Online Appendix

July 23, 2023

This online appendix presents a number of complementary results. Section A.1 presents auxiliary propositions and lemmas for theorems and corollaries in the paper, while A.2 presents proofs of the theorems and corollaries. A.3 discusses why it's possible to extend LP to handle time-varying parameters. A.4 presents additional robustness checks and Monte Carlos. A.5 discusses bootstrapping theory and inference with LP GLS. A.6 discusses structural identification. A.7 presents an application to Gertler and Karadi (2015). Lastly, A.8 is a how to section for the code.

Preliminaries

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, and assuming that $\Omega(k,h,OLS)$ is positive definite, then 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 by showing

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

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 $\delta = 1/\nu$. □

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

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

Proof. To show

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

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 $k^{1/2} \| \hat{\Gamma}_k^{-1} - \Gamma_k^{-1} \|_1 = o_p(1)$ (see their Lemma A.1). From Proposition 2 in Jordà and Kozicki (2011), we know that $\| \sqrt{T-k-H} U_{1T}\Gamma_k^{-1} \| \xrightarrow{p} 0$. So to complete the proof, I just need to show

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

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

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

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

$$\times \{k^{1/2} \| (\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \|_1\}.$$

By Theorem 2, $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$ 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 $1 \le l \le h-1$.

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

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

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

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

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

$$\le \underbrace{\Big(\frac{T-k-H}{k}\Big)^{1/2} \parallel (\hat{B}(k,1) - B(k,1)) \parallel}_{bounded} \underbrace{\parallel (T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t+h-l-1,k} X'_{t,k} \parallel_1}_{bounded} .$$

□

Proposition 5. *Under assumption 4,*

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

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

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

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

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

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

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

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

It follows that

$$\| \sqrt{T-k-H}vech[\hat{\Sigma}-\Sigma]-\sqrt{T-k-H}vech\left[\frac{\sum_{t=k}^{T-H}\varepsilon_t\varepsilon_t'-\Sigma}{T-k-H}\right]\|=o_p(1).$$

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, V_{22} is positive definite by assumption. □

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

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

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

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 converge to a normal distribution due to the Mixingale CLT (see proof of Proposition 3 for setup). Therefore

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

□

A.1.2 Lemmas

Lemma 1. *If Assumption 3 holds,*

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

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

Proof. From Proposition 4 we know that

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

and

$$(T-k-H)^{1/2}l(k)'vec[U_{2T}\Gamma_k^{-1}+U_{3T}\Gamma_k^{-1}-U_{4T}\Gamma_k^{-1}]$$
$$= (T-k-H)^{-1/2}l(k)'vec\Big\{(\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k})\Gamma_k^{-1}+\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l})X'_{t,k}\Gamma_k^{-1}-\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\hat{\Theta}_l\hat{\varepsilon}_{t+h-l})X'_{t,k}\Gamma_k^{-1}\Big\}.$$

Note that

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

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

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

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

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

and

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

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

□

Lemma 2. *Under Assumption 5, 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

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

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

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

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,*

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

Proof. Let $\hat{V}_{11}(k,H) = (T-k-H)^{-1}\sum_{t=k}^{T-H} R\hat{Score}^{(H)}_{t+1} R\hat{Score}^{(H)'}_{t+1}$. 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 $\parallel \hat{s}_{k,h} \parallel \xrightarrow{p} \parallel s_{k,h} \parallel < \infty$, $\parallel \hat{\Gamma}_k^{-1} \parallel \xrightarrow{p} \parallel \Gamma_k^{-1} \parallel < \infty$, $\parallel F_{k,m,n} \parallel < \infty$, $\parallel l(k)' \parallel < \infty$, and $2H+1$ is finite, then showing $\parallel \hat{V}_{11}(k,H) \parallel \xrightarrow{p} \parallel V_{11}(k,H) \parallel$ simplifies to showing $\parallel \hat{F}_{k,m,n} - F_{k,m,n} \parallel \xrightarrow{p} 0$ 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,*

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

where

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

$$\hat{V}^{lr}_{11}(k,H) = \sum_{p=-\ell}^{\ell}(T-k-H)^{-1}\sum_{t=k}^{T-H} R\hat{score}^{(H)}_{t+1} R\hat{score}^{(H)'}_{t+1-p},$$

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

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

Proof. First note that

$$\| [(T-k-H)^{-1}\sum_{t=k+1}^{T-H+1}RStrucScore_t^{(H)}RStrucScore_{t-p}^{(H)}{}'] - E[RStrucScore_t^{(H)}RStrucScore_{t-p}^{(H)}{}'] \|$$

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

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 $RStrucScore_t^{(H)} = [RScore_t^{(H)}{}', vech(\varepsilon_{t+1}\varepsilon'_{t+1}-\Sigma)']'$, 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

$$\| \sum_{p=-\ell}^{\ell}\{[(T-k-H)^{-1}\sum_{t=k+1}^{T-H+1}RStrucScore_t^{(H)}RStrucScore_{t-p}^{(H)}{}'] - E[RStrucScore_t^{(H)}RStrucScore_{t-p}^{(H)}{}']\} \|$$

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

Therefore, I just need to show

$$\| \sum_{p=-\ell}^{\ell}(T-k-H)^{-1}\sum_{t=k+1}^{T-H+1}[RStr\hat{u}cScore_t^{(H)}RStr\hat{u}cScore_{t-p}^{(H)}{}' - RStrucScore_t^{(H)}RStrucScore_{t-p}^{(H)}{}'] \| \xrightarrow{p} 0.$$

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

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

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

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

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

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

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

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

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

Proof.

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

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

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

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

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

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

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

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

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

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

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

□

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 $\| U_{4T} \| \overset{p}{\to} 0$ 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 $1 \le l \le h-1$. 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 $1 \le l \le h-1$. Since $h-1$ is finite and $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$,

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

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

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

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

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 $1 \le l \le h-1$. 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). $\square$

Proof of Theorem 3

Proof. By Lemma 1 and proposition 5 we know that

$$\| \left(l(k,H)' \begin{bmatrix} \sqrt{T-k-H} vec[\hat{B}(k,H,GLS)-B(k,H)] \\ \vdots \\ \sqrt{T-k-H} vec[\hat{B}(k,2,GLS)-B(k,2)] \\ \sqrt{T-k-H} vec[\hat{B}(k,1,OLS)-B(k,1)] \end{bmatrix} \atop \sqrt{T-k-H} vech[\hat{\Sigma}-\Sigma] \right) - (T-k-H)^{-1/2} \sum_{t=k}^{T-H} StrucScore_{t+H}^{(H)} \| = o_p(1).$$

where

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

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 $\delta > 0$

$\zeta = [\zeta_{11}', \zeta_{21}']'$ 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 $\gamma_i = O_p(i^{-(\nu+1)/\nu})$. $\square$

Proof of Theorem 4

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

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

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

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

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

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

□

Proof of Theorem 5

Proof. Note that $\hat{s}_{k,h}$ can replace $\hat{s}^*_{k,h}$ in $RStru\hat{c}Score^{(H),*}_{t+1}$ since $plim\{(T-k-H)^{-1/2}\sum_{t=k}^{T-H}RStru\hat{c}Score^{(H),*}_{t+1}\}$ 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.

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

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

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 $E^*[Q^*_j] = 0$, $E^*[Q^*_jQ^*_s] = 0$ for $s \neq j$, and $E^*[Q^*_jQ^*_j] = \frac{\hat{V}^{lr}(k,H)}{N}$. 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. $$\sum_{j=1}^{N} E^*(Q_j^* Q_j^{*\prime}) \xrightarrow{p^*} V(k,H),$$

2. $$\frac{\sum_{j=1}^{N} E^*(|\,\zeta' Q_j^*\,|^{2+\xi})}{[\sum_{j=1}^{N} E^*((\zeta' Q_j^*)^2)]^{(2+\xi)/2}} \xrightarrow{p^*} 0,$$

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

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

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 in probability to 0, note that since η has finite fourth moments and $(\zeta' Q_j^*)^2 = O_p(N^{-1})$,

$$\sum_{j=1}^{N} E^*(|\,\zeta' Q_j^*\,|^4) = \sum_{j=1}^{N} E^*[(\zeta' Q_j^*)^2 (\zeta' Q_j^*)^2] = O_p(N^{-1}) \xrightarrow{p} 0.$$

□

A.3 Structural Breaks and Time-Varying Parameter LP

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.

It is worth reiterating that the GLS procedure presented in Section 2 and the consistency and asymptotic normality of the procedure assumes stationarity.[1] Nonstationarity can be caused by unit roots or structural 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).[2] 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 $\widetilde{y}_{t+h}^{(h)} = y_{t+h} - \hat{B}_{1,t}^{(h-1)}\hat{\varepsilon}_{t+1,k} - \ldots - \hat{B}_{1,t}^{(1)}\hat{\varepsilon}_{t+h-1,k}$. 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

[1] 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.

[2] 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).

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 may 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 better job capturing the time variation in the impulse responses at each horizon.[3]

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.[4] 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.

A.4 Robustness Checks and Additional Monte Carlos

The coverage distortion results for the bootstrap and analytical VARs in the empirically calibrated VAR(16) Monte Carlos is alarming. Due too sample size limitations, it could be the case that the some of the estimates of the empirically calibrated VAR(16)'s are erratic and are the source of the distortions. As a robustness check, I generate the data using a VAR(10), VAR(12), and VAR(14), and estimated the data for all of the models using 8 lags. For the fiscal VAR when the data was generated using the VAR(10), coverage could drop as low as approximately 80%, but would be at least in the mid 80's overall. Distortions show up significantly when estimating the VAR(12) and VAR(14). For the technology VAR, there are coverage distortions for some parameters and horizons for the VAR(10) and they only get worse for the VAR(12) and VAR(14). As another robustness check, I also generate the data using a VAR(8), then estimate the models using 4 lags and then 6 lags. The severe coverage distortions for the fiscal and technology VAR exist when both 4 and 6 lags are used in estimation.

As discussed in Poskitt and Yao (2017), the two sources of estimation error for structural impulse responses identified via long-run restrictions can be broken down into "truncation bias" and "identification

[3] This will depend on the nature of the time variation in the Wold representation and how time variation is modeled when estimating (i.e. random walk, nonparametric, etc).

[4] 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.

bias", which we can think of as estimation error for the Wold coefficients and estimation error for the long-run identification restriction. To help pinpoint how this affects the Poskitt and Yao (2017) Monte Carlo, I present the results for just the Wold coefficients. Select results are presented in Figure 1.

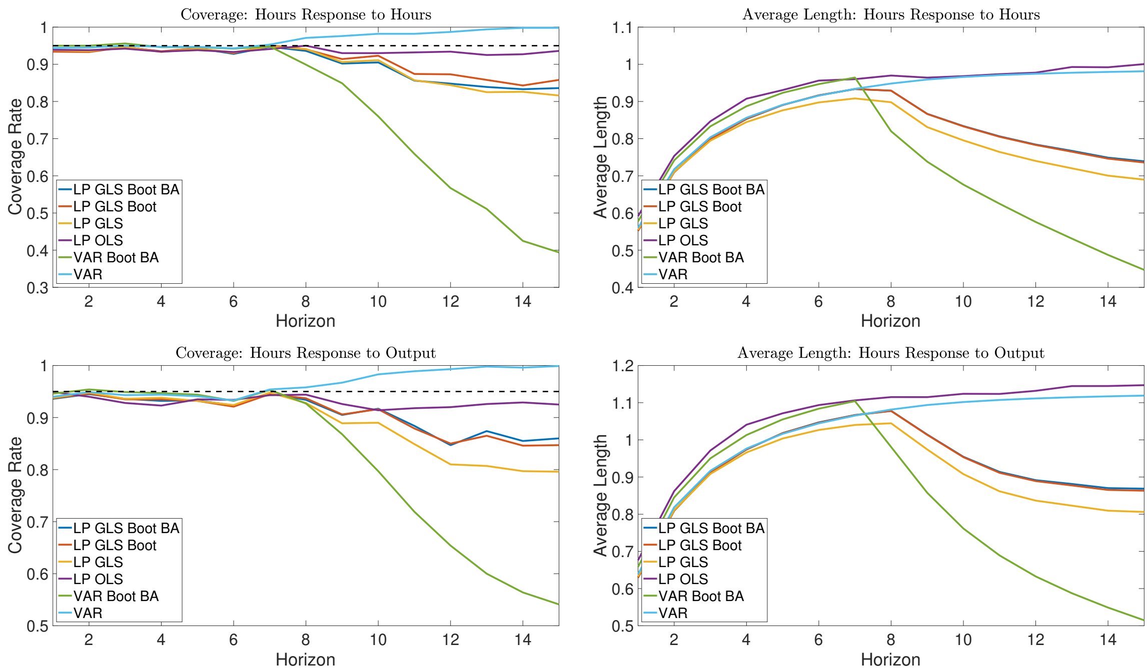

Figure 1: Coverage Rates for 95% Confidence Intervals and Average Length for RBC VARMA

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).

LP OLS and the analytical VAR have the best coverage, which is at or near the nominal level throughout. The different LP GLS bootstrap estimators have close to nominal coverage for around the first 8 or so horizons for all of the impulse responses, but at longer horizons coverage would dip for 2 of the four impulse responses to the mid 80s for the non bias adjusted estimator and low to mid 80s for the bias adjusted estimator.[5] For the analytical LP GLS estimator, coverage would be at or near the nominal level for the first 8 or so horizons for all of the impulse responses, while it could drop as low as approximately 80% for higher horizons for 2 of the 4 impulse responses. The VAR bootstrap had serious coverage distortions with coverage falling as low as 38% for one impulse response and 55% for another. The VAR bootstrap had the shortest average length throughout followed by the LP GLS estimators. The average length of the analytic VAR and LP OLS, however, could be significantly larger than all of the other estimators at longer horizons.

Inspired by Barnichon and Matthes (2018), Jordà et al. (2020), I include the following MA(35) generated using gaussian basis functions:

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

[5]Additional Monte Carlo evidence shows that the dip is only temporary and coverage returns to approximately nominal level after horizon 15.

where

$$\theta_i = \frac{\theta_i^*}{\sum_{i=1}^{35} \theta_i^*}, \quad \theta_i^* = \alpha exp\Big\{-\Big(\frac{j-\beta}{\delta}\Big)^2\Big\} \quad for\ j = 1, \ldots, 35\ , \quad and \quad \alpha = 1, \qquad \beta = 6, \qquad \delta = 12.$$

The parameters are chosen so that the true impulse response is hump shaped, which is thought to be a common occurrence in macro Auclert et al. (2020), and the cumulative impulse response sums up to 1. The results are presented in Figure 2. Here, the LP estimators have at least 88% coverage for all horizons and approximately 95% coverage for most horizons. The bootstrap VAR estimator, on the other hand, have about 90% coverage for the first 1 or 2 horizons before coverage drops precipitously. The analytic VAR estimator had coverage comparable if not slightly better than the LP estimators. There are little to no efficiency gains from using the LP GLS estimators as all of the estimators except for the VAR bootstrap have approximately the same average length throughout.

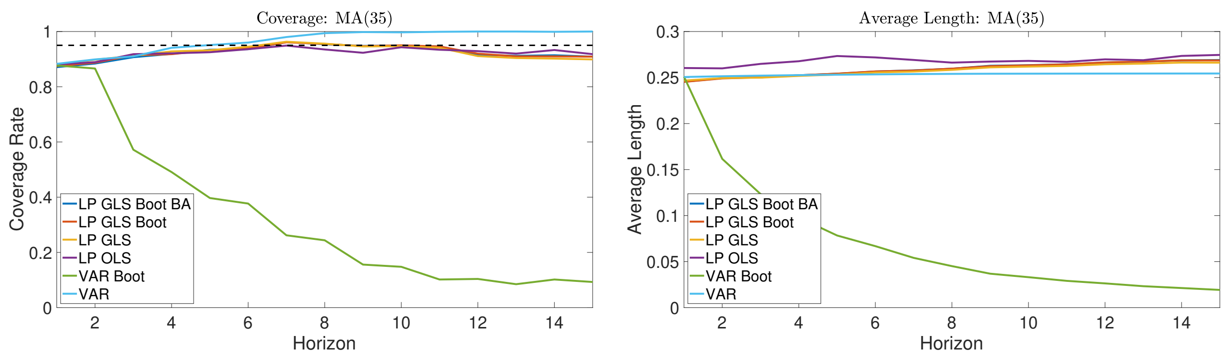

Figure 2: Coverage Rates for 95% Confidence Intervals and Average Length for MA(35)

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).

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. The LP GLS estimators are more efficient than the LP OLS estimator. Select results are presented in Figure 3.

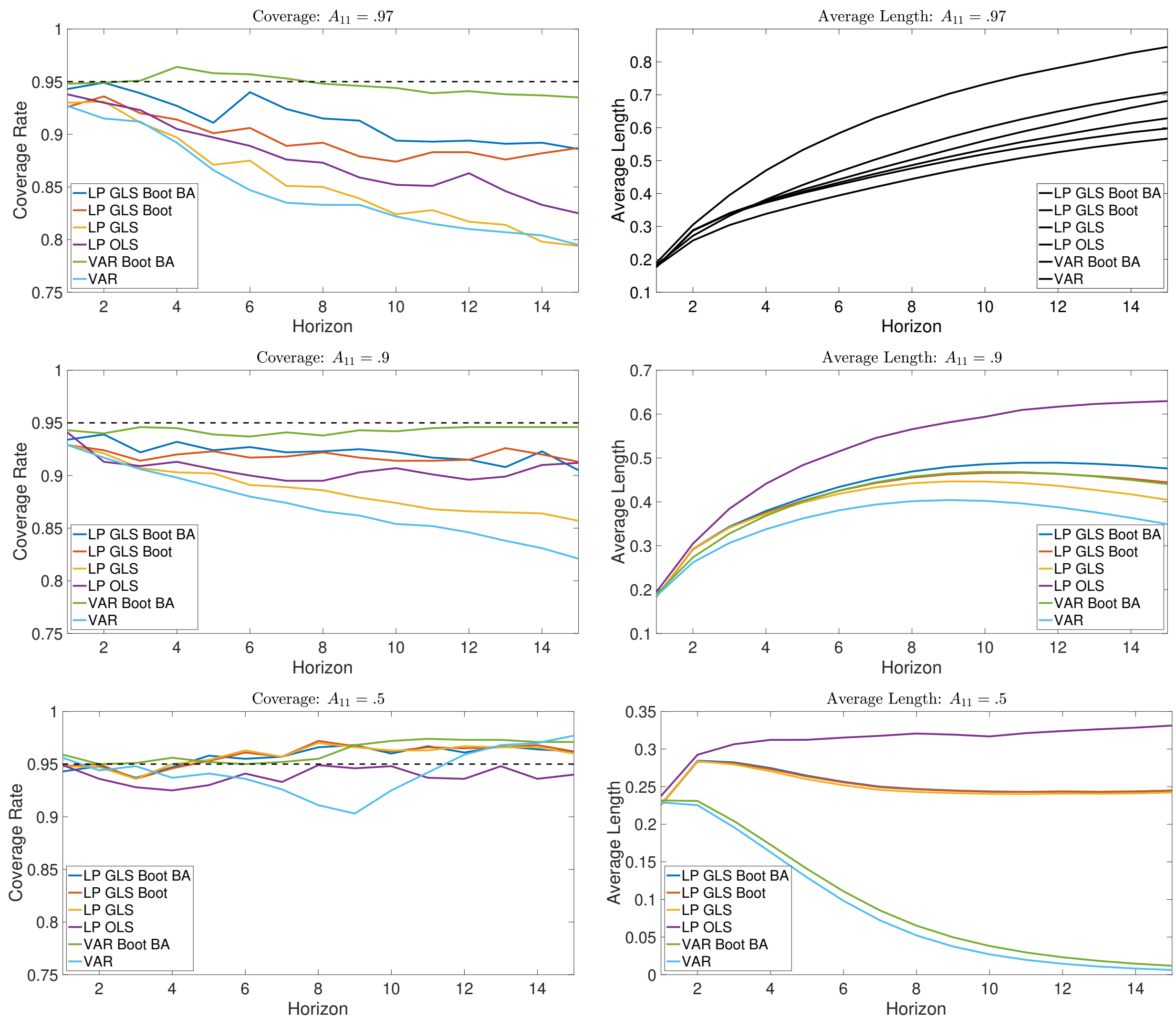

Figure 3: Coverage Rates for 95% Confidence Intervals and Average Length for VAR(1) Models

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).

The following is from Kilian and Kim (2011) 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 $\varepsilon_t \sim N(0, PP')$. All of the estimators have at least 90% for all parameters. The average length for the LP GLS estimators are quite a bit shorter than the LP OLS estimator. Also note that the average length of the VAR analytic confidence intervals are much wider than the bootstrap VAR confidence intervals and even the LP GLS intervals. Select results are presented in Figure 4.

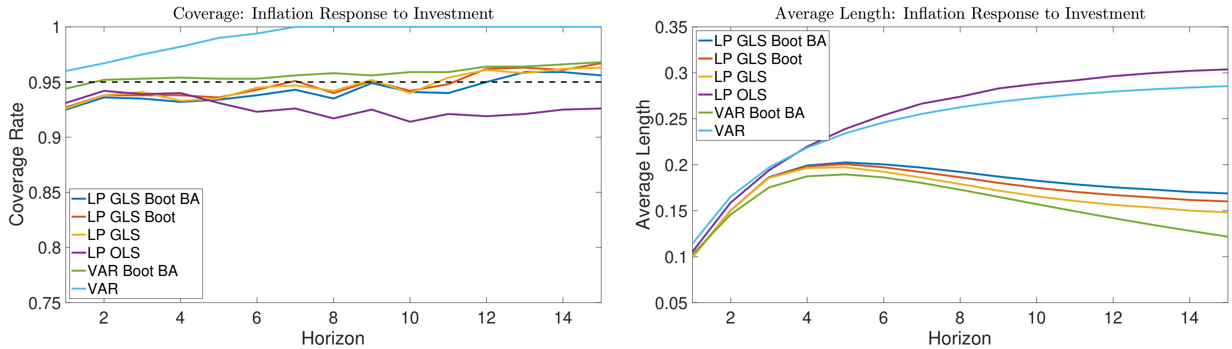

Figure 4: Coverage Rates for 95% Confidence Intervals and Average Length for VARMA(1,1)

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).

In summary for the Kilian and Kim (2011) data generating processes, the LP GLS bootstraps performed well, but not as well as the bias-adjusted bootstrap VAR. The LP GLS analytical estimator tended to perform better than the LP OLS, but all three of these estimators had coverage that tended to fall off more when estimators had more persistent eigenvalues. Surprisingly for the VARMA(1,1) DGP, most estimators had average lengths shorter than the infinite order analytic VAR estimator. 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 are discussed in the next section (Section A.5). Moreover, 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 Kilian and Kim (2011) Monte Carlos.

A.5 Bootstrapping 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 Kline and Santos (2012), 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. This can be seen by using the structural version of the rearranged the scaled score since. Note that

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

where $RStrucScore_{t+1}^{(H)}$ 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. Let η be a zero mean unit variance random variable with finite fourth moments. For simplicity assume $T-k-H=N\ell$ where N is the number of blocks of length ℓ is the length of each block. Define $RStruc\hat{S}core_t^{(H)}$ as the sample analogue of $RStrucScore_{t+1}^{(H)}$; anywhere a parameter is not known, the estimated sample analogue would be used, i.e. $\hat{\Gamma}_k^{-1}$ for Γ_k^{-1}. Instead of multiplying the rearranged scaled scores by the i.i.d. $\{\eta_{k+1},\dots,\eta_{T-H+1}\}$, yielding

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

one would create

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

In other words, cut $\{RStruc\hat{S}core_{k+1}^{(H)},\dots,RStruc\hat{S}core_{T-H+1}^{(H)}\}$ into N blocks of length ℓ and multiply the jth block by η_j to get the bootstrap sample $\{RStru\hat{c}Score_{k+1}^{(H),*},\dots,RStru\hat{c}Score_{T-H+1}^{(H),*}\}$. 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 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 3 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

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

Note that $\hat{\Theta}_l^* = \hat{B}_1^{*(l),GLS}$. 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.[6] 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.[7]

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.

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

[7]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).

Assumption 5. *Let Assumption 4 hold. Assume that in addition that* $\eta_{[t/\ell]}$ *is i.i.d. with* $E|\eta_{[t/\ell]}|^4 \leq \Delta < \infty$,

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

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

Theorem 5. *(Validity of Bootstrap for Structural Inference) Under Assumption 5*

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

Proof. See section A.2 of the 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 5 includes the sieve VAR as a special case, thus the bootstrap also provides a sieve extension of Brüggemann et al. (2016).[8]

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. In the case where one is only interested in the reduced form impulse responses, one can simply set $\ell = 1$ and not sample $\hat{\Sigma}^*$. Using standard arguments the Delta method can applied.[9]

A.6 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, confidence intervals 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). Going back to the infinite order horizon 1 LP

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

[8]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. Note that the score bootstraps avoids the second order bias created by recursive design bootstraps (see Kilian (1998) for a demonstration of the second order bias).

[9]It is possible to write a more general delta method theorem to include a wider range of statistics (albeit it may require additional conditions). See for example Corollary 1 in Inoue and Kilian (2002).

and let $\varepsilon_t = R(L)s_t$ where s_t is a vector of structural shocks and $R(L)$ is a lag polynomial.

It is often the case that the researcher may not know all of the identifying restrictions in $R(L)$, 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). Plagborg-Møller and Wolf (2022) 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

$$(i)\ E[s_{1,t}\widetilde{z}_t] \neq 0,$$

$$(ii)\ E[s_{2,t}\widetilde{z}_t] = 0,$$

$$(iii)\ E[s_{t+j}\widetilde{z}_t] = 0\ for\ j \neq 0.$$

where $\widetilde{z}_t$ is the population residual from projecting z_t on all lags of z_t, y_t. 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, $\widetilde{z}_t$, is only identifying the impulse response of the shock $s_{1,t}$. If the third condition is not satisfied, then $\widetilde{z}_t$ will amalgamate the impulse responses at different horizons. In the case where $\widetilde{z}_t$ is a vector of multiple instruments, the conditions are easily extended (Plagborg-Møller and Wolf, 2022).

Researchers typically estimate LP-IV via two-stage least squares (2SLS). For example, say I want to estimate the structural 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

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

via 2SLS and using $\widetilde{z}_t$ as an instrument for mp_t.[10] 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, 2021).[11] 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

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

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.[12] Going back to the monetary policy example, the impulse response $g^{(h)}$ can be constructed as the impulse response of $output_{t+h}$ to $\mathring{\varepsilon}_{1,t}$ divided by 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

[10] The increasing variance problem may be particularly problematic with LP-IV, because the increasing variance can weaken the strength of instrument for $h \geq 1$ if one is estimating a cumulative multiplier directly.

[11] 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.

[12] 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.

identification scheme, one can just estimate the reduced form impulse responses of $\mathring{y}_t$ via their preferred LP GLS method and construct the structural impulse responses of interest.

A.7 Application to Gertler and Karadi (2015)

For an empirical application, I redo the analysis of Gertler and Karadi (2015). One of the reasons why the Gertler and Karadi (2015) analysis is so interesting is because there has been tension in the literature about the results. Using a Proxy 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, 2022).[13] Treating the high frequency identification instrument as the structural monetary policy shock and using Newey-West standard errors, Ramey (2016) finds that an increase in the the high frequency identification instrument leads to a significant decrease in CPI and a significant increase in industrial production.[14] 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.

To standardize comparisons so all of the methods are valid under the same assumptions, I will compare the results of the invertibility robust LP GLS IV, the invertibility robust LP-IV with the equal-weighted cosine HAC standard errors, and the invertibility robust recursive SVAR. All of these methods are valid in the presence of conditional heteroskedasticity, are invertibility robust (that is, they are consistent in the presence of omitted variables), and are estimated over the same sample period. The system of macroeconomic variables I use includes output (growth rate of industrial production), inflation (growth rate of CPI), the one-year Treasury yield, the excess bond premium spread, and a high frequency identification instrument.[15] I do the baseline analysis using the surprise to the three-month ahead fed funds futures (ff4_tc) as the instrument, and the data spans 1990M1-2012M6. All of the data was obtained from Ramey's (2016) data repository.

First I estimate the system using the LP GLS bootstrap (without bias adjustment) based on 20,000 replications using the procedure detailed in section A.5. The high frequency instrument is ordered first in the system. I use 12 lags in estimation as in Gertler and Karadi (2015) and a block length of 10, but the results are qualitatively similar across alternative choices for these parameters. For each bootstrap draw, I apply the recursive decomposition as discussed in section A.6, and the 90% confidence intervals for each horizon are the are the 5% and 95% quantiles of the structural impulse response of interest across all of the bootstrap draws for that horizon. Section A.8 discusses the software suite used for LP GLS IV in context of this specific application.

[13] Jentsch and Lunsford (2022, 2019) 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 3.3. Gertler and Karadi's (2015) Proxy SVAR assumes invertibility, which is akin to assuming that the SVAR system doesn't have any omitted variables (Stock and Watson, 2018).

[14] As noted in Stock and Watson (2018), constructed measures of shocks have measurement error, which in general leads to bias if the measure is treated as the true shock.

[15] These are the same variables used in Gertler and Karadi (2015), Ramey (2016), except they use logs of industrial production and CPI. The assumptions of these techniques are based on stationary data and Monte Carlos in this paper only compare stationary data generating processes. Moreover, the equal-weighted cosine HAC standard errors weren't designed to handle persistent regressors (Lazarus et al., 2018).

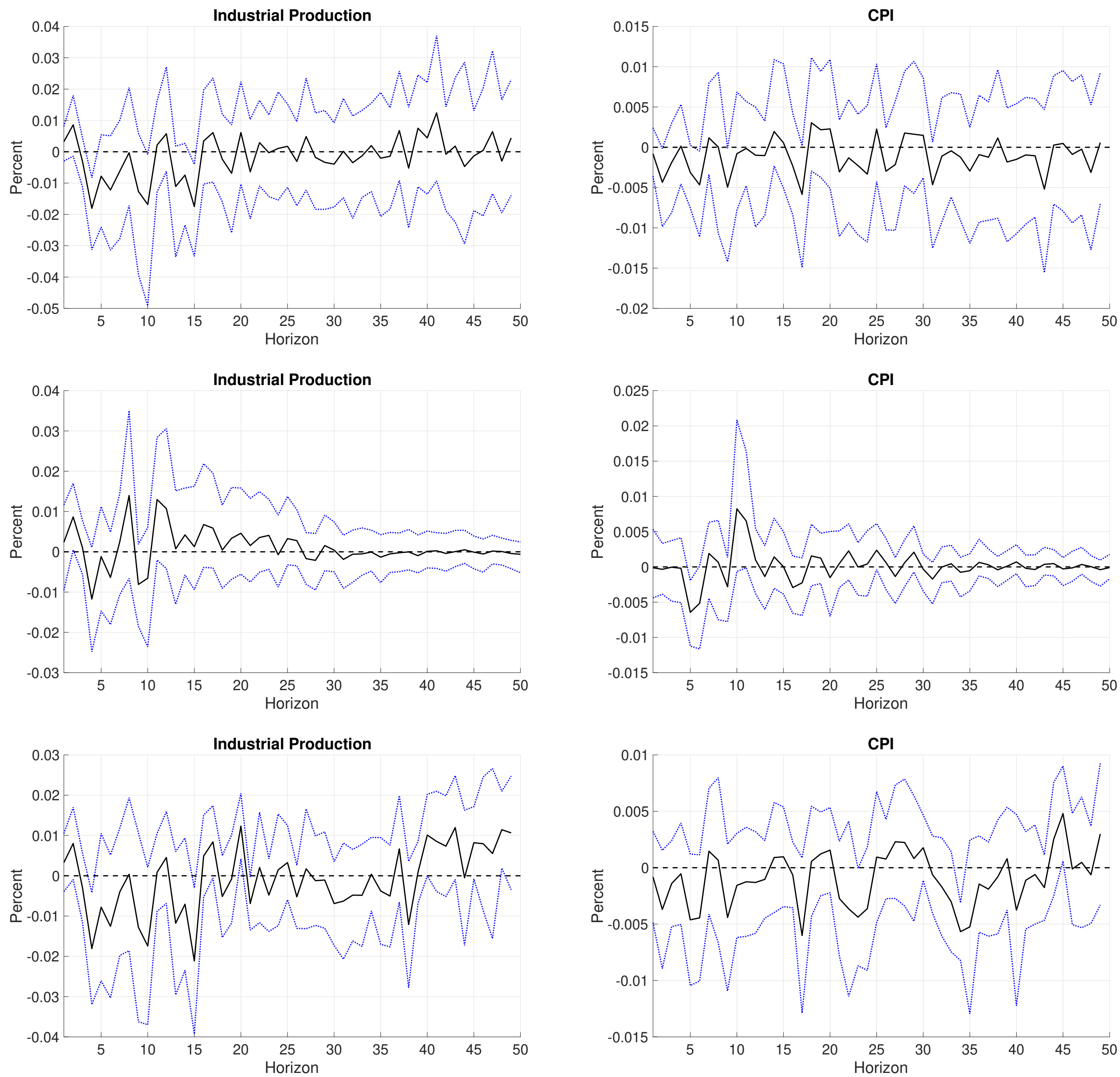

Figure 5: LP GLS bootstrap results with 90% confidence intervals (top panel). Invertibility robust invertibility-robust recursive SVAR results (middle panel) and the LP-IV results (bottom panel) with 90% confidence intervals.

The LP GLS IV bootstrap results for output and inflation are presented in the top panel of Figures 5. 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. F-test indicates the instrument is relevant, with a bootstrap F-statistic of approximately 25.[16]

Next I estimate the invertibility-robust recursive SVAR. To implement, I repeat the procedure discussed

[16] Instrument strength varied with block length and the number of lags, but the instrument strength was generally approximately 20 or higher.

in section A.5, but only estimate the horizon 1 LP (since the horizon 1 LP is the VAR). The VAR is not bias adjusted. I also use 12 lags, 20,000 bootstrap replications, and a block length of 10. The bootstrap draws are used to construct the VAR approximation of the Wold coefficients, then for each bootstrap draw, I apply the recursive decomposition as discussed in section A.6 to trace out the structural impulse responses. The LP-IV with the equal-weighted cosine HAC standard errors results are estimated using the 2SLS procedure discussed in section A.6, where the control variables are the 12 lags of output, inflation, the one-year Treasury yield, the excess bond premium, and the high frequency identification instrument. The results for the invertibility-robust recursive SVAR and the LP-IV are presented in Figure 5 (in the middle and bottom panels respectively). As can be seen, one would in general still fail to reject the null hypothesis that monetary policy has no affect on output and inflation.[17]

The results are consistent with what Nakamura and Steinsson (2018) 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 probably 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 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.

A.8 How to Section for Code

This section details how to use the LP GLS IV code and the reduced form LP GLS bootstrap code and tries to be as self contained as possible. To illustrate how to use the LP GLS IV code, I'll walk through how the Gertler and Karadi (2015) application figures were constructed using LP GLS IV. The code is

[*Structural_IRF, fstat*] = lp_gls_boot_iv(*y, arp, nstraps, maxh, blocksize, badj*)

where the inputs are:

- *y* is the $r \times T$ matrix of data,
- *arp* is the desired lag length,
- *nstraps* is the number of bootstrap draws,

[17] As a robustness check I also used the change in one-month ahead fed futures as an instrument. The results were qualitatively and quantitative similar and the conclusions to follow remain unchanged.

- *maxh* is the max horizon one wants to estimate,
- *blocksize* is the block size for the block bootstrapping,
- *badj* is an indicator variable which is 1 if one wants to bias adjust and 0 if one does not want to bias adjust.[18]

The outputs are:

- *Structural_IRF* is the $r \times maxh \times nstraps$ matrix structural impulse responses
- *fstat* is the bootstrap first stage F-statistic.

For this code, the order of the variables in y is important. The first row is the instrument (e.g. HFI instrument). The second row is the independent variable that needs to be instrumented (one year treasury). The order of the rest of the variables doesn't matter. What's outputted is the structural impulse responses of the other variables to the instrumented independent variables (e.g. output's responses to an exogenous change in the 1-year treasury, inflation's responses to a exogenous change in the 1-year treasury, etc). The code is

[*Structural_IRF, fstat*] = lp_gls_boot_iv(*y, 12, 20000, 48, 10, 0*).

Structural_IRF(i,h,:) gives the bootstrap draws of structural impulse responses for the impact of 1 unit increase in the 1-year Treasury has on variable i at horizon h , and 90% confidence intervals can be calculated by quantile(*Structural_IRF*(i,h,:),[.05 .95]). For example, since output is the 3rd variable, the bootstrap draws of structural impulse responses for the impact of 1 unit increase in the 1-year Treasury has on output for horizon 15 is given by *Structural_IRF*(3,15,:), and 90% confidence intervals can be calculated by quantile(*Structural_IRF*(3,15,:),[.05 .95]). Since inflation is the 4th variable, the bootstrap draws of structural impulse responses for the impact of 1 unit increase in the 1-year Treasury has on inflation for horizon 27 is given by *Structural_IRF*(4,27,:) and 90% confidence intervals can be calculated by quantile(*Structural_IRF*(4,27,:),[.05 .95]).

For those interested in the reduced form LP GLS code, the scripts *Testing_Code_VAR(1)* and *Testing_Code_VARMA(1,1)* in the data replication files walks through examples for how to estimate the Wold impulse responses using the VAR(1) and VARMA(1,1) from section A.4 of the appendix . The function to estimate the Wold irfs is

[*Beff, Sigma_hold*] = lp_gls_boot(*y, arp, nstraps, maxh, blocksize, badj, full*)

where the inputs are mostly the same as the LP GLS IV code

- *y* is the $r \times T$ matrix of data,
- *arp* is the desired lag length,
- *nstraps* is the number of bootstrap draws,
- *maxh* is the max horizon one wants to estimate,
- *blocksize* is the block size for the block bootstrapping,

[18] The econometrics toolbox is required for bias adjustment. It's debatable whether you should bias adjustment see Efron and Tibshirani (1993). So depending on what your doing you may or may not want to bias adjust. Uses bias adjustment from West and Zhao (2019). See bias code folder and this link for more details https://www.ssc.wisc.edu/~kwest/publications/2010/Adjusting%20for%20Bias%20in%20Long%20Horizon%20Regressions%20Using%20R.pdf

- *badj* is an indicator variable which is 1 if one wants to bias adjust and 0 if one does not want to bias adjust.

- *full* is an optional input where *full*=0 gives the Wold coefficients while *full*=1 gives all of the LP parameters for each horizon. If *full* isn't included as an input, the default will just give the Wold coefficients.

the outputs are

- *Beff* is the $r \times r \times maxh \times nstraps$ matrix of the wold impulse responses bootstrap replications (default). For example *Beff*(:,:,h,n)' gives the nth bootstrap draw of the horizon h Wold impulse. If *full*=1 is included as an input, *Beff* gives the $(1 + kr) \times r \times maxh \times nstraps$ matrix of all of the LP estimates. For example if *full*=1, *Beff*(:,:,h,n) gives the nth bootstrap draw of $B(k,h)' = (B_1^{(h)}, \ldots, B_k^{(h)})'$ as defined in the paper, but with the addition of an intercept (which would be the first column).

- *Sigma_hold* is the $r \times r \times nstraps$ matrix of the covariance matrix bootstrap horizon 1 LP (VAR) residuals replications.

References

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

Auclert, A., M. Rognlie, and L. Straub (2020). Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model. *NBER Working Paper 26647*.

Barnichon, R. and C. Matthes (2018). Functional approximation of impulse responses. *Journal of Monetary Economics 99*, 41–55.

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.

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

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

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 (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.

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

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

Inoue, A. and L. Kilian (2002). Bootstrapping smooth functions of slope parameters and innovation variances in var â models. *International Economic Review 43*(2), 309–331.

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

Jentsch, C. and K. G. Lunsford (2022). Asymptotically valid bootstrap inference for proxy svars. *Journal of Business and Economic Statistics 40*(4), 1876–1891.

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.

Jordà, Ò., S. R. Singh, and A. M. Taylor (2020). The long-run effects of monetary policy. *NBER Working paper 26666*.

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.

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.

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.

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

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

Plagborg-Møller, M. and C. K. Wolf (2022). Instrumental variable identification of dynamic variance decompositions. *Journal of Political Economy 130*(8), 2164–2202.

Poskitt, D. S. and W. Yao (2017). Vector autoregressions and macroeconomic modeling: An error taxonomy. *Journal of Business and Economic Statistics 35*(3), 407–419.

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.

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.