{T-H} \hat{\varepsilon}^{*'}_{n+h,k} \hat{\varepsilon}^*_{m+h,k} X'_{m,k} X_{n,k}\},$$

by the cyclic property of traces. Note that

$$E^* trace\{\sum_{m=k}^{T-H} \sum_{n=k}^{T-H} \hat{\varepsilon}^{*'}_{n+h,k} \hat{\varepsilon}^*_{m+h,k} X'_{m,k} X_{n,k}\} = trace\{\sum_{m=k}^{T-H} \hat{\varepsilon}'_{m+h,k} \hat{\varepsilon}_{m+h,k} X'_{m,k} X_{m,k}\} = O_p((T-k-H)k),$$

since $E^*[\hat{\varepsilon}^{*'}_{n+h,k} \hat{\varepsilon}^*_{m+h,k}] = 0$ for $m \neq n$. It follows that

$$E^* \| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \hat{\varepsilon}^*_{t+h,k} X'_{t,k}\} \|^2 \leq (T-k-H)^{-2} O_p([T-k-H]k) = O_p\Big(\frac{k}{T-k-H}\Big) \xrightarrow{p^*} 0.$$

To complete the proof, note that the horizon 1 LP is a VAR, and the proof of consistency is provided by

Lemma A.5 in Goncalves and Kilian (2007). □

Lemma 3. *Under Assumption 3,*

$$\| \hat{V}_{11}(k,H) \| \xrightarrow{p} \| V_{11}(k,H) \| .$$

Proof. Let $\hat{V}_{11}(k,H) = (T-k-H)^{-1} \sum_{t=k}^{T-H} R\hat{Score}_{t+1}^{(H)} R\hat{Score}_{t+1}^{(H)\prime}$. Define

$$F_{k,m,n} = E[(X_{t-m,k} \otimes I_r)\varepsilon_{t+1}\varepsilon_{t+1}'(X_{t-n,k} \otimes I_r)'].$$

Let $\hat{F}_{k,m,n} = (T-k-H)^{-1} \sum_{t=k}^{T-H} [(X_{t-m,k} \otimes I_r)\varepsilon_{t+1}\varepsilon_{t+1}'(X_{t-n,k} \otimes I_r)']$ and $s_{k,0} = I_{kr^2 \times kr^2}$. For $m,n = 0,\ldots,H-1$ and $i,j = 0,\ldots,H$

$$l(k)'\hat{s}_{k,i}(T-k-H)^{-1} \sum_{t=k}^{T-H} [(\hat{\Gamma}_k^{-1} X_{t-m,k} \otimes I_r)\hat{\varepsilon}_{t+1}\hat{\varepsilon}_{t+1}'(\hat{\Gamma}_k^{-1} X_{t-n,k} \otimes I_r)']\hat{s}_{k,j}' l(k)$$

$$-l(k)' s_{k,i} E[(\Gamma_k^{-1} X_{t-m,k} \otimes I_r)\varepsilon_{t+1}\varepsilon_{t+1}'(\Gamma_k^{-1} X_{t-n,k} \otimes I_r)'] s_{k,j}' l(k)$$

$$= l(k)'\hat{s}_{k,i}(\hat{\Gamma}_k^{-1} \otimes I_r)\hat{F}_{k,m,n}(\hat{\Gamma}_k^{-1} \otimes I_r)'\hat{s}_{k,i}' l(k) - l(k)' s_{k,i}(\Gamma_k^{-1} \otimes I_r) F_{k,m,n}(\Gamma_k^{-1} \otimes I_r)' s_{k,j} l(k)$$

$$= l(k)'\hat{s}_{k,i}(\hat{\Gamma}_k^{-1} \otimes I_r)[\hat{F}_{k,m,n} - F_{k,m,n}](\Gamma_k^{-1} \otimes I_r)' s_{k,j}' l(k)$$

$$+l(k)'[\hat{s}_{k,i}(\hat{\Gamma}_k^{-1} \otimes I_r) - s_{k,i}(\Gamma_k^{-1} \otimes I_r)] F_{k,m,n}'(\Gamma_k^{-1} \otimes I_r)' s_{k,j}' l(k)$$

$$+l(k)' s_{k,i}'(\Gamma_k^{-1} \otimes I_r)\hat{F}_{k,m,n}[\hat{s}_{k,j}(\hat{\Gamma}_k^{-1} \otimes I_r) - s_{k,j}(\Gamma_k^{-1} \otimes I_r)]' l(k).$$

Since $\| \hat{s}_{k,h} \| \xrightarrow{p} \| s_{k,h} \| < \infty$, $\| \hat{\Gamma}_k^{-1} \| \xrightarrow{p} \| \Gamma_k^{-1} \| < \infty$, $\| F_{k,m,n} \| < \infty$, $\| l(k)' \| < \infty$, and $2H+1$ is finite, then showing $\| \hat{V}_{11}(k,H) \| \xrightarrow{p} \| V_{11}(k,H) \|$ simplifies to showing $\| \hat{F}_{k,m,n} - F_{k,m,n} \| \xrightarrow{p} 0$ for $m,n = 0,\ldots,H-1$. Convergence follows same argument as the proof of Theorem 2.2 in Goncalves and Kilian (2007). □

Lemma 4. *Under Assumption 5,*

$$\| \hat{V}^{lr}(k,H) \| \xrightarrow{p} \| V(k,H) \|,$$

where

$$\hat{V}^{lr}(k,H) = \begin{bmatrix} \hat{V}_{11}^{lr}(k,H) & \hat{V}_{12}(k,H) \\ \hat{V}_{21}(k,H) & \hat{V}_{22} \end{bmatrix},$$

$$\hat{V}_{11}^{lr}(k,H) = \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k}^{T-H} R\hat{score}_{t+1}^{(H)} R\hat{score}_{t+1-p}^{(H)\prime},$$

$$\hat{V}_{12}(k,H) = \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k}^{T-H} \{R\hat{score}_{t+1}^{(H)} vec\left[\hat{\varepsilon}_{t+1-p}\hat{\varepsilon}_{t+1-p}' - \hat{\Sigma}\right]' L_r\},$$

$$\hat{V}_{22} = \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} L_r' \{ \sum_{t=k}^{T-H} (vec(\hat{\varepsilon}_{t+1}\hat{\varepsilon}_{t+1}'), vec(\hat{\varepsilon}_{t+1-p}\hat{\varepsilon}'_{t+1-p})') - vec(\hat{\Sigma})vec(\hat{\Sigma})' \} L_r.$$

Proof. First note that

$$\| [(T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} RStrucScore_t^{(H)} RStrucScore_{t-p}^{(H)\prime}] - E[RStrucScore_t^{(H)} RStrucScore_{t-p}^{(H)\prime}] \|$$

$$= O_p(\frac{k}{(T-k-H)^{1/2}}).$$

Proof follows the same argument as the proof of Theorem 2.2 in Goncalves and Kilian (2007) (in particular their proof that $A_3 = O_p(\frac{k}{(T-k-H)^{1/2}})$). Before applying their proof, replace the setup in the beginning of their proof with the setup in Lemma 3 but applied to $RStrucScore_t^{(H)}$. Note that $RStrucScore_t^{(H)} = [RScore_t^{(H)\prime}, vech(\varepsilon_{t+1}\varepsilon_{t+1}' - \Sigma)']'$, and convergence is not affected by also accounting for $vech(\varepsilon_{t+1}\varepsilon_{t+1}' - \Sigma)'$ due to cumulant condition on ε and since it's finite dimensional. The explicit setup of $RStrucScore_t^{(H)}$ is omitted due to brevity. It follows that

$$\| \sum_{p=-\ell}^{\ell} \{ [(T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} RStrucScore_t^{(H)} RStrucScore_{t-p}^{(H)\prime}] - E[RStrucScore_t^{(H)} RStrucScore_{t-p}^{(H)\prime}] \} \|$$

$$= O_p(\frac{k\ell}{(T-k-H)^{1/2}}) \xrightarrow{p} 0.$$

Therefore, I just need to show

$$\| \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} [RStru\hat{c}Score_t^{(H)} RStru\hat{c}Score_{t-p}^{(H)\prime} - RStrucScore_t^{(H)} RStrucScore_{t-p}^{(H)\prime}] \| \xrightarrow{p} 0.$$

The proof will proceed in 3 parts. First I'll show $\| \hat{V}_{22} \| \xrightarrow{p} \| V_{22} \|$, second $\| \hat{V}_{12}(k,H) \| \xrightarrow{p} \| V_{12}(k,H) \|$, and lastly $\| \hat{V}_{11}^{lr}(k,H) \| \xrightarrow{p} \| V_{11}(k,H) \|$.

To show $\| \hat{V}_{22} \| \xrightarrow{p} \| V_{22} \|$, note since $\| (\hat{B}(k,1) - B(k,1) \| = O_p(\frac{k^{1/2}}{T^{1/2}})$, $\| X_{t-1,k} \| = O_p(k^{1/2})$, and by Theorem 1 and in Lewis and Reinsel (1985)

$$\| \hat{\varepsilon}_{t,k} \| \leq \underbrace{\| \varepsilon_t \|}_{O_p(1)} + \underbrace{\| (\sum_{j=k+1}^{\infty} A_j y_{t-j}) \|}_{O_p(\sum_{j=k+1}^{\infty} A_j)} + \underbrace{\| -(\hat{B}(k,1) - B(k,1)) X_{t-1,k}) \|}_{O_p\left(\frac{k}{T^{1/2}}\right)},$$

implying $\| \hat{\varepsilon}_{t,k} - \varepsilon_t \| = O_p(\frac{k}{T^{1/2}})$. It follows that $\| \hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t-p,k}\hat{\varepsilon}_{t-p,k} - \varepsilon_t\varepsilon_t\varepsilon_{t-p}\varepsilon_{t-p} \| = O_p(\frac{k}{T^{1/2}})$. Therefore

$$\| (T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} [\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t-p,k}\hat{\varepsilon}_{t-p,k} - \varepsilon_t\varepsilon_t\varepsilon_{t-p}\varepsilon_{t-p}] \| = O_p(\frac{k}{T^{1/2}})$$

$$\implies \| \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} [\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t-p,k}\hat{\varepsilon}_{t-p,k} - \varepsilon_t\varepsilon_t\varepsilon_{t-p}\varepsilon_{t-p}] \| = O_p(\frac{k\ell}{T^{1/2}}).$$

Since $\| \hat{\Sigma} - \Sigma \| = O_p([T-k-H]^{-1})$ by Proposition 5, $\| \hat{V}_{22} - V_{22} \| = O_p(\frac{k\ell}{T^{1/2}}) = O_p((\frac{k^4}{T}\frac{\ell^4}{T})^{1/4}) \xrightarrow{p} 0$.

Now to show $\| \hat{V}_{12}(k,H) \| \xrightarrow{p} \| V_{12}(k,H) \|$. Using an analogous set up as Lemma 3, it suffices to show that

$$\| \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} \{(X_{t-m,k} \otimes I_r)[\hat{\varepsilon}_t vec\left[\hat{\varepsilon}_{t-p}\hat{\varepsilon}'_{t-p} - \hat{\Sigma}\right]' - \varepsilon_t vec\left[\varepsilon_{t-p}\varepsilon'_{t-p} - \Sigma\right]']L_r\} \| \xrightarrow{p} 0.$$

for $m = 1, \ldots, H$. $\| \hat{\varepsilon}_{t,k} - \varepsilon_t \| = O_p(\frac{k}{T^{1/2}})$ implies $\| \hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t-p,k}\hat{\varepsilon}_{t-p,k} - \varepsilon_t\varepsilon_{t-p}\varepsilon_{t-p} \| = O_p(\frac{k}{T^{1/2}})$. Therefore,

$$\| (X_{t-m,k} \otimes I_r)[\hat{\varepsilon}_t vec\left[\hat{\varepsilon}_{t-p}\hat{\varepsilon}'_{t-p} - \hat{\Sigma}\right]' - \varepsilon_t vec\left[\varepsilon_{t-p}\varepsilon'_{t-p} - \Sigma\right]']L_r \| = O_p(\frac{k^2}{T^{1/2}}),$$

since $(X_{t-m,k} \otimes I_r)$ is $kr^2 \times r$. It follows that

$$\| \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} \{l(k)'(X_{t-m,k} \otimes I_r)\hat{\varepsilon}_t vec\left[\hat{\varepsilon}_{t-p}\hat{\varepsilon}'_{t-p} - \hat{\Sigma}\right]' L_r\} \|$$

$$= \| \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} \{l(k)'(X_{t-m,k} \otimes I_r)\varepsilon_t vec\left[\varepsilon_{t-p}\varepsilon'_{t-p} - \Sigma\right]' L_r\} \| + O_p(\frac{k^2\ell}{T^{1/2}}),$$

which implies $\| \hat{V}_{12}(k,H) - V_{12}(k,H) \| = O_p(\frac{k^2}{T^{1/2}}\ell) = O_p((\frac{k^8}{T}\frac{\ell^4}{T})^{1/4}) \xrightarrow{p} 0$.

Now to show $\| \hat{V}^{lr}_{11}(k,H) \| \xrightarrow{p} \| V_{11}(k,H) \|$. Using an analogous set up as Lemma 3, it suffices to show that

$$\| \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} (X_{t-m,k} \otimes I_r)[\hat{\varepsilon}_t\hat{\varepsilon}'_{t-p} - \varepsilon_t\varepsilon'_{t-p}](X_{t-p-n,k} \otimes I_r)' \| \xrightarrow{p} 0,$$

for $m, n = 1, \ldots, H$. Since $\| \hat{\varepsilon}_{t,k}\hat{\varepsilon}_{t-p,k} - \varepsilon_t\varepsilon_{t-p} \| = O_p(\frac{k}{T^{1/2}})$, it follows that

$$\| (X_{t-m,k} \otimes I_r)[\hat{\varepsilon}_t\hat{\varepsilon}'_t - \varepsilon_t\varepsilon'_{t-p}](X_{t-p-n,k} \otimes I_r)' \| = O_p(\frac{k^3}{T^{1/2}}),$$

since $(X_{t-m,k} \otimes I_r)$ is $kr^2 \times r$. It follows that

$$\| \sum_{p=-\ell}^{\ell} (T-k-H)^{-1} \sum_{t=k+1}^{T-H+1} (X_{t-m,k} \otimes I_r)[\hat{\varepsilon}_t\hat{\varepsilon}'_{t-p} - \varepsilon_t\varepsilon'_{t-p}](X_{t-p-n,k} \otimes I_r)' \| = O_p(\frac{k^3}{T^{1/2}}\ell).$$

This implies that $\| \hat{V}^{lr}_{11}(k,H) - V_{11}(k,H) \| = O_p(\frac{k^3}{T^{1/2}}\ell) = O_p((\frac{k^8}{T}\frac{k^8}{T}\frac{k^8}{T}\frac{\ell^8}{T})^{1/8}) \xrightarrow{p} 0$. □

Lemma 5. *Under the assumptions used for Proposition 6, for any integer* $1 \le p \le h-1$,

$$\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\left[a^p\varepsilon_{t+h-p} - \hat{b}^{(p),GLS}\hat{\varepsilon}_{t+h-p}\right]}{\hat{\Gamma}} \xrightarrow{p} \underbrace{\hat{b}^{(p),GLS}a^{h-p-1}}_{plim=a^{h-1}}\underbrace{\sqrt{T-H}(\hat{a}-a)}_{\xrightarrow{d}} + o_p(1).$$

Proof.

$$\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\left[a^p\varepsilon_{t+h-p} - \hat{b}^{(p),GLS}\hat{\varepsilon}_{t+h-p}\right]}{\hat{\Gamma}} = \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t a^p\varepsilon_{t+h-p}}{\hat{\Gamma}} - \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\hat{b}^{(p),GLS}\hat{\varepsilon}_{t+h-p}}{\hat{\Gamma}}.$$

Substitute out $\hat{\varepsilon}_{t+h-p} = (a-\hat{a})y_{t+h-p-1} + \varepsilon_{t+h-p}$

$$= \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t a^p\varepsilon_{t+h-p}}{\hat{\Gamma}} - \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\hat{b}^{(p),GLS}\left((a-\hat{a})y_{t+h-p-1} + \varepsilon_{t+h-p}\right)}{\hat{\Gamma}},$$

$$= \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t a^p\varepsilon_{t+h-p}}{\hat{\Gamma}} - \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\hat{b}^{(p),GLS}(a-\hat{a})y_{t+h-p-1}}{\hat{\Gamma}} - \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\hat{b}^{(p),GLS}\varepsilon_{t+h-p}}{\hat{\Gamma}},$$

$$= \underbrace{\left(a^p - \hat{b}^{(p),GLS}\right)}_{plim=0}\underbrace{\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+h-p}}{\hat{\Gamma}}}_{\xrightarrow{d}} - \hat{b}^{(p),GLS}(a-\hat{a})\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t y_{t+h-p-1}}{\hat{\Gamma}},$$

where convergence in distribution is due to the Mixingale Central Limit Theorem. It follows that

$$\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\left[a^p\varepsilon_{t+h-p} - \hat{b}^{(p),GLS}\hat{\varepsilon}_{t+h-p}\right]}{\hat{\Gamma}} = o_p(1) + \hat{b}^{(p),GLS}(\hat{a}-a)\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t y_{t+h-p-1}}{\hat{\Gamma}}.$$

Substituting out $y_{t+h-p-1} = a^{h-p-1}y_t + a^{h-p-2}\varepsilon_{t+1} + \ldots + a\varepsilon_{t+h-p-2} + \varepsilon_{t+h-p-1}$

$$= o_p(1) + \underbrace{\hat{b}^{(p),GLS}a^{h-p-1}}_{plim=a^h}\underbrace{\sqrt{T-H}(\hat{a}-a)}_{\xrightarrow{d}} + \underbrace{\hat{b}^{(p),GLS}(\hat{a}-a)}_{plim=0}\underbrace{\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\left(a^{h-p-2}\varepsilon_{t+1} + \ldots + a\varepsilon_{t+h-p-2} + \varepsilon_{t+h-p-1}\right)}{\hat{\Gamma}}}_{\xrightarrow{d}},$$

where convergence in distribution is due to the Mixingale Central Limit Theorem. Consequently

$$\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\left[a^p\varepsilon_{t+h-p} - \hat{b}^{(p),GLS}\hat{\varepsilon}_{t+h-p}\right]}{\hat{\Gamma}} = \underbrace{\hat{b}^{(p),GLS}a^{h-p-1}}_{plim=a^{h-1}}\underbrace{\sqrt{T-H}(\hat{a}-a)}_{\xrightarrow{d}} + o_p(1).$$

□

A.2 Proofs of Theorems and Corollaries

A.2.1 Proofs of Theorems

Proof of Theorem 2

Proof. To show consistency of LP GLS it suffices to show that $\| U_{4T} \| \xrightarrow{p} 0$ because

$$\| \hat{B}(k,h,GLS) - B(k,h) \| \leq (\| U_{1T} \| + \| U_{2T} \| + \| U_{3T} \| - \| U_{4T} \|) \| \hat{\Gamma}_k^{-1} \|_1 .$$

From Proposition 2 we know $\| \hat{\Gamma}_k^{-1} \|_1$ is bounded in probability and that $\| U_{1T} \|, \| U_{2T} \|$, and $\| U_{3T} \|$ converge in probability to 0. The proof showing $\| U_{4T} \| \xrightarrow{p} 0$ will be a proof by induction. Assume the consistency for the previous $h-1$ horizons has been proven. Hence $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$ for $1 \leq l \leq h-1$. Note $\hat{\varepsilon}_{t,k} = \varepsilon_t + (\sum_{j=k+1}^{\infty} A_j y_{t-j}) - (\hat{B}(k,1) - B(k,1))X_{t-1,k})$. Therefore

$$U_{4T} = \sum_{l=1}^{h-1} \hat{\Theta}_l \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\varepsilon_{t+h-l} + (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1))X_{t+h-l-1,k})X'_{t,k}\}.$$

By Lemma A.2 part C in Goncalves and Kilian (2007), we know that $\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k}\} \| \xrightarrow{p} 0$, for $1 \leq l \leq h-1$. Since $h-1$ is finite and $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$,

$$\| \sum_{l=1}^{h-1} \hat{\Theta}_l \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k}\} \| \leq \sum_{l=1}^{h-1} \underbrace{\| \hat{\Theta}_l \|}_{bounded} \underbrace{\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k}\} \|}_{plim=0} \xrightarrow{p} 0.$$

To show $\| U_{4T} \| \xrightarrow{p} 0$ it now suffices to show that

$$\| \sum_{l=1}^{h-1} \hat{\Theta}_l \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1))X_{t+h-l-1,k})X'_{t,k}\} \| \xrightarrow{p} 0.$$

Owing to $h-1$ is finite and $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$, this simplifies to showing

$$\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j})X'_{t,k}\} - \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\hat{B}(k,1) - B(k,1))X_{t+h-l-1,k})X'_{t,k}\} \| \xrightarrow{p} 0.$$

By Theorem 1 in Lewis and Reinsel (1985), $\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}))X'_{t,k}\} \| \xrightarrow{p} 0$. Now all that is left to show is $\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\hat{B}(k,1) - B(k,1))X_{t+h-l-1,k})X'_{t,k}\} \| \xrightarrow{p} 0$. Note that

$$\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\hat{B}(k,1) - B(k,1))X_{t+h-l-1,k})X'_{t,k}\} \|$$

$$\leq \underbrace{\| \{(\hat{B}(k,1) - B(k,1)) \|}_{plim=0} \underbrace{\| (T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t+h-l-1,k} X'_{t,k}\} \|_1}_{bounded} \xrightarrow{p} 0.$$

Since this is a proof by induction, it was assumed that the first $h-1$ horizons are consistent, so the first term converges in probability to 0. The second term is bounded due to $\| \hat{\Gamma}_k \|_1 = \| (T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t,k} X'_{t,k} \|_1$ being bounded and since the autocovariances are absolutely summable. It follows that

$$\| \hat{\Theta}_l \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\varepsilon_{t+h-l} + (\sum_{j=1}^{\infty} A_j y_{t+h-l-j}) - (\sum_{i=1}^{k} \hat{A}_i y_{t+h-l-i})))X'_{t,k}\} \| \xrightarrow{p} 0,$$

for each $1 \leq l \leq h-1$. Therefore, $\| U_{4T} \| \xrightarrow{p} 0$. To complete the proof by induction, note that the horizon 1 LP is a VAR, and the consistency results for the VAR were proved in Goncalves and Kilian (2007) (Lemma A.2). □

Proof of Theorem 3

Proof. By Proposition 4 and Lemma 1 we know that

$$\sqrt{T-k-H} l(k)' vec[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p}$$

$$(T-k-H)^{-1/2} l(k)' vec[(\sum_{t=k}^{T-H} \varepsilon_{t+h} X'_{t,k})\Gamma_k^{-1}] + (T-k-H)^{-1/2} l(k)' s_{k,h} vec[(\sum_{t=k}^{T-H} \varepsilon_{t+1} X'_{t,k})\Gamma_k^{-1}].$$

Define

$$r_{t+h}^{(h),GLS} = l(k)' vec[\varepsilon_{t+h} X'_{t,k} \Gamma_k^{-1}] + l(k)' s_{k,h} vec[\varepsilon_{t+1} X'_{t,k} \Gamma_k^{-1}].$$

To show asymptotic normality of $r_{t+h}^{(h),GLS}$ the mixingale CLT will be used. The argument that $\{r_t^{(h),GLS}, \mathcal{F}_t\}$ is an adapted mixingale with $\gamma_i = O_p(i^{-1-\delta})$ for some $\delta > 0$, follows the exact same reasoning as the OLS case (Proposition 3), and is omitted for brevity. Positive definiteness of $\sum_{m=-\infty}^{\infty} cov(r_t^{(h),GLS}, r_{t-m}^{(h),GLS})$ follows the same argument used in the OLS case (Proposition 3). □

Proof of Theorem 4

Proof. Note that $\hat{s}_{k,h}$ can replace $\hat{s}^*_{k,h}$ in $RS\hat{c}ore_t^{(H),*}$ since $plim\{(T-k-H)^{-1/2} \sum_{t=k}^{T-H} RS\hat{c}ore_{t+1}^{(H),*}\}$ is unaffected by the change. To see why note that

$$\| \Big(\sum_{l=1}^{h-1} \{\hat{\Gamma}_k^{-1} \hat{\Gamma}'_{(h-l-1),k} \otimes \hat{\Theta}^*_l - \hat{\Theta}_l\} \Big) (T-k-H)^{-1/2} \sum_{t=k}^{T-H} (\Gamma_k^{-1} X_{t,k} \otimes I_r) \varepsilon_{t+1} \|$$

$$\leq \underbrace{\| (T-k-H)^{1/2}\Big(\sum_{l=1}^{h-1} \{\hat{\Gamma}_k^{-1}\hat{\Gamma}'_{(h-l-1),k} \otimes \hat{\Theta}_l^* - \hat{\Theta}_l\}\Big) \|_1}_{O_p(k^{1/2})} \underbrace{\| (T-k-H)^{-1} \sum_{t=k}^{T-H} (\Gamma_k^{-1} X_{t,k} \otimes I_r)\varepsilon_{t+1} \|}_{O_p(\frac{k^{1/2}}{T^{1/2}})} = O_p(\frac{k}{T^{1/2}}).$$

$\| j(k,H)' \|$ is bounded by assumption and can be ignored. The first term on the last line is $O_p(k^{1/2})$ because $\| \hat{\Gamma}_k^{-1} \|_1$ and $\| \hat{\Gamma}'_{(h-l-1),k} \|_1$ are consistent and bounded in probability and $\| \hat{\Theta}_l^* - \hat{\Theta}_l \| = O_p(\frac{k^{1/2}}{T^{1/2}})$ by Lemma 2. The second term is $O_p(\frac{k^{1/2}}{T^{1/2}})$ by Proposition 2. Since $E^*[R\hat{Score}_t^{(H),*}] = 0$,

$$var^*((T-k-H)^{-1/2} \sum_{t=k}^{T-H} R\hat{Score}_{t+1}^{(H),*}) = (T-k-H)^{-1} \sum_{t=k}^{T-H} R\hat{Score}_{t+1}^{(H)} R\hat{Score}_{t+1}^{(H)\prime} = \hat{V}_{11}(k,H),$$

where the last equality is due to Lemma 3. Moreover, by Lemma 3 $\| \hat{V}_{11}(k,H) \| \xrightarrow{p} \| V_{11}(k,H) \|$. To show asymptotic normality, the mixingale CLT will be used. Conditional on the original data, $R\hat{Score}_t^{(H),*}$ is an adapted mixingale. Let $c_t = (E^*(E^*(R\hat{Score}_t^{(H),*} \big| \mathcal{F}_{t-i}^*)^2))^{1/2}\Delta$, where $\mathcal{F}_{t-1}^* = \sigma(\eta_{t-1}, \eta_{t-2}, \ldots)$ is the bootstrap sigma field conditional on the original data, $\Delta = H^{\nu/(\nu+1)}$ and set $\gamma_i = i^{-(\nu+1)/\nu}$ for any $\nu > 0$. □

Proof of Theorem 5

Proof. To use the mixingale CLT, I need to show:

1. $\{\zeta StrucScore_t^{(H)}, \mathcal{F}_t\}$ is an adapted mixingale with $\gamma_i = O_p(i^{-1-\delta})$ for some $\delta > 0$,
2. $\sum_{p=-\infty}^{\infty} cov(\zeta StrucScore_t^{(H)}, \zeta StrucScore_{t-p}^{(H)}) > 0$.

$\zeta = [\zeta'_{11}, \zeta'_{21}]'$ is a $\{[r(r+1)/2]+1\} \times 1$ Cramer-Wold device where ζ_{11} is a scalar. Note that from Corollary 1 $\{Score_t^{(H)}, \mathcal{F}_t\}$ is an adapted mixingale with $c_t = (E(E(Score_t^{(H)} \big| \mathcal{F}_{t-i})^2))^{1/2}\Delta$, where $\Delta = H^{\nu/(\nu+1)}$ for any $\nu > 0$, and $\gamma_i = i^{-(\nu+1)/\nu}$. By Theorem 3.49 and Lemma 6.16 in White (2001)

$$(E(E(\zeta_{21} vech(\varepsilon_t \varepsilon'_t - \Sigma) \big| \mathcal{F}_{t-i})^2))^{1/2} \leq 2(2^{1/2}+1)\alpha(i)^{1/4}(E[(\zeta_{21} vech(\varepsilon_t \varepsilon'_t - \Sigma))^4])^{1/4}.$$

By Assumption 4, $\alpha(i)^{1/4} = O_p(i^{-(\nu+1)/\nu})$ since $\alpha(m) = O_p(m^{-4(\nu+1)/\nu})$. So $\{\zeta_{21} vech(\varepsilon_t \varepsilon'_t - \Sigma), \mathcal{F}_t\}$ is an adapted mixingale sequence with $c_t = 2(2^{1/2}+1)(E[(\zeta_{21} vech(\varepsilon_t \varepsilon'_t - \Sigma))^4])^{1/4}$ and $\gamma_i = i^{-(\nu+1)/\nu}$. It follows by Minkowski's inequality that $\{\zeta StrucScore_t^{(H)}, \mathcal{F}_t\}$ is an adapted mixingale sequence with

$$c_t = (E(E(\zeta_{11} StrucScore_t^{(H)} \big| \mathcal{F}_{t-i})^2))^{1/2}\Delta + 2(2^{1/2}+1)(E[(\zeta_{21} vech(\varepsilon_t \varepsilon'_t - \Sigma))^4])^{1/4},$$

where $\Delta = H^{\nu/(\nu+1)}$ and $\gamma_i = i^{-(\nu+1)/\nu}$. Therefore $\{\zeta StrucScore_t^{(H)}, \mathcal{F}_t\}$ is a mixingale of size $\gamma_i = O_p(i^{-(\nu+1)/\nu})$.

Lastly, the proof $\sum_{p=-\infty}^{\infty} cov(\zeta StrucScore_t^{(H)}, \zeta StrucScore_{t-p}^{(H)}) > 0$ follows the same argument as Proposition 3 if it can be shown that $StrucScore_t^{(H)}$ has absolutely summable autocovariances. We already know $Score_t^{(H)}$ and $\varepsilon_t \varepsilon'_t$ each have absolutely summable autocovariances. To show absolute summability of the autocovariances for $StrucScore_t^{(H)}$, it suffices to show $\sum_{p=-\infty}^{\infty} \parallel cov(Score_t^{(H)}, vech(\varepsilon_{t-p}\varepsilon'_{t-p} - \Sigma)') \parallel < \infty$. Since $X'_{t,k} = [\sum_{j=0}^{\infty}(\varepsilon'_{t-j}\Theta'_j, \ldots, \varepsilon'_{t-k+1-j}\Theta'_j)]$,

$$\sum_{p=-\infty}^{\infty} \parallel cov(Score_t^{(H)}, vech(\varepsilon_{t-p}\varepsilon'_{t-p} - \Sigma)') \parallel$$

$$= \sum_{p=-\infty}^{\infty} \parallel E\{j(k,H)' \begin{bmatrix} (\Gamma_k^{-1} \otimes I_r) vec([\sum_{j=0}^{\infty}(\varepsilon_{t+H}\varepsilon'_{t-j-1}\Theta'_j, \ldots, \varepsilon_{t+H}\varepsilon'_{t-j-k}\Theta'_j)]) \\ s_{k,H} vec([\sum_{j=0}^{\infty}(\varepsilon_{t+1}\varepsilon'_{t-j-1}\Theta'_j, \ldots, \varepsilon_{t+1}\varepsilon'_{t-j-k}\Theta'_j)]) \\ \vdots \\ (\Gamma_k^{-1} \otimes I_r) vec([\sum_{j=0}^{\infty}(\varepsilon_{t+2}\varepsilon'_{t-j-1}\Theta'_j, \ldots, \varepsilon_{t+2}\varepsilon'_{t-j-k}\Theta'_j)]) \\ s_{k,2} vec([\sum_{j=0}^{\infty}(\varepsilon_{t+1}\varepsilon'_{t-j-1}\Theta'_j, \ldots, \varepsilon_{t+1}\varepsilon'_{t-j-k}\Theta'_j)]) \\ (\Gamma_k^{-1} \otimes I_r) vec([\sum_{j=0}^{\infty}(\varepsilon_{t+1}\varepsilon'_{t-j-1}\Theta'_j, \ldots, \varepsilon_{t+1}\varepsilon'_{t-j-k}\Theta'_j)]) \end{bmatrix} vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} - \Sigma \end{bmatrix}' L_r\} \parallel .$$

Since $\parallel (\Gamma_k^{-1} \otimes I_r) \parallel$, $\parallel s_{k,h} \parallel$, and $\parallel \Sigma \parallel$ are bounded and H is finite, it suffices to show that

$$\sum_{p=-\infty}^{\infty} \parallel E\{vec([\sum_{j=0}^{\infty}(\varepsilon_{t+h}\varepsilon'_{t-j-1}\Theta_j, \ldots, \varepsilon_{t+h}\varepsilon'_{t-j-k}\Theta_j)]) vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} \end{bmatrix}'\} \parallel$$

$$\leq \sum_{s=1}^{k} \sum_{p=-\infty}^{\infty} \sum_{j=0}^{\infty} \parallel E\{vec(\varepsilon_{t+h}\varepsilon'_{t-j-s}\Theta_j) vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} \end{bmatrix}'\} \parallel < \infty,$$

for $h = 1, 2, \ldots, H$. Note that $E\{vec(\varepsilon_{t+h}\varepsilon'_{t-j-s}\Theta_j) vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} \end{bmatrix}'\}$, is a $r^2 \times [r(r+1)/2]$ matrix. For ease of notation, assume for the rest of this proof $r = 1$ so that ε and Θ are scalars, and note that for the multivariate case when $r > 1$, since r is finite, the sum of $r^2 \times [r(r+1)/2]$ finite constants is finite. Now

$$\sum_{s=1}^{k} \sum_{p=-\infty}^{\infty} \sum_{j=0}^{\infty} \parallel E\{vec(\varepsilon_{t+h}\varepsilon'_{t-j-s}\Theta_j) vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} \end{bmatrix}'\} \parallel$$

$$= \sum_{s=1}^{k} \sum_{p=-\infty}^{\infty} \parallel E\{vec(\varepsilon_{t+h}\varepsilon'_{t-s}\Theta_0) vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} \end{bmatrix}'\} \parallel + \sum_{s=1}^{k} \sum_{p=-\infty}^{\infty} \parallel E\{vec(\varepsilon_{t+h}\varepsilon'_{t-1-s}\Theta_1) vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} \end{bmatrix}'\} \parallel + \ldots$$

$$\leq \parallel \Theta_0 \parallel \underbrace{\sum_{s=1}^{k} \sum_{p=-\infty}^{\infty} \parallel E\{vec(\varepsilon_{t+h}\varepsilon'_{t-s}) vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} \end{bmatrix}'\} \parallel}_{bounded} + \parallel \Theta_1 \parallel \underbrace{\sum_{s=1}^{k} \sum_{p=-\infty}^{\infty} \parallel E\{vec(\varepsilon_{t+h}\varepsilon'_{t-1-s}) vech\begin{bmatrix} \varepsilon_{t-p}\varepsilon'_{t-p} \end{bmatrix}'\} \parallel}_{bounded} + \ldots$$

$$\leq \sum_{j=0}^{\infty} \parallel \Theta_j \parallel \times constant < \infty,$$

where the boundedness is due to absolute summability of the fourth order cumulants (which implies the absolute summability of the fourth order moments). □

Proof of Theorem 6

Proof. Using same argument as Theorem 4, $\hat{s}_{k,h}$ can replace $\hat{s}^*_{k,h}$ in $RStru\hat{c}Score^{(H),*}_{t+1}$. Due to independence of the blocks, $var^*((T-k-H)^{-1/2}\sum_{t=k}^{T-H} RStru\hat{c}Score^{(H),*}_{t+1}) = \hat{V}^{lr}(k,H)$ by Lemma 4. Moreover, by Lemma 4 we know that $\| \hat{V}^{lr}(k,H) \| \overset{p}{\to} \| V(k,H) \|$. Note that

$$(T-k-H)^{-1/2}\sum_{t=k}^{T-H} RStru\hat{c}Score^{(H),*}_{t+1} = \sum_{j=1}^{N}(N)^{-1/2}(\ell)^{-1/2}\sum_{s=1}^{\ell} RStru\hat{c}Score^{(H),*}_{k+s+(j-1)\ell} = plim\sum_{j=1}^{N} Q^*_j,$$

where $Q^*_j = (N)^{-1/2}(\ell)^{-1/2}\sum_{s=1}^{\ell} RStru\hat{c}Score^{(H),*}_{k+s+(j-1)\ell}$. It follows that $E^*[Q^*_j] = 0$, $E^*[Q^*_j Q^*_s] = 0$ for $s \neq j$, and $E^*[Q^*_j Q^*_j] = \frac{\hat{V}^{lr}(k,H)}{N}$. To show asymptotic normality, I will use the CLT for triangular arrays of independent random variables, Theorem 27.3 in Billingsley (1995). Need to show

1. $$\sum_{j=1}^{N} E^*(Q^*_j Q^{*\prime}_j) \overset{p^*}{\to} V(k,H),$$

2. $$\frac{\sum_{j=1}^{N} E^*(|\, \zeta Q^*_j \,|^{2+\xi})}{[\sum_{j=1}^{N} E^*((\zeta Q^*_j)^2)]^{(2+\xi)/2}} \overset{p^*}{\to} 0,$$

where ζ is a $\{[r(r+1)/2]+1\} \times 1$ Cramer-Wold device. The first condition has already been proven. For Lyapunov condition, set $\xi = 2$. The denominator is bounded since

$$E^*[(\zeta Q^*_j)^2] = (N)^{-1}\underbrace{(\ell)^{-1}\sum_{n=1}^{\ell}\sum_{m=1}^{\ell}\zeta RStru\hat{c}Score^{(H)}_{m+(r-1)\ell}(\zeta RStru\hat{c}Score^{(H)}_{n+(r-1)\ell})'}_{\zeta\hat{V}^{lr}(k,H)\zeta' = O_p(1)} = O_p(N^{-1}).$$

Therefore

$$\sum_{j=1}^{N} E^*[(\zeta Q^*_j)^2] = O_p(1) \implies [\sum_{j=1}^{N} E^*((\zeta Q^*_j)^2)]^2 = O_p(1).$$

To show the numerator converges to probability 0, note that since η has finite fourth moments and $(\zeta Q^*_j)^2 = O_p(N^{-1})$,

$$\sum_{j=1}^{N} E^*(|\, \zeta Q^*_j \,|^4) = \sum_{j=1}^{N} E^*[(\zeta Q^*_j)^2(\zeta Q^*_j)^2] = O_p(N^{-1}) \overset{p}{\to} 0.$$

□

Proof of Theorem 7

Proof. Let

$$y_{t+1} = a y_t + \varepsilon_{t+1},$$

$|a| < 1$ and ε_t is an i.i.d. process with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma^2$. This implies that $E(y_t) = 0$ and the $var(y_t) = E(y_t' y_t) = \frac{\sigma^2}{(1-a^2)}$. The LP GLS model at horizon h is:

$$y_{t+h} - \hat{b}^{(h-1),GLS}\hat{\varepsilon}_{t+1} - \ldots - \hat{b}^{(1),GLS}\hat{\varepsilon}_{t+h-1} = b^{(h)} y_t + \widetilde{u}_{t+h}^{(h)}.$$

Note that

$$lim_{T\to\infty} var[\sqrt{T-H}(\hat{b}^{(h),OLS} - a^h)]$$

$$= lim_{T\to\infty}\{var[\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h) - \sqrt{T-H}(\hat{b}^{(h),GLS} - \hat{b}^{(h),OLS})]$$

$$= lim_{T\to\infty}\{var[\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h)] + var[\sqrt{T-H}(\hat{b}^{(h),OLS} - \hat{b}^{(h),GLS})]$$

$$+2cov[\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h), \sqrt{T-H}(\hat{b}^{(h),OLS} - \hat{b}^{(h),GLS})]\}.$$

In order to show that the GLS estimator is at least as efficient, it suffices to show that

$$lim_{T\to\infty}\{2cov[\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h), \sqrt{T-H}(\hat{b}^{(h),OLS} - \hat{b}^{(h),GLS})]\} \geq 0.$$

By Proposition 6

$$\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h) \xrightarrow{p} (h-1)a^{h-1}\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+1}}{\hat{\Gamma}} + \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+h}}{\hat{\Gamma}}.$$

By Proposition 6, we also know that

$$\sqrt{T-H}[\hat{b}^{(h),OLS} - \hat{b}^{(h),GLS}] \xrightarrow{p} \frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t(\sum_{p=1}^{h-1} a^p\varepsilon_{t+h-p})}{\hat{\Gamma}} - (h-1)a^{h-1}\frac{\frac{1}{\sqrt{T-H}}\sum_{t=1}^{T-H} y_t\varepsilon_{t+1}}{\hat{\Gamma}}.$$

So

$$lim_{T\to\infty} cov[\sqrt{T-H}[\hat{b}^{(h),OLS} - \hat{b}^{(h),GLS}], \sqrt{T-H}\{\hat{b}^{(h),GLS} - a^h\}]$$

$$= E[(h-1)a^{h-1}\frac{\frac{1}{\sqrt{T-H}}\sum_{m=1}^{T-H} y_m\varepsilon_{m+1}}{\hat{\Gamma}} \times \frac{\frac{1}{\sqrt{T-H}}\sum_{n=1}^{T-H} y_n(\sum_{p=1}^{h-1} a^p\varepsilon_{n+h-p})}{\hat{\Gamma}}]$$

$$-E[(h-1)a^{h-1}\frac{\frac{1}{\sqrt{T-H}}\sum_{m=1}^{T-H} y_m\varepsilon_{m+1}}{\hat{\Gamma}} \times (h-1)a^{h-1}\frac{\frac{1}{\sqrt{T-H}}\sum_{n=1}^{T-H} y_n\varepsilon_{n+1}}{\hat{\Gamma}}]$$

$$+E[\frac{\frac{1}{\sqrt{T-H}}\sum_{m=1}^{T-H} y_m\varepsilon_{m+h}}{\hat{\Gamma}} \times \frac{\frac{1}{\sqrt{T-H}}\sum_{n=1}^{T-H} y_n(\sum_{p=1}^{h-1} a^p\varepsilon_{n+h-p})}{\hat{\Gamma}}]$$

$$-E[\frac{\frac{1}{\sqrt{T-H}}\sum_{m=1}^{T-H} y_m\varepsilon_{m+h}}{\hat{\Gamma}} \times (h-1)a^{h-1}\frac{\frac{1}{\sqrt{T-H}}\sum_{n=1}^{T-H} y_n\varepsilon_{n+1}}{\hat{\Gamma}}].$$

Since $\hat{\Gamma} \xrightarrow{p} \frac{\sigma^2}{(1-a^2)}$, we have

$$= lim_{T\to\infty}\Big(\frac{(1-a^2)}{\sigma^2}\Big)^2\{(h-1)a^{h-1}\frac{1}{T-H}E[\sum_{m=1}^{T-H}\sum_{n=1}^{T-H} y_m\varepsilon_{m+1}y_n(\sum_{p=1}^{h-1} a^p\varepsilon_{n+h-p})]$$

$$-a^{2(h-1)}(h-1)^2\frac{1}{T-H}E[\sum_{m=1}^{T-H}\sum_{n=1}^{T-H} y_m\varepsilon_{m+1}y_n\varepsilon_{n+1}]$$

$$+\frac{1}{T-H}E[\sum_{m=1}^{T-H}\sum_{n=1}^{T-H} y_m\varepsilon_{m+h}y_n(\sum_{p=1}^{h-1} a^p\varepsilon_{n+h-p})] - \frac{1}{T-H}a^{h-1}(h-1)E[\sum_{m=1}^{T-H}\sum_{n=1}^{T-H} y_m\varepsilon_{m+h}y_n\varepsilon_{n+1}]\}.$$

$$= \Big(\frac{(1-a^2)}{\sigma^2}\Big)^2\{(h-1)a^{h-1}\sum_{p=1}^{h-1} a^{h-1}\frac{\sigma^4}{(1-a^2)} - a^{2(h-1)}(h-1)^2\frac{\sigma^4}{(1-a^2)} + \sum_{p=1}^{h-1} a^{2p}\frac{\sigma^4}{(1-a^2)} - a^{2(h-1)}(h-1)\frac{\sigma^4}{(1-a^2)}\}.$$

$$= \underbrace{(1-a^2)}_{positive}\underbrace{\{(\sum_{p=1}^{h-1} a^{2p}) - a^{2(h-1)}(h-1)\}}_{non-negative},$$

where the second to last line is due to independence of the errors. Note that the last term is non-negative since

$$\frac{(\sum_{p=1}^{h-1} a^{2p})}{a^{2(h-1)}(h-1)} = \frac{\sum_{p=1}^{h-1} a^{2(p-h+1)}}{h-1} \geq 1,\ for\ h = 2, 3, \ldots$$

where the inequality is due to $p+1 \leq h$ and $|a| < 1$. Therefore GLS is more efficient since

$$lim_{T\to\infty} var[\sqrt{T-H}(\hat{b}^{(h),OLS} - a^h)]$$

$$= lim_{T\to\infty}\{\underbrace{var[\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h)]}_{positive} + \underbrace{var[\sqrt{T-H}(\hat{b}^{(h),OLS} - \hat{b}^{(h),GLS})]}_{positive}$$

$$+ \underbrace{2cov[\sqrt{T-H}(\hat{b}^{(h),GLS} - a^h), \sqrt{T-H}(\hat{b}^{(h),OLS} - \hat{b}^{(h),GLS})]\}}_{non-negative}.$$

□

A.2.2 Proofs of Corollaries

Proof of Corollary 1

Proof. Proof follows the exact same lines as Proposition 3, so only the broad strokes will be discussed. Note that when $i > H-1$, $E[Score_t^{(H)}\big|\mathcal{F}_{t-i}] = 0$ by the martingale difference sequence assumption on the errors. $c_t = (E(E(Score_t^{(H)}\big|\mathcal{F}_{t-i})^2))^{1/2}\Delta$ where $\Delta = H^{\nu/(\nu+1)}$ for any $\nu > 0$, and $\gamma_i = i^{-(\nu+1)/\nu}$. Lastly, the proof $V_{11}(k,H) > 0$ follows the same argument as Proposition 3. □

A.2 Monte Carlo Evidence Figures

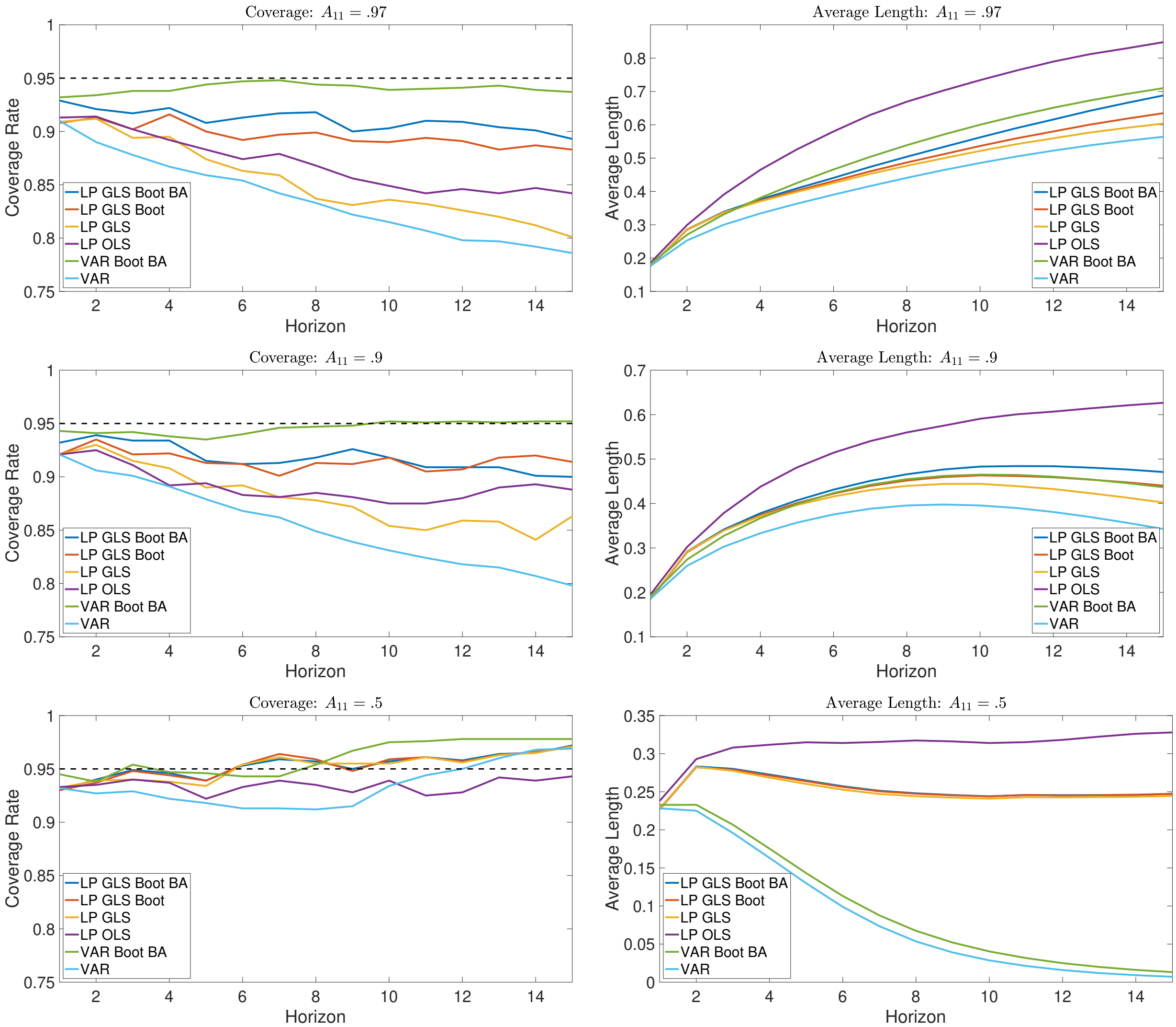

Figure 4: Coverage Rates for 95% Confidence Intervals and Average Length for VAR(1) Models

Note: Bias-adjusted LP GLS bootstrap (LP GLS Boot BA), LP GLS bootstrap (LP GLS Boot), Analytical LP GLS estimator (LP GLS), LP OLS with equal-weighted cosine HAC standard errors (LP OLS), Bias-adjusted VAR bootstrap (VAR Boot BA), Analytical VAR estimator (VAR).

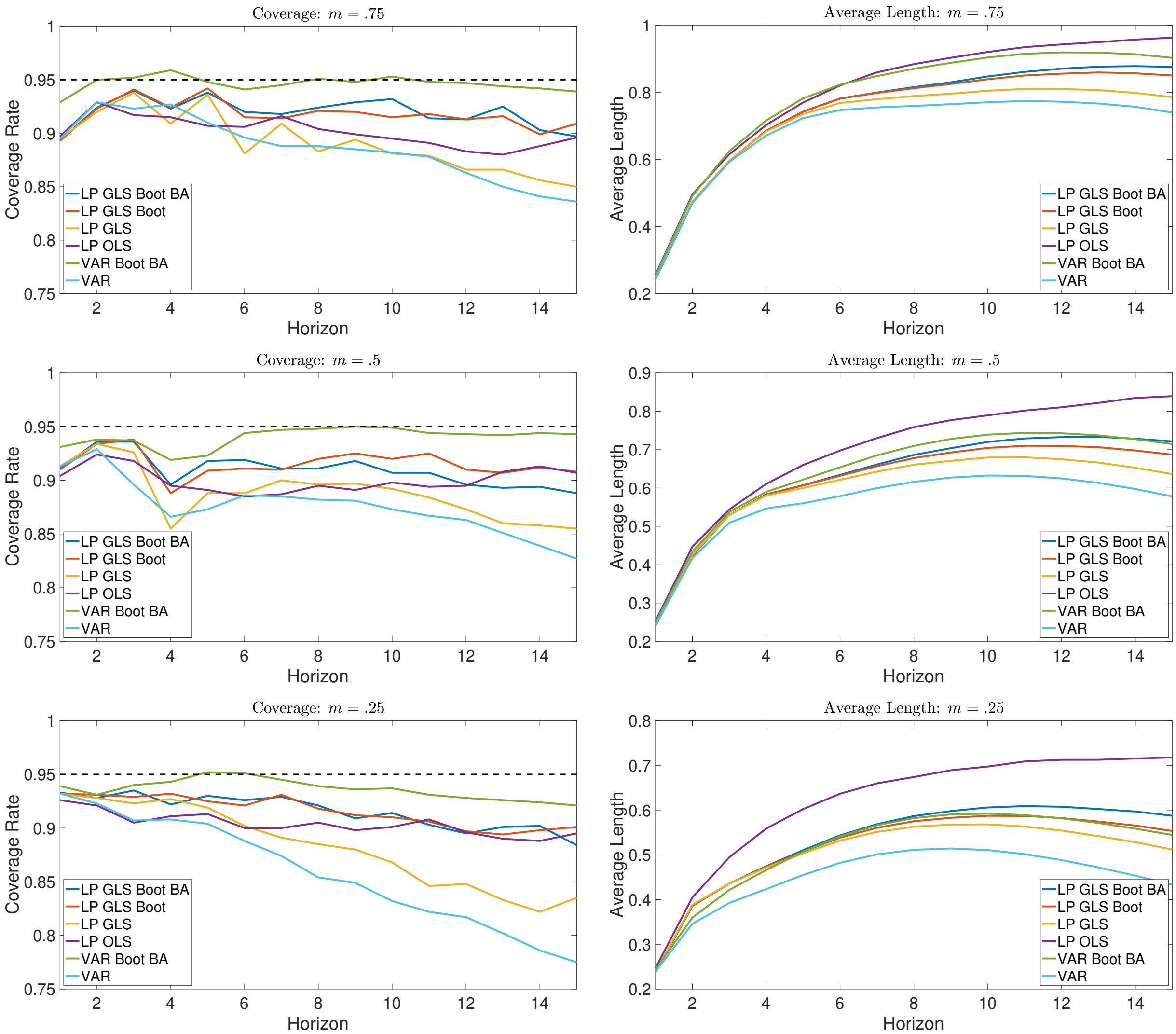

Figure 5: Coverage Rates for 95% Confidence Intervals and Average Length for ARMA(1,1) Models

Note: Bias-adjusted LP GLS bootstrap (LP GLS Boot BA), LP GLS bootstrap (LP GLS Boot), Analytical LP GLS estimator (LP GLS), LP OLS with equal-weighted cosine HAC standard errors (LP OLS), Bias-adjusted VAR bootstrap (VAR Boot BA), Analytical VAR estimator (VAR).

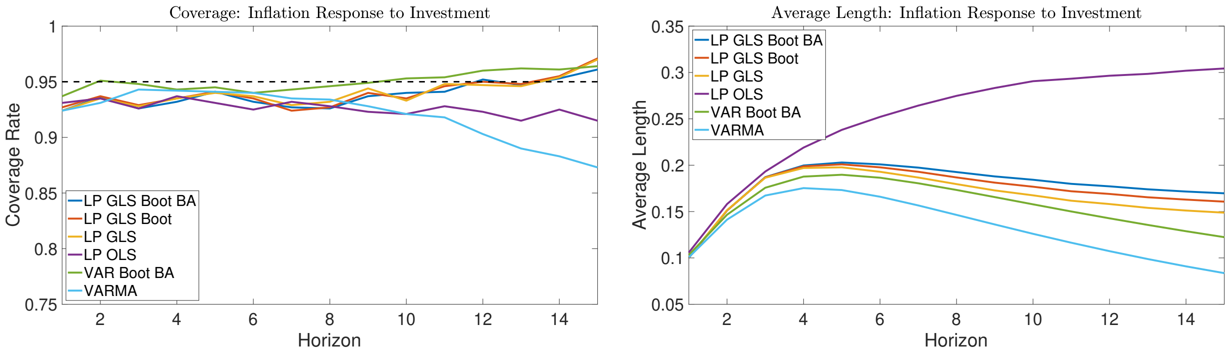

Figure 6: Coverage Rates for 95% Confidence Intervals and Average Length for VARMA(1,1)

Note: Bias-adjusted LP GLS bootstrap (LP GLS Boot BA), LP GLS bootstrap (LP GLS Boot), Analytical LP GLS estimator (LP GLS), LP OLS with equal-weighted cosine HAC standard errors (LP OLS), Bias-adjusted VAR bootstrap (VAR Boot BA), Analytical VAR estimator (VAR).

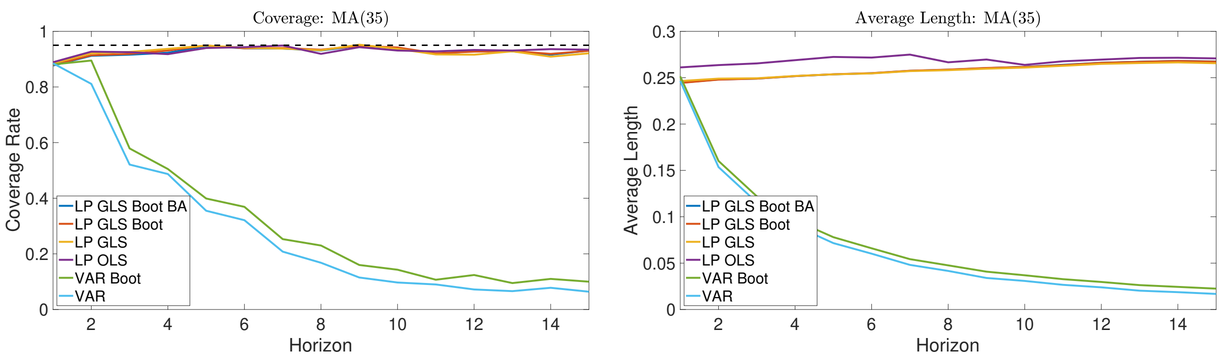

Figure 7: Coverage Rates for 95% Confidence Intervals and Average Length for MA(35)

Note: Bias-adjusted LP GLS bootstrap (LP GLS Boot BA), LP GLS bootstrap (LP GLS Boot), Analytical LP GLS estimator (LP GLS), LP OLS with equal-weighted cosine HAC standard errors (LP OLS), Bias-adjusted VAR bootstrap (VAR Boot BA), Analytical VAR estimator (VAR).

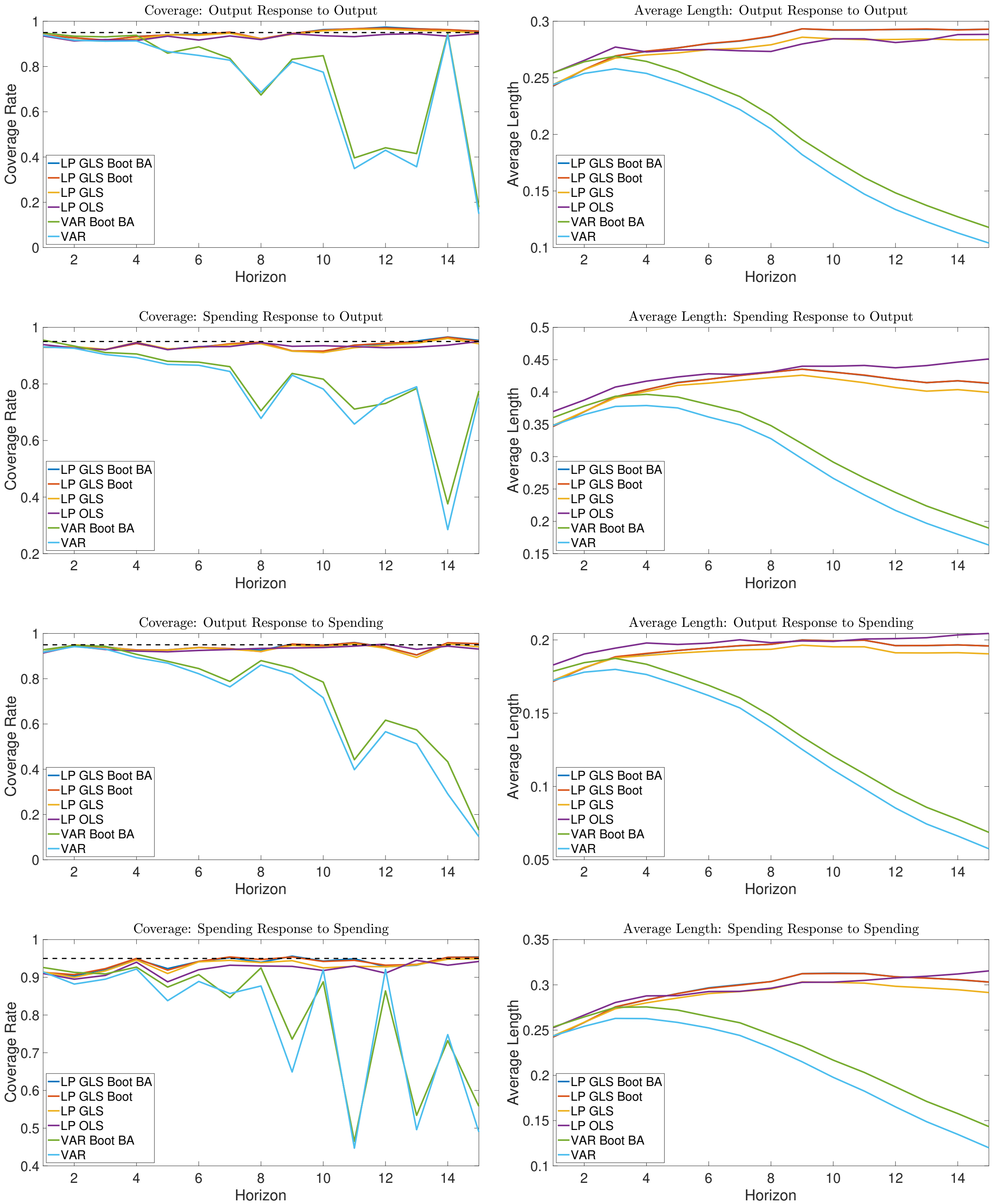

Figure 8: Coverage Rates for 95% Confidence Intervals and Average Length for Fiscal VAR

Note: Bias-adjusted LP GLS bootstrap (LP GLS Boot BA), LP GLS bootstrap (LP GLS Boot), Analytical LP GLS estimator (LP GLS), LP OLS with equal-weighted cosine HAC standard errors (LP OLS), Bias-adjusted VAR bootstrap (VAR Boot BA), Analytical VAR estimator (VAR).

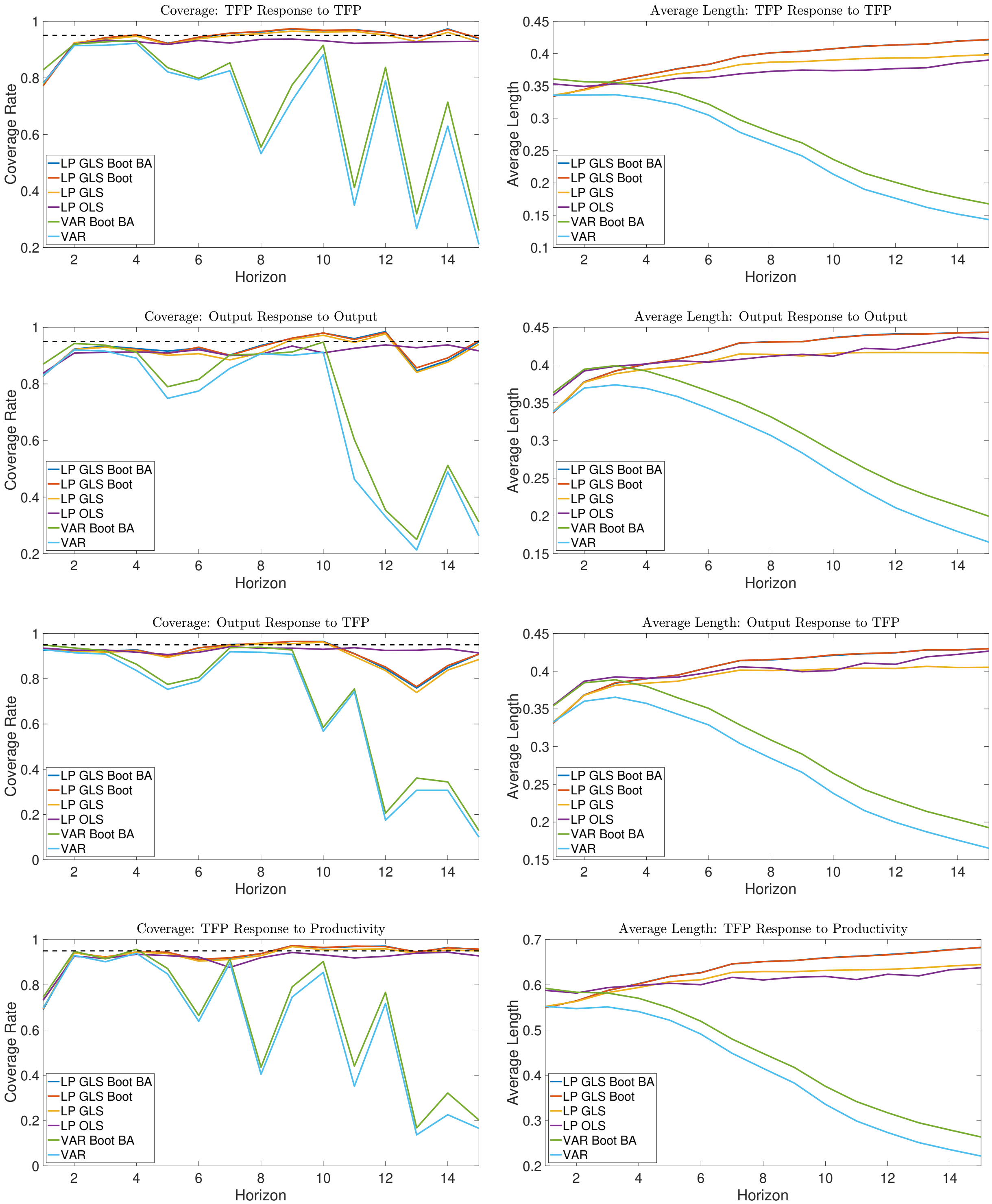

Figure 9: Coverage Rates for 95% Confidence Intervals and Average Length for Technology VAR

Note: Bias-adjusted LP GLS bootstrap (LP GLS Boot BA), LP GLS bootstrap (LP GLS Boot), Analytical LP GLS estimator (LP GLS), LP OLS with equal-weighted cosine HAC standard errors (LP OLS), Bias-adjusted VAR bootstrap (VAR Boot BA), Analytical VAR estimator (VAR).