Local Projection Based Inference under General Conditions*

Ke-Li Xu†

Indiana University Bloomington

January 4, 2023

Abstract

This paper provides the uniform asymptotic theory for local projection (LP) regression when the true lag order of the model is unknown, possibly infinity. The theory allows for various persistence levels of the data, growing response horizons, and general conditionally heteroskedastic shocks. Based on the theory, we make two contributions. First, we show that LPs are semiparametrically efficient under classical assumptions on data and horizons if the controlled lag order diverges. Thus the commonly perceived efficiency loss of running LPs is asymptotically negligible with many controls. Second, we propose LP-based inferences for (individual and cumulated) impulse responses with robustness properties not shared by other existing methods. Inference methods using two different standard errors are considered, and neither involves HAR-type correction. The uniform validity for the first method depends on a zero fourth moment condition on shocks, while the validity for the second holds more generally for martingale-difference heteroskedastic shocks.

*The author is grateful to participants in Econometric Society meetings in different regions in 2021 and 2022, Symposium on Econometric Theory and Applications 2022, Cemmap/SNU Workshop on Advances in Econometrics 2022, NBER-NSF Time Series Conference 2022, ECMFE Workshop on Predictability, Forecasting and Monitoring 2021, and departmental seminars at LSE, University of Nottingham, McGill University, Indiana University, University of California-Irvine, and University of Illinois at Urbana-Champaign, especially Mikkel Plagborg-Møller, for insightful comments and suggestion.

†*Address*: Department of Economics, Indiana University, Wylie Hall 255, 100 South Woodlawn Avenue, Bloomington, IN 47405-7104, USA. *E-mail*: `kelixu@indiana.edu`. *Weblink*: `https://sites.google.com/view/kelixu`.

Keywords: Impulse response; local projection; persistence; semiparametric efficiency; uniform inference.

JEL Classification: C32.

1 Introduction

Impulse response analysis is a major tool in applied macroeconomic analysis. In this paper, we aim to develop statistical inference methods on impulse responses which are robust to model specification, model parameters, the range of propagation horizons under investigation, and the dependence structure of the shock process. Thus the procedures developed can accommodate stylized features in macro data such as stochastic trends, persistence, long-range dependence, and volatility clustering.

Our approach is based on local projections (LPs) (Jordà, 2005, Dufour and Renault, 1998). Consider K endogenous variables of interest, stacked in the vector y_t, and the primary interest is in the responses of its first entry y_{1t} to various economic shocks after h propagation horizons, where $h \geq 1$. These responses are collected in the $K \times 1$ vector $\beta_1(h)$. The method of LP runs the regression

$$y_{1,t+h} = \beta_1(h)' y_t + \sum_{\ell=1}^{p-1} \theta_{1\ell}(h)' y_{t-\ell} + \eta_{1t}(h), \tag{1}$$

for a given integer $p \geq 1$, where $\eta_{1t}(h)$ is the regression error. The number of controlled lags $p-1$ needs to be determined. A common method is to apply some data-dependent information criterion to select p for the horizon-one regression ($h = 1$), and then use the same p for regressions at all other horizons.

Our robust inference is based on a new standard error of the ordinary least squares (OLS) estimator of the LP regression (1). The new standard error is constructed via first partialling out controls and then estimating the variance of the martingale-transformed (effective) regression score. The estimated score variance is simply a sum of squares, thus does not require the selection of a tuning parameter as usually needed in the long-run variance estimation, even that original score contributions (with or without partialling out) are in general serially correlated (for $h \geq 2$). In the paper we establish uniform validity of the inference procedure robust to data features mentioned above under the vector autoregression (VAR) model with an unknown and potentially infinite number of lags.

Robust inference of impulse responses which are potentially generated from a VAR(∞) process is new. Existing literature which allows models with infinitely many lags (e.g. Inoue and Kilian, 2002, Chang and Sakata, 2007, Jordà and Kozicki, 2011, Kilian and Lütkepohl, 2017, chapter 12, and Lusompa, 2022) mostly focuses on stationary models and fixed horizons. Leaving the true lag order unrestricted in the VAR model has been argued as fundamentally important in the modern empirical macro literature (Kilian and Lütkepohl, 2017, chapter 6, Nakamura and Steinsson, 2018).

When the VAR model is (plausibly) viewed as an approximation of the unknown true data generating process (DGP), a large model lag order is often used. Even when the true DGP has a finite number of lags, a conservative (large) model lag order is recommended for robust inference; see Montiel Olea and Plagborg-Møller (2021). When the true DGP is VAR(∞), the model order is typically required in theory to diverge (at an appropriate rate) for reasonable model approximation quality. A sufficiently large model order has an important implication. Our analysis shows that when the model order diverges, the LP estimator of the impulse response is semiparametrically efficient, under classical assumptions on data and design. The result extends those of Plagborg-Møller and Wolf (2021), who show that iterative model-implied and LP estimators have the same population estimand and probability limit under VAR(∞) model, to the equivalence of their asymptotic distributions. The equivalence result is important to motivate our consideration of LP for the *inference* purpose as the efficiency loss is (asymptotically) negligible, in contrast to the common wisdom that LP is inefficient under the finite small-order VAR model.

We now discuss recent work in the literature which is closely related. Montiel Olea and Plagborg-Møller (2021, MOPM hereinafter), which clearly motivates the current study, is concerned about robust inference for local projection regression under the finite-order VAR model. Other than assuming the model order to be finite and known, a fundamental assumption underlying their uniform validity result is the mean independence of the shock process, which we relax in the current paper when developing the new theory and methods. As the authors mentioned, "... What matters is that we include enough control variables so that the effective regressor of interest approximately satisfies the conditional mean independence condition", (MOPM, p.1809). The current paper thus contributes to the line of work in the following five aspects, allowance for a more general data generating process (with potentially infinite lags), a more flexible controlled lag order (which may diverge) in the LP regression, a new asymptotic theory for general martingale difference shocks, a new standard error (the consistency of which does not require mean-independent shocks), and a unified inferential framework which covers both individual and cumulated impulse responses (thereby permitting higher-order integrated data via differencing).

Our paper is complementary to the literature on robust model-implied inference for impulse responses (Mikusheva, 2012, Inoue and Kilian, 2020, and references therein). These methods, mostly developed under the univariate autoregression (AR) model with finite and known order, specify a dynamic model (and sometime estimate model parameters) and draw inference for impulse responses (as a function of model parameters) based on the likelihood principle or the delta method. Compared with model-implied inference methods in the literature, the advantage of our methods lies in the following aspects, the validity uniform over a larger class of models and larger spaces of model parameters and horizons, accommodation of more general shock processes, and extendability (in both theory and computational feasibility) to multivariate models. For example, Mikusheva (2012) focuses on the univariate model, and considers the extension to the VAR model but only allows for a unit root for one endogenous variable. Her uniformity results allow for heteroskedastic shocks but only when data are highly persistent (i.e. under the local-to-unity model). Under the univariate finite-order AR model, the methods of Inoue and Kilian (2020) require some rank condition on the parameter space (see also Dufour, et al., 2021), which we dispense with. They mainly focus on a fixed horizon and allow the horizon to grow but only under the local-to-unity model.

In a recent working paper, Lusompa (2022) proposed an alternative LP estimator utilizing the structure of the LP regression error. The paper discusses the efficiency comparison with the standard LP estimator, and shows that in a simplified setting of AR(1) model, the alternative estimator is asymptotically more efficient across horizons in general. In the current paper we show that in a similar simple setting, the efficiency gain does not carry over if a sufficiently large lag order is used; the alternative estimator has the same asymptotic distribution with the LP estimator. Importantly, the inference discussed in Lusompa (2022) requires a much stronger set of assumptions: stationarity, fixed horizons, model lag order with a more restricted rate. It thus remains unclear regarding the uniform validity of inference based on the alternative estimator under general conditions and parameter spaces considered in the current paper.

The rest of the paper is organized as follows. In Section 2, we introduce the inferential framework, and assumptions on the data generating process, the parameter space, and the approximation model. These assumptions are illustrated via a potentially highly persistent VARMA (vector autoregression and moving average) model. Section 3 gives the uniform asymptotic theory for the LP estimator and discusses a few implications especially on semi-parametric efficiency. In Section 4 we prove the consistency of two different standard errors, neither of which involves selection of a tuning parameter as in the usual long run variance estimation. The new standard error proposed in the paper is shown to enjoy robust properties to the form of dynamic dependence of shocks. Another potential advantage of the new standard error regarding the lag order flexibility under finite-order VAR model is also discussed. Simulation results are presented in Section 5 and concluding remarks in Section 6. In Appendices A and B we sketch the proofs of some main results. Technical details and additional results are provided in the on-line Supplement to the paper (available at the author's website).

A word on notations. For a matrix (vector, scalar) x, we use $|x|$ to denote its Frobenius norm, i.e. $|x| = [\text{trace}(x'x)]^{1/2}$. We use C with a decoration ($C_1, C_u, C', \overline{C}_w$, etc.) to denote a positive constant, which does not depend on the model coefficients $\{a_1, a_2, \cdots, \}$, the LP lag order p, the horizon h, or the sample size n. For a square matrix D, denote $\lambda_{\min}(D)$ as its smallest eigenvalue, and $\lambda_{\max}(D)$ as its spectral norm (i.e. the largest eigenvalue in magnitude). Denote $\text{diag}(x_1, \cdots, x_K)$ as the block diagonal matrix with $x_1, \cdots, x_K$ on the diagonal. We use $\otimes$ to denote the Kronecker product of two matrices, and $\text{vec}(D)$ to denote the vectorization vector of the matrix D. Write the $K \times K$ identity matrix as I_K.

2 Data generating process and parameter space

Consider the K-dimensional vector autoregression process of infinite order (VAR(∞)):

$$\underset{K\times 1}{y_t} = \sum_{j=1}^{\infty} a_j y_{t-j} + u_t, \tag{2}$$

where u_t is a serially uncorrelated shock process. Initial conditions are set as $y_t = 0$, for $t \leq 0$. In the theoretical development, we assume that y_t contains no deterministic part, but in implementation we always include an intercept in the regression, e.g. (1) or (4); see Section 5 of the paper, and Section S6 of the Supplement.

Write the data generating process (DGP) (2) as $a(\mathrm{L})y_t = u_t$, where $a(\mathrm{L}) = \mathrm{I}_K - \sum_{j=1}^{\infty} a_j \mathrm{L}^j$ and L is the lag operator such that $\mathrm{L}y_t = y_{t-1}$. Under such DGP, the horizon-h impulse response matrix $\beta(h) = [\beta_1(h), \cdots, \beta_K(h)]'$ is the coefficient of u_{t-h} in the *formal* vector moving average (VMA(∞)) representation of the model (2) (obtained by formal inversion of $a(\mathrm{L})y_t = u_t$), $y_t = \sum_{h=0}^{\infty} \beta(h) u_{t-h}$, with $\beta(0) = \mathrm{I}_K$. Focusing on the responses of y_{1t} (to K shocks in u_t), the parameter of interest in this paper is defined as

$$\beta_1(h, \mu) = \sum_{j=1}^{h} \mu_j \beta_1(j),$$

for $h \geq 1$, where $\mu = (\mu_1, \cdots, \mu_h)'$ is an h-dimensional nonzero vector of known constants. The inferential framework developed below applies to linear combinations of all impulse responses (IRs) up to the horizon h, including individual and cumulative ones, corresponding to μ set as $\mu_{\mathrm{IR}} = (0, \cdots, 0, 1)'$ and $\mu_{\mathrm{CIR}} = (1, \cdots, 1)'$, respectively. To reconcile with earlier notations, we write $\beta_1(h, \mu_{\mathrm{IR}})$ simply as $\beta_1(h)$. In what follows we will use this notational convention by implicitly imposing $\mu = \mu_{\mathrm{IR}}$ whenever we drop the argument μ, e.g. in $\beta_1(h, \mu)$, $\pi_1(h, \mu)$ and $y_{1t}(h, \mu)$, etc.

Cumulative level responses can be interesting *per se* (long-run responses of the economy). Cumulative responses recover orginal level responses if the response variable of interest is transformed into differences (e.g. growth rates, returns); see e.g. Lunsford (2020) for such applications.

We first define the parameter space $\mathcal{A}$ of coefficients in the model (2), where $\mathcal{A} = \{a \in \mathbb{R}^{\infty} : a = \mathrm{vec}(a_1, a_2, \cdots)\}$ is a subset in the $\mathbb{R}^{\infty}$ space. We introduce the following scalar quantity, which eventually determines the convergence rate,

$$\pi_k(h, \mu) = \sum_{i=1}^{h} |\varphi_{ki}|^2,$$

for $k = 1, \cdots, K$, where $\varphi_{ki} = \sum_{j=i}^{h} \mu_j \beta_k(j - i)$. The form of φ_{ki} will become clear in the population local projection regression (3) introduced below. The restrictions on coefficient matrices $\{a_j : j \geq 1\}$ in the model (2) are imposed through conditions on impulse response matrices, as stated in the following Assumption 1.

Assumption 1. The following is true for $k = 1, \cdots, K$ and given μ.
(i) $\sup_{a\in\mathcal{A}} \sup_{i\geq 0} |\beta_k(i)| \leq C_1$.

(ii) $\sup_{a\in\mathcal{A}} \sup_{h\geq 1} \frac{|\beta_k(h,\mu)|^2 + \sum_{i=1}^{h} |\varphi_{ki}|}{\pi_k(h,\mu)} \leq C_2$.

(iii) $\sup_{a\in\mathcal{A}} \sup_{N\geq 2} \sum_{i=1}^{N-1} |\beta_k(i) - \beta_k(i-1)| \leq C_3$.

Assumption 1 is a high-level one. For illustration, if the data generating process follows scalar AR(1), the assumption requires $|a_1| \leq 1$. For the AR(2) process, the assumption is satisfied via reparameterization $a(\mathrm{L}) = 1 - a_1\mathrm{L} - a_2\mathrm{L}^2 = (1 - b\mathrm{L})(1 - \rho\mathrm{L})$, where $|\rho| \leq 1$ and $|b| \leq 1 - \varepsilon$, for a constant $\varepsilon \in (0, 1]$. Assumption 1 thus rules out explosive roots (violating (i) and (ii)) and integration of order larger than one (violating (iii)) for the data y_t. If some response variables are potentially integrated of order two, taking differences (to form y_t) is necessary to apply our framework. If this is the case, setting $\mu = \mu_{\mathrm{CIR}}$ will recover level responses of original variables. From a technical point of view, the conditions in Assumption 1 are especially useful in establishing uniform bounds for moments in the asymptotic analysis.

Write $y_{1t}(h,\mu) = \sum_{j=1}^{h} \mu_j y_{1,t+j}$. Under the model (2), by recursive substitution we obtain the following LP(∞) form (local projection with infinite lags), with the focus on responses of the first entry of y_t,

$$y_{1t}(h,\mu) = \beta_1(h,\mu)' y_t + \sum_{\ell=1}^{\infty} \theta_{1\ell}(h,\mu)' y_{t-\ell} + \xi_{1t}(h,\mu), \tag{3}$$

where $\xi_{1t}(h,\mu)$ is the LP(∞) regression error,

$$\xi_{1t}(h,\mu) = \sum_{i=1}^{h} \varphi_{1i}' u_{t+i}.$$

To estimate $\beta_1(h,\mu)$, we run the truncated regression

$$y_{1t}(h,\mu) = \beta_1(h,\mu)'y_t + \sum_{\ell=1}^{p-1} \theta_{1\ell}(h,\mu)'y_{t-\ell} + \eta_{1t}(h,\mu), \tag{4}$$

for a chosen model order p. The LP(p) regression error $\eta_{1t}(h,\mu)$ in (4) thus takes the form $\eta_{1t}(h,\mu) = \sum_{\ell=p}^{\infty} \theta_{1\ell}(h,\mu)'y_{t-\ell} + \xi_{1t}(h,\mu)$.

Our inferential framework is valid for a range of model orders p, such that $\underline{p} \leq p \leq \overline{p}$, where $\underline{p}$ and $\overline{p}$ are two positive integers such that the following Assumption 2 is satisfied. On top of controlling the bounds on p, Assumption 2 further restricts the coefficient parameter space. While Assumption 1 imposes restrictions on the behavior of all a_j, Assumption 2 is only concerned about tail coefficients a_j, for $j \geq \underline{p}$. For the horizon h in (4), we consider the range $1 \leq h \leq \overline{h}$, where the upper bound $\overline{h}$ may grow with the sample size. Denote $\overline{\mu} = \sup_{1\leq h\leq \overline{h}} |\mu|_1$, where we write the L^1-norm as $|\mu|_1 = \sum_{j=1}^{h} |\mu_j|$.

Assumption 2. (i). $\overline{h}\overline{\mu}^2\overline{p}^2/n \to 0$.

(ii). $\lim_{n\to\infty} \sup_{a\in\mathcal{A}} \overline{p}n^{1/2} \sum_{j=1}^{\infty} j|a_{\underline{p}-1+j}| = 0$.

Assumption 2(i) restricts the upper bounds $\overline{p}$ and $\overline{h}$. There is a tension between the ranges of model orders and response horizons; the upper bound on p is tighter if a wider range of horizons is allowed. Given such an upper bound $\overline{p}$, Assumption 2(ii) restricts the lower bound $\underline{p}$; $\underline{p}$ should be sufficiently large such that the remote sum $\sup_{a\in\mathcal{A}} \sum_{j=1}^{\infty} j|a_{\underline{p}-1+j}|$ shrinks to zero faster than $\overline{p}^{-1}n^{-1/2}$. If the data generating process involves infinite lags, these conditions require $\underline{p} \to \infty$ so that the misspecification bias (via truncation) diminishes.

Assumption 2 (ii) is stated for interpretation transparency rather than being the weakest possible; see the weaker Assumption 2B in the Supplement which is sufficient for asymptotic results developed below. To illustrate Assumptions 1 and 2, consider the following example.

Example. Suppose that y_t follows the vector autoregression and moving average model with finite orders $q+1$ and r (VARMA($q+1, r$)):

$$Q(\mathrm{L})y_t = d(\mathrm{L})u_t, \tag{5}$$

where $Q(\mathrm{L}) = b(\mathrm{L})(\mathrm{I}_K - \rho \mathrm{L})$ and $d(\mathrm{L}) = \mathrm{I}_K - \sum_{j=1}^{r} d_j \mathrm{L}^j$, with $\rho = \mathrm{diag}\{\rho_1, \cdots, \rho_K\}$, $\rho_k \in [-1, 1]$ for each $1 \leq k \leq K$, and $b(\mathrm{L}) = \mathrm{I}_K - \sum_{j=1}^{q} b_j \mathrm{L}^j$. Suppose $\lambda_{\max}(B) \leq 1 - \varepsilon_b$ and $\lambda_{\max}(D) \leq 1 - \varepsilon_d$, for $0 < \varepsilon_b < 1$ and $0 < \varepsilon_d < 1$, where B and D are companion matrices (see Section S4 in the Supplement for definitions) of VARMA coefficients $\{b_1, \cdots, b_q\}$ and $\{d_1, \cdots, d_r\}$, respectively.

The following proposition shows that this type of VARMA models can be subsumed within the class of models permitted by our Assumptions 1 and 2.

Proposition 1. Suppose that y_t follows the VARMA process defined in the Example.
(i). Assumption 1 holds.
(ii). Suppose that y_t does not have a finite-order VAR form. If

$$\overline{p} n^{1/2} (1 - \varepsilon_d)^{\underline{p}} \to 0, \tag{6}$$

then Assumption 2 holds.

Note that for the VARMA process defined in the Example the condition (6) requires the lower bound $\underline{p}$ to diverge, but it can diverge as slowly as $\log n$; see the remark following the proof of Proposition 1 in the Supplement.

If the data y_t follow the VAR process with a finite number p_{true} of lags, with $a_{p_{\text{true}}} \neq 0$ and $a_j = 0$ for $j \geq p_{\text{true}} + 1$, where the true lag order p_{true} is not necessarily known, Assumption 2(ii) is trivially satisfied if $\underline{p} \geq p_{\text{true}} + 1$.

We next state assumptions on the shock process u_t.

Assumption 3. (i) $\mathrm{E}(u_t | u_s, s \leq t-1) = 0$, almost surely (a.s.).

(ii). u_t is covariance-stationary and strong mixing with mixing numbers $\{\alpha(j) : j \geq 1\}$. There exist $\zeta > 2$, $\epsilon > 1$, and $C_\alpha < \infty$, such that $\alpha(j) \leq C_\alpha j^{-2\zeta\epsilon/(\zeta-2)}$, for all $j \geq 1$. [In other words, u_t is mixing of size $-2\zeta/(\zeta-2)$].

(iii). $\lambda_{\min}(\mathrm{E}(u_t u_t' | u_s, s \leq t-1)) \geq C_\lambda > 0$, a.s.

(iv). For ζ defined in (ii), $\mathrm{E}|u_t|^{8\zeta} \leq C_u < \infty$.

In Assumption 3, the condition (i) assumes that u_t is a martingale difference sequence

(MDS), and in other words, the dynamic model (2) for the conditional expectation of y_t is correctly specified. The mixing conditions in (ii) are regularity conditions, and they are only required in the proofs to establish uniform bounds for moments of sums (e.g. in proving the law of large numbers for squared sequences). These conditions are stronger than necessary. An alternative method is to directly impose assumptions on summability of cross-moments of the sequence $\{u_t\}$. We find that the approach based on mixing conditions is more intuitive and makes proofs relatively transparent. The mixing conditions are widely used in the literature to restrict the higher order serial dependence of time series, e.g. recently by Andrews and Guggenberger (2012, 2014). These conditions impose relatively weak restrictions on the variance dynamics, e.g. those generated by stationary GARCH and stochastic volatility models; see Carrasco and Chen (2002). The condition (iii) rules out singular conditional variances. The condition (iv) on existence of moments of u_t can be substantially weakened if stronger assumptions are imposed on the serial dependence of u_t, e.g. mean independence or conditional homoskedasticity.

3 Local projection regression

We now provide details of statistical inference. Throughout the paper, the time series data available to the econometrician are indexed by $t = 1, \cdots, n$. Assume that h and p are small enough so that $n \geq 3h + p - 3$. The LP estimator of $\beta_1(h, \mu)$, $\widehat{\beta}_1(h, \mu)$, is obtained by OLS (ordinary least squares) of (4) (for $t = p, \cdots, n-h$).[1]

By the partialling-out theorem,

$$\widehat{\beta}_1(h, \mu) = [\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)']^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h) y_{1t}(h, \mu), \tag{7}$$

where $\widehat{u}_t(h)$ is the *residualized* focal regressor y_t, obtained as OLS residuals of the VAR($p-1$) regression using the data $\{y_t : t = p, \cdots, n-h\}$.

The crux of the development of uniform distributional theory for $\widehat{\beta}_1(h, \mu)$ lies in the

[1] To estimate the cumulated response at the horizon h, empirical researchers can adopt the multiple-step method by running local projection regressions for all horizons up to h and then summing these estimates. For this method each regression uses a different sample size; the sample size reduces by one as the horizon increases by one. Our method based on the one-step regression (4) is asymptotically equivalent to the multi-step method and is neater for inference purpose.

fact that the effective regressor $\widehat{u}_t(h)$ asymptotically recovers the true shock u_t, as $p \to \infty$ (equivalently, the controlled lag order $p-1 \to \infty$) at an appropriate rate, so that the inference based on the estimator $\widehat{\beta}_1(h,\mu)$ is not asymptotically affected by the persistence level of the data y_t. This observation was made by MOPM, under the finite-order VAR(p_{true}) model, who achieved such *robustness via residualization* by proposing the lag-augmented LP regression (i.e. setting $p = p_{\text{true}} + 1$).

To study the asymptotic properties, replacing the focal regressor y_t in the LP(∞) regression (3) with the expression in the model (2), we can write

$$y_{1t}(h,\mu) = \beta_1(h,\mu)'u_t + \sum_{\ell=1}^{\infty} \gamma_{1\ell}(h,\mu)'y_{t-\ell} + \xi_{1t}(h,\mu), \tag{8}$$

where $\gamma_{1\ell}(h,\mu)' = \theta_{1\ell}(h,\mu)' + \beta_1(h,\mu)'a_\ell$. Combining (7) and (8) and applying the least squares algebra give that

$$\widehat{\beta}_1(h,\mu) - \beta_1(h,\mu) = [\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)']^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h)\psi_{1t}(h,\mu), \tag{9}$$

where $\psi_{1t}(h,\mu) = \sum_{\ell=p}^{\infty} \gamma_{1\ell}(h,\mu)'y_{t-\ell} + \xi_{1t}(h,\mu)$.

Before stating the main theorem of the paper, we impose the following assumption. Assumption 4 is similar to a technical assumption used in MOPM (their assumption 3), who argued as necessary for uniform inference results.[2]

Assumption 4. Let $\widetilde{X}_{t-1}(1) = y_{t-1}$ if $p = 1$, and $\widetilde{X}_{t-1}(p) = (y'_{t-1}, \Delta y'_{t-1}, \cdots, \Delta y'_{t-p+1})'$ if $p \geq 2$, where $\Delta y_t = y_t - y_{t-1}$. Denote $\pi_k(n) = \sum_{i=0}^{n-1} |\beta_k(i)|^2$, for $k = 1, \cdots, K$, and $\Pi(n) = \text{diag}\{\pi_1(n), \cdots, \pi_K(n)\}$. Let $\Upsilon_n(p) = (n-p)^{1/2}\text{diag}(\Pi(n)^{1/2}, \underbrace{\text{I}_K, \cdots, \text{I}_K}_{p-1 \text{ times}})$. Then

$$\lim_{M\to\infty} \lim_{n\to\infty} \sup_{\underline{p}\leq p\leq \overline{p}} \sup_{a\in\mathcal{A}} \text{P}\left(\lambda_{\min}\left(\Upsilon_n^{-1}(p) \textstyle\sum_{t=p+1}^{n} \widetilde{X}_{t-1}(p)\widetilde{X}'_{t-1}(p)\Upsilon_n^{-1}(p)\right) \geq 1/M\right) = 1.$$

We now establish the asymptotic normality of $\nu_1'\widehat{\beta}_1(h,\mu)$, where ν_1 is a known $K \times 1$ vector such that $|\nu_1| \neq 0$. Although we allow drifting sequences in the parameter space and

[2] See also Montiel Olea and Plagborg-Møller (2022).

allow the horizon and model order to increase with n, for notational simplicity we do not write a_i, p, or h explicitly as functions of n.

Theorem 1. Let $\pi_1(h,\mu) = \sum_{i=1}^{h} |\varphi_{1i}|^2$, where $\varphi_{1i} = \sum_{j=i}^{h} \mu_j \beta_1(j-i)$, and

$$V = \nu_1' \Sigma^{-1} \mathrm{Var}\left((n-h-p+1)^{-1} \sum_{t=p}^{n-h} u_t \xi_{1t}(h,\mu)\right) \Sigma^{-1} \nu_1 > 0,$$

where $\Sigma = \mathrm{E} u_t u_t' > 0$. Suppose that Assumptions 1, 2, 3 and 4 hold. Then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{\underline{p}\le p\le \overline{p}} \sup_{1\le h\le \overline{h}} \sup_{a\in\mathcal{A}} \left| \mathrm{P}\left(V^{-1/2}[\nu_1'\widehat{\beta}_1(h,\mu) - \nu_1'\beta_1(h,\mu)] \le x\right) - \Phi(x)\right| = 0, \tag{10}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Moveover,

$$\underline{C}_V \le \pi_1(h,\mu)^{-1}(n-h-p+1)V \le \overline{C}_V, \tag{11}$$

for two positive constants $\underline{C}_V$ and $\overline{C}_V$.

Remark 1 (Convergence rate). The pointwise convergence rate is given by $\pi_1(h,\mu)^{-1/2}n^{1/2}$, which depends on the persistence level of data, the horizon, and linear combination coefficients μ. The convergence rate for the cumulated response estimator is generally slower than that for the individual response estimator. If y_t follows an AR(1) process, the rates are $(\sum_{i=0}^{h-1} a_1^{2i})^{-1/2}n^{1/2}$ and $(\sum_{i=1}^{h}(\sum_{j=0}^{h-i}|a_1|^j)^2)^{-1/2}n^{1/2}$, respectively. The uniform rates are $h^{-1/2}n^{1/2}$ and $h^{-3/2}n^{1/2}$, respectively.

Remark 2. Under the finite-order VAR data generating process, MOPM developed uniform asymptotic theory for the individual response estimator. Specialized to the framework of MOPM, Theorem 1 extends their theory by using a more general form of the asymptotic variance, the validity of which does not require mean-independent shocks, thereby allowing for general MDS shocks.

3.1 LP is semiparametrically efficient

We now discuss the efficiency of LP estimators $\widehat{\beta}_1(h)$ of individual responses ($\mu = \mu_{\mathrm{IR}}$). We frame most of the discussions under *classical assumptions* on data and the design, i.e.

homoskedastic MDS shocks, and slope coefficients satisfying the stationarity assumption, and a fixed horizon. Instead of running the LP regression for each horizon, an alternative estimator of the impulse response is based on VAR implication. The method first estimates a truncated VAR model with p lags and then induces impulse response estimates recursively as a function of past responses and estimated VAR slopes. Under the VAR(∞) model, classical assumptions on data, and a fixed horizon, Plagborg-Møller and Wolf (2021) show that the LP and the VAR-implied estimators converge to the same estimand. Then, as the natural next question, how do LP and VAR-implied estimators compare, in term of *efficiency*?

Theorem 1 sheds light on this question. We first state a corollary. If u_t is conditionally homoskedastic ($\mathrm{E}(u_t u_t' | u_s, s \leq t-1) = \Sigma$), the expression for V can be simplified, and the pointwise version of Theorem 1 becomes

$$n^{1/2} \left(\nu_1' [\textstyle\sum_{i=0}^{h-1} \beta_1(i)' \Sigma \beta_1(i)] \Sigma^{-1} \nu_1 \right)^{-1/2} [\nu_1' \widehat{\beta}_1(h) - \nu_1' \beta_1(h)] \xrightarrow{d} \mathcal{N}(0,1). \tag{12}$$

Under classical assumptions on data and a fixed horizon, we can show that as the VAR lag order diverges at an appropriate rate, the VAR-implied estimator of $\beta_1(h)$ has the *same* asymptotic distribution as the one given in (12), by slightly extending the argument in Lütkepohl (1990) (see also Lütkepohl, 2005, eqn. (15.4.1)).

Thus LP possesses optimality properties which the VAR-implied estimator enjoys. In particular, under classical assumptions mentioned above, LP reaches the asymptotic efficiency bound in the sense of Chamberlain (1987) under the semiparametric (conditional) moment condition model $\mathrm{E}(y_t - \sum_{j=1}^{\infty} a_j y_{t-j} | y_{t-s}, s \geq 1) = 0$. If one further assumes Gaussianity, the LP is asymptotically Cramér-Rao efficient.

The optimality property of LP is in contrast to the well known result that the direct regression is less efficient than the iterative (VAR-implied) estimator, under the *finite-order* VAR model; see e.g. Bhansali (1997), Marcellino, Stock and Watson (2006), Xu (2020), among others, in slightly different contexts. The intuition behind such efficiency gain is in a sense similar to that of a dimension-reduction factor model; all impulse responses are functions of a relatively small number of common parameters, VAR slopes, so imposing such functional relation (if correctly specified) should yield the efficient estimator. However, the efficiency gain generated from such extraction via a parsimonious model diminishes as the dimension of the model grows, which eventually leads to our efficiency equivalence result.[3]

[3] Probably led by the inefficiency result on LP under finite-order VAR model, the literature has mixed

The discussions so far in this subsection assume stationarity. Note that the stationarity subspace is where the VAR-implied method potentially enjoys the most efficiency gain (under the finite-order model), so the equivalence result above is expected to hold more broadly. Nevertheless, the asymptotic distribution for the LP estimator given in Theorem 1 holds under much more general conditions on the parameter space, the range of horizons and shock dependence properties, than the classical assumptions maintained above. In Section 4 we will consider the inference based on LP, the generality of which is not shared so far by other existing approaches in the literature.

3.2 An alternative LP estimator

Instead of performing OLS on the regression (4), one can construct an alternative estimator utilizing the serial dependence structure of the error term $\xi_{1t}(h, \mu)$, when $h \geq 2$, in the hope of achieving "better" efficiency. On way is to replace the outcome variable $y_{1t}(h, \mu)$ with

$$\check{y}_{1t}(h, \mu) = y_{1t}(h, \mu) - \sum_{i=1}^{h-1} \mathring{\varphi}_{1i}' \mathring{u}_{t+i}$$

and then run OLS with the same set of regressors as in (4), where $\mathring{\varphi}_{1i} = \sum_{j=i}^{h} \mu_j \mathring{\beta}_1(j - i)$ with some preliminary estimates of $\beta_1(i)$ and u_t, denoted by $\mathring{\beta}_1(i)$ and $\mathring{u}_t$, respectively. This estimator is referred to as the alternative LP estimator, denoted by $\check{\beta}_1(h, \mu)$.

Lusompa (2022) showed that, for individual impulse responses ($\mu = \mu_{\text{IR}}$), under the stationary homoskedastic AR(1) model (with the true lag order known) the alternative estimator $\check{\beta}_1(h)$ is asymptotically more efficient than the LP estimator $\widehat{\beta}_1(h)$ across h in general. In Appendix B (Proposition 2), we show that in a simple setting such efficiency gain does *not* extend if a sufficiently large lag order is used; $\check{\beta}_1(h)$ and $\widehat{\beta}_1(h)$ are equivalently efficient. The discrepancy between the asymptotic distribution of $\check{\beta}_1(h)$ and that of the infeasible estimator (obtained by using true values of $\beta_1(i)$ and u_t in forming $\check{y}_{1t}(h, \mu)$) is due to the estimation error in $\mathring{u}_t$. Suppose that $\mathring{u}_t$ is obtained by a VAR(p_u) regression of y_t on their lags. If the

conjecture on the efficiency of LP under the *infinite-order model*. Lusompa (2022, footnote 5) mentioned that "... in the infinite lag case ... most people would probably assume this (LP is less efficient than VAR-implied estimator)". Focusing on identification and consistency, Plagborg-Møller and Wolf (2021, section 2.5) conjectured that LP and VAR are equally efficient under stationary VAR(∞) model for a fixed horizon, but did not provide the analysis.

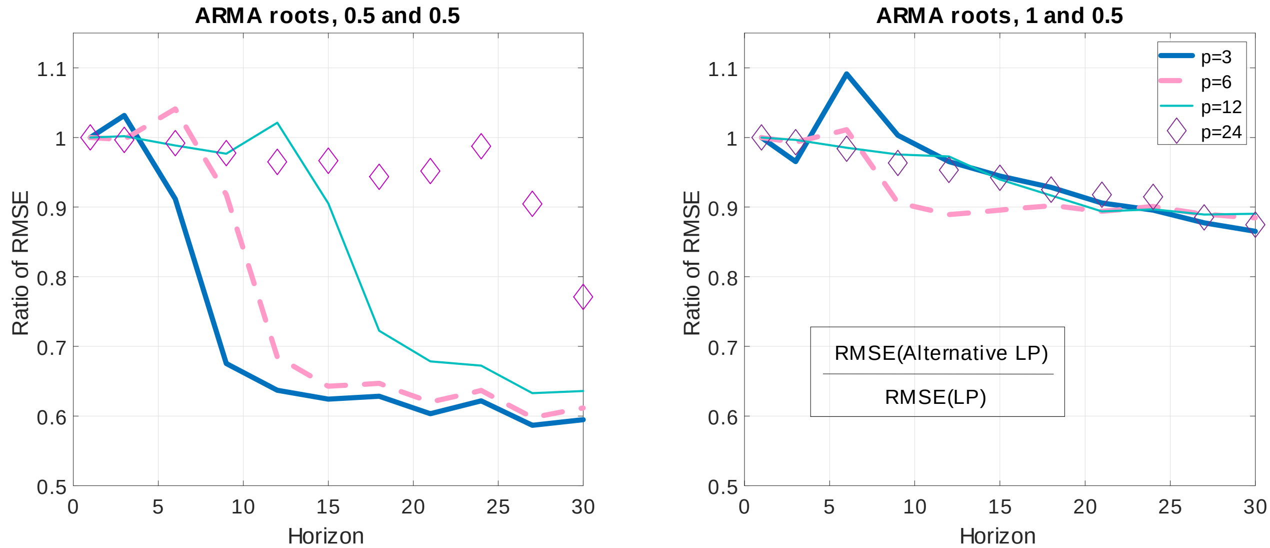

Figure 1: Ratio of RMSEs of $\check{\beta}_1(h)$ and $\widehat{\beta}_1(h)$ in simulations. The model is ARMA(1,1). The sample size $n = 240$. The LP regression order $p \in \{3, 6, 12, 24\}$.

lag order p_u is sufficiently large, estimation errors of the first $h-1$ VAR(p_u) slopes enter the asymptotic distribution of $\check{\beta}_1(h)$, which leads to the distributional equivalence of $\check{\beta}_1(h)$ and $\widehat{\beta}_1(h)$.

In Figure 1 we report a simple simulation study. The figure shows the ratio of finite-sample root mean square errors (RMSEs) of the standard and alternative LP estimators, $\widehat{\beta}_1(h)$ and $\check{\beta}_1(h)$, as a function of response horizons. The data are generated from a scalar ARMA(1,1) model, $y_t = \rho y_{t-1} + u_t + 0.5u_{t-1}$, where u_t is I.I.D. (independent and identically distributed) with the standard normal distribution. With a sample size $n = 240$, we use the lag order $p \in \{3, 6, 12, 24\}$. For highly persistent data ($\rho = 1$), the RMSE advantage of $\check{\beta}_1(h)$, if there any, is fairly small over all horizons and all lag orders. For mildly persistent data ($\rho = 0.5$), $\check{\beta}_1(h)$ has a smaller MSE than $\widehat{\beta}_1(h)$ at medium and long horizons but the difference clearly shrinks as the lag order increases, which corroborates our theoretical equivalence result above.

While we defer more comprehensive investigation of $\check{\beta}_1(h)$ (especially under general conditions of Theorem 1) to further study, given the optimality result discussed in Section 3.1, we do not expect the efficiency gain of the alternative estimator over LP asymptotically in the VAR(∞) model at least under the classical setting of Plagborg-Møller and Wolf (2021) (stationarity, homoskedasticity and fixed horizons). Nevertheless, as mentioned in the introduction, how to draw robust inference based on the alternative LP estimator remains unknown.

3.3 Further remarks on Theorem 1

Conditions on the lag order. Under fixed horizon we only require $\overline{p}^2/n \to 0$ (Assumption 2(i)). Even with stronger assumptions on data dependence (stationarity) and sometime on shocks, the asymptotic theory for VAR-implied impulse response estimators often assumes a more stringent restriction on the lag order, $\overline{p}^3/n \to 0$. Such restriction is inherited from the joint limit theory for slope matrices of all p lags of VAR model (Lewis and Reinsel, 1985, Gonçalves and Kilian, 2007), while our LP method only estimates one single slope matrix (i.e. coefficient matrix of y_t) and treats other $p-1$ lags as control variables.

For asymptotic pivotalness of test statistics, it typically assumes $\overline{p}^4/n \to 0$ even under stationarity (Gonçalves and Kilian, 2007, theorem 2.2) for VAR-implied estimators. The alternative LP-based estimator considered by Lusompa (2022) also requires $\overline{p}^4/n \to 0$ for his asymptotic theory under stationarity. For our estimator, we do not strengthen the restrictions on the lag order when establishing the asymptotic distribution for test statistics; see Section 4.

Asymptotic variance. Note that Theorem 1 holds under both the finite-order VAR and VAR(∞) models, and we will show (in Section 4) that the asymptotic variance will be estimated in the same way under either model. This is in contrast to the VAR-implied impulse response estimator, the asymptotic variance of which is estimated differently under two different models (Kilian and Lütkepohl, 2017, section 12.1.3).[4] Such continuity of the asymptotic distribution over models further reinforces our consideration of LP for uniform inference.

Martingale representation of the score. The key tool to the proof of Theorem 1 is a martingale representation. Examining (9), and ignoring smaller-order terms, the (effective)

[4] Such discontinuity of inference methods based on VAR-implied estimator can be mitigated by applying the bootstrap method to a *non-pivotal* statistic which does not involve variance estimation (Inoue and Kilian, 2002). Nevertheless, existing literature (which mostly focuses on stationary univariate AR model) establishes the validity over narrower spaces of model parameters, response horizon, and model lag order than those our setting allows.

score $\sum_{t=p}^{n-h} u_t \xi_{1t}(h,\mu)$ which plays a major role in the asymptotic theory can be rewritten as (see Lemma MART for a more general form)

$$\sum_{t=p}^{n-h} u_t \xi_{1t}(h,\mu) = \sum_{t=p+1}^{n} w_t, \tag{13}$$

where

$$w_t = \left[\sum_{i=1}^{h} \mathbb{I}_{\{p \le t-i \le n-h\}} u_{t-i} \varphi_{1i}'\right] u_t.$$

The advantage of the representation (13) is that, even that score contributions $u_t \xi_{1t}(h,\mu)$ are generally autocorrelated, the transformed summands w_t are (conditionally and unconditionally heteroskedastic) martingale differences with respect to their natural filtration, given that u_t is assumed to be an MDS. This representation provides basis for establishing the asymptotic theory, by allowing a straight application of the martingale central limit theorem (CLT).

The approach of MOPM, in a more restricted setting, deals directly with the raw (untransformed) score. To apply the CLT the approach relies on a reverse-time argument, which consequentially requires shocks to be mean independent with the future, $\mathrm{E}(u_t | u_s, s \ge t+1) = 0$. Dispensing with such assumption of mean independence, although desirable for application purposes, brings a technicality cost of more involved calculations of uniform bounds for moments in the proof; see Section S1 in the Supplement to the paper.

Although the equality (13) is a simply algebraic rearrangement, it has a deeper root. It connects with the classical argument in obtaining the central limit theorem for generic stationary and serially correlated sequences, like the score contributions $u_t \xi_{1t}(h,\mu)$ (for a given h) in our context. A general method is known as Gordin's (1969) approach, in which a martingale approximation plays a central role.[5] Adapting this tool to score contributions $u_t \xi_{1t}(h,\mu)$ gives the representation (13). Note that the representation (13) is exact, instead of being approximate, due to the moving-average form of the LP(∞) error $\xi_{1t}(h,\mu)$. In the next section we show that the martingale representation (13) is not just a technical trick, but also provides basis for construction of the standard error detailed there.

[5] See Beveridge and Nelson (1981), Phillips and Solo (1992), Wu and Woodroofe (2004), and Cuny and Merlevède (2014), among others, for applications of the tool of martingale approximation and further development.

4 Two standard errors

To estimate the asymptotic variance V and form test statistics for inference, the general HAR methods (Heteroskedasticity and Autocorrelation Robust methods) appear natural; see e.g. Lazarus, et al. (2018), for recent development on the general topic. However, there are prevalent reservations about their applications to the LP regression, especially when the data are highly persistent and the horizon is long; see Herbst and Johannsen (2021) for a recent discussion. In this section we consider two simple standard errors which adapt to the LP regression and, unlike HAR methods, do not involve determination of tuning parameters.

The first variance estimator for $\nu_1'\widehat{\beta}_1(h,\mu)$ is of Eicker-Huber-White sandwich type,

$$\widehat{V}_{HC} = (n-h-p+1)^{-2}\nu_1'\widehat{\Sigma}(h)^{-1} \cdot \left[\sum_{t=p}^{n-h} \widehat{\eta}_{1t}(h,\mu)^2 \widehat{u}_t(h)\widehat{u}_t(h)'\right] \cdot \widehat{\Sigma}(h)^{-1}\nu_1,$$

where $\widehat{\eta}_{1t}(h,\mu)$ is the OLS residual of (4), and $\widehat{\Sigma} = (n-h-p+1)^{-1}\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)'$. The variance estimator $\widehat{V}_{HC}$, specialized to the one for the individual response, was proposed by MOPM under the finite-order VAR(p_{true}) model for the so-called lag-augmented local projection regression, i.e. the regression (1) where the model order p is set as $p = p_{\text{true}} + 1$.[6]

We study the variance estimator $\widehat{V}_{HC}$ under the general framework introduced in Section 2. It is shown that $\widehat{V}_{HC}$ can recover V asymptotically even for the regression (4) with serially dependent errors, but *only when* such serial dependence in errors does not cause serial correlation in score contributions, which can be generated by conditional heteroskedasticity of unknown form. Thus the variance estimator $\widehat{V}_{HC}$, despite its remarkable simplicity and robustness, is not truly "heteroskedasticity-consistent (HC)", even that the notation carries the usual subscript HC.

The t-statistic using the variance estimator $\widehat{V}_{HC}$ is constructed as $\widehat{S}_{HC} = \widehat{V}_{HC}^{-1/2}[\nu_1'\widehat{\beta}_1(h,\mu) - \nu_1'\beta_1(h,\mu)]$. The following theorem describes the asymptotic behavior of $\widehat{S}_{HC}$ under the VAR(∞) model.

[6] Under finite-order VAR model, Dufour et al. (2006) also discussed the robustness of the lag augmented regression to persistent data but suggested using a more complicated standard error involving the choice of a tuning parameter.

Theorem 2. Suppose that Assumptions 1, 2, 3 and 4 hold. Then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{\underline{p}\leq p\leq\overline{p}} \sup_{1\leq h\leq\overline{h}} \sup_{a\in\mathcal{A}} \left| \mathrm{P}\left((V_{HC}/V)^{1/2}\widehat{S}_{HC} \leq x \right) - \Phi(x) \right| = 0,$$

where $V_{HC} = (n-h-p+1)^{-2}\nu_1'\Sigma^{-1}\sum_{t=p}^{n-h} \mathrm{E}u_t u_t' \xi_{1t}(h,\mu)^2\Sigma^{-1}\nu_1$.

Theorem 2 shows that the consistency of the variance estimator $\widehat{V}_{HC}$ clearly requires the equality $\mathrm{Var}(\sum_{t=p}^{n-h} u_t\xi_{1t}(h,\mu)) = \sum_{t=p}^{n-h} \mathrm{E}\xi_{1t}(h,\mu)^2 u_t u_t'$, which holds under the following assumption.

Assumption 5. The score contribution process $\{u_t\xi_{1t}(h,\mu),\ t = 1, 2, \cdots\}$ is serially uncorrelated.

Note that Assumption 5 is implied by, thus weaker than, commonly used assumptions on u_t such as conditional homoskedasticity or mean independence. Assumption 5 is relatively easy to check empirically.

To give concrete examples, consider a simple MDS $u_t = e_t e_{t-1}$, where e_t is I.I.D. with zero mean, unit variance and $\mathrm{E}e_t^3 \neq 0$. Simple calculations show that $\mathrm{Cov}(u_t, u_{t+1}^2 u_{t-1}^2) = (\mathrm{E}e_t^3)^2 \neq 0$. Thus u_t violates the mean independence, which by definition would require u_t to be uncorrelated with any measurable function of random variables in the set $\{u_s : s \neq t\}$. Nevertheless, the process $u_t\xi_t(h,\mu)$ is still serially uncorrelated in this example, thus Assumption 5 is satisfied.

Assumption 5 can be violated. Consider a slightly modified example, the MDS $u_t = e_t|e_{t-1}|$, where e_t is the same as in the last example. Examination shows that the process $u_t\xi_t(h,\mu)$ is serially correlated. This example is a simplification of an MDS with GARCH-type and stochastic volatility functions, which generally violate Assumption 5 if the innovation has non-zero third moment. Non-zero or even time-varying third moment is not uncommon in applications (Hansen, 1994, Conrad, et al., 2013, Colacito, et al., 2016). In the Supplement to the paper (Section S6), we provide empirical evidence that Assumption 5 may not be satisfied.

To derive lower-level conditions for Assumption 5, denote score contributions as $s_t(h,\mu) =$

$u_t\xi_{1t}(h,\mu)$. Then the autocovariance matrices of $s_t(h,\mu)$ are given by

$$\mathrm{E}s_t(h,\mu)s_{t-j}(h,\mu)' = \begin{cases} \sum_{i=j+1}^{h} \mathrm{E}[u_{t+j}u_t'\varphi_{1i}'u_{t+i}u_{t+i}'\varphi_{1,i-j}], & \text{if } 1 \le j \le h-1, \\ 0, & \text{if } j \ge h. \end{cases}$$

Although score contributions are serially uncorrelated beyond $h-1$ lags, the autocorrelations can play a role for smaller lags. Assumption 6 below imposes restrictions directly on the fourth moments of u_t, and is sufficient for Assumption 5 for all $h \ge 1$.

Assumption 6. Write $u_t = (u_{1t}, \cdots, u_{Kt})'$. Assume $\mathrm{E}u_{t-i}u_{t-j}'u_{k_1t}u_{k_2t} = 0$ for $i > j > 0$ and $k_1, k_2 = 1, \cdots, K$.

Like Assumption 5, Assumption 6 is weaker than assuming conditional homoskedasticity or future mean independence on u_t. The scalar version of Assumption 6 appeared earlier in the literature, and was imposed for inference validity of time series models in different contexts (e.g. Deo, 2000, condition A(vii); Gonçalves and Kilian, 2004, assumption A'(iv')). For example, Gonçalves and Kilian (2004) show that a condition like Assumption 6 is required for the validity of the recursive residual-based wild bootstrap scheme for stationary AR model with conditional heteroskedasticity of unknown form.

We now introduce a new standard error which does not require Assumption 5. By the martingale representation (13), we have

$$\mathrm{Var}\left(\sum_{t=p}^{n-h} u_t\xi_{1t}(h,\mu)\right) = \sum_{t=p+1}^{n} \mathrm{E}w_tw_t'. \tag{14}$$

The equality (14) motivates the following estimator of V, referred to as the martingale (MG) variance estimator,

$$\widehat{V} = (n-h-p+1)^{-2}\nu_1'\widehat{\Sigma}(h)^{-1}\left(\sum_{t=p+1}^{n} \widehat{w}_t\widehat{w}_t'\right)\widehat{\Sigma}(h)^{-1}\nu_1, \tag{15}$$

where $\widehat{w}_t = [\sum_{i=1}^{h} \mathbb{I}_{\{p\le t-i\le n-h\}}\widehat{u}_{t-i}(h)\widetilde{\varphi}_{1i}']\widetilde{u}_t$ and $\widetilde{\varphi}_{1i} = \sum_{j=i}^{h} \mu_j\widetilde{\beta}_1(j-i)$. In (15), $\widetilde{\beta}_1(i)$ and $\widetilde{u}_t$ are preliminary estimates of $\beta_1(i)$ and u_t, respectively, and both estimates should converge sufficiently fast so that the following Assumption 7 is satisfied. Note that $\widehat{V}$ differs from the sandwich variance estimator $\widehat{V}_{HC}$ only in the middle part.

Assumption 7. The preliminary estimates $\widetilde{\beta}_1(i)$ and $\widetilde{u}_t$ in (15) are such that

(i) $\lim_{n\to\infty} \sup_{\underline{p}\leq p\leq \overline{p}} \sup_{1\leq h\leq \overline{h}} \sup_{a\in\mathcal{A}} \mathrm{P}\left(h\pi_1(h,\mu)^{-1}\sum_{i=1}^{h}|\widetilde{\varphi}_{1i}-\varphi_{1i}|^2 > M\right) = 0,$ for all $M > 0$;

(ii) $\lim_{M\to\infty}\lim_{n\to\infty} \sup_{\underline{p}\leq p\leq \overline{p}} \sup_{a\in\mathcal{A}} \mathrm{P}\left(p^{-2}\sum_{t=p+1}^{n}|\widetilde{u}_t - u_t|^2 > M\right) = 0.$

For an example of $\widetilde{\beta}_1(i)$, suppose that our initial interest is in the individual impulse response ($\mu = \mu_{\mathrm{IR}}$). Consider the LP estimator for $\widetilde{\beta}_1(i)$. Theorem 1 shows that $\sum_{i=0}^{h-1}|\widetilde{\beta}_1(i) - \beta_1(i)|^2 = O_p(n^{-1}\sum_{i=0}^{h-1}\pi_1(i))$, thus Assumption 7 (i) is satisfied provided $h^2/n \to 0$. The fitted errors $\widetilde{u}_t$ can be obtained by running OLS on the VAR(p) regression. The following theorem shows the uniform validity of the test based on the MG variance estimator $\widehat{V}$.

Theorem 3. Suppose that Assumptions 1, 2, 3, 4 and 7 hold. Let $\widehat{S} = \widehat{V}^{-1/2}[\nu_1'\widehat{\beta}_1(h,\mu) - \nu_1'\beta_1(h,\mu)]$. Then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{\underline{p}\leq p\leq \overline{p}} \sup_{1\leq h\leq \overline{h}} \sup_{a\in\mathcal{A}} \left|\mathrm{P}(\widehat{S}\leq x) - \Phi(x)\right| = 0.$$

Remark 3. Theorem 3 shows that estimation of unknown parameters in the asymptotic variance of $\widehat{\beta}_1(h,\mu)$ has asymptotically negligible effects on the test based on the MG standard error (as long as these estimates satisfy the constraints in Assumption 7). This is in contrast to the method of using these estimates (under similar constraints) in constructing the point estimator, as shown in Section 3.2, which generally has non-trivial effects on (increases) the asymptotic variance.

Remark 4. The MG variance estimator can be extended to infer about cross-equation restrictions. Such restrictions are useful when empiricists want to learn responses of several macro variables to an economic shock. The interest is thus in the response matrix $\beta(h,\mu) = \sum_{j=1}^{h}\mu_j\beta(j)$, which can be estimated by OLS on the regression (4) for each response variable $y_{kt}(h,\mu)$, where $k = 1,\cdots,K$. Denote the estimator as $\widehat{\beta}(h,\mu) = \sum_{t=p}^{n-h} y_t(h,\mu)\widehat{u}_t(h)'[\sum_{t=p}^{n-h}\widehat{u}_t(h)\widehat{u}_t(h)']^{-1}$, where $y_t(h,\mu) = (y_{1t}(h,\mu),\cdots,y_{Kt}(h,\mu))'$. Let ν be a $d_\nu \times K^2$ matrix of constants. To draw inference for the $d_\nu \times 1$ vector $\nu\mathrm{vec}(\beta(h,\mu))$, we can construct the MG variance matrix estimator $\widehat{\Omega}$ as

$$\widehat{\Omega} = (n-h-p+1)^{-2}\nu(\widehat{\Sigma}(h)^{-1}\otimes \mathrm{I}_K)\left(\sum_{t=p+1}^{n}\widehat{W}_t\widehat{W}_t'\right)(\widehat{\Sigma}(h)^{-1}\otimes \mathrm{I}_K)\nu',$$

where $\widehat{W}_t = [\sum_{i=1}^{h}\mathbb{I}_{\{p\le t-i\le n-h\}}(\widehat{u}_{t-i}(h)\otimes \mathrm{I}_K)\widetilde{\varphi}_i]\widetilde{u}_t$, $\widetilde{\varphi}_i = (\widetilde{\varphi}_{1i},\cdots,\widetilde{\varphi}_{Ki})'$ and $\widetilde{\varphi}_{ki} = \sum_{j=i}^{h}\mu_j\widetilde{\beta}_k(j-i)$, for $k=1,\cdots,K$, with $\widetilde{\beta}_k(j)$ and $\widetilde{u}_t$ being preliminary estimates. The Wald statistic is then formed as $\{\nu\mathrm{vec}[\widehat{\beta}(h,\mu)-\beta(h,\mu)]\}'\widehat{\Omega}^{-1}\nu\mathrm{vec}[\widehat{\beta}(h,\mu)-\beta(h,\mu)]$. If $\nu\mathrm{vec}(\beta(h,\mu))$ reduces to the scalar $\nu_1'\beta_1(h,\mu)$ for a K-dimensional vector ν_1, the Wald statistic reduces to the square of $\widehat{S}$ defined in Theorem 3.

Remark 5 (Finite-order VAR and lag augmentation). In the literature it is often assumed that the data follow a VAR process with a finite number p_{true} of lags. Under such finite-order VAR model, Assumption 2(ii) requires lag augmentation, $\underline{p} \ge p_{\text{true}}+1$. While this assumption is essential for the consistency of $\widehat{V}_{HC}$, it can be weakened as $\underline{p} \ge p_{\text{true}}$ (i.e. lag augmentation is not needed) for the validity of the MG variance estimator $\widehat{V}$, provided that at least two lags are used in the LP regression (4) (i.e. $p \ge 2$).

To illustrate, let R_t be the regression residual of y_t on $y_{t-1},\cdots,y_{t-p+1}$. The requirement $p\ge 2$ guarantees that for inference of $\beta_1(h,\mu)$, the *effective* regressor R_t in the local projection regression (4) is stationary if the data y_t are not integrated or nearly integrated of order two or more (Assumption 1). The condition $p \ge 2$ essentially requires the presence of control variables in the regression. The effective regression score is now $\sum_{t=p}^{n-h} R_t\xi_{1t}(h,\mu)$, and importantly, $R_t \ne u_t$ if we set $p = p_{\text{true}}$ and the model is VAR(p_{true}) (i.e. $a_{p_{\text{true}}} \ne 0$). Note that the martingale representation (13) holds algebraically, which now becomes $\sum_{t=p}^{n-h} R_t\xi_{1t}(h,\mu) = \sum_{t=p+1}^{n} w_t^R$, with $w_t^R = (\sum_{i=1}^{h}\mathbb{I}_{\{p\le t-i\le n-h\}}R_{t-i}\varphi_{1i}')u_t$. The key argument to justify the MG variance estimator $\widehat{V}$ still goes through even that R_t does not recover the shock u_t: w_t^R is an MDS as long as u_t is an MDS, since R_t only depends on current and past values of u_t. We thus continue to have the equality

$$\mathrm{Var}\left(\sum_{t=p}^{n-h} R_t\xi_{1t}(h,\mu)\right) = \sum_{t=p+1}^{n}\mathrm{E}w_t^R w_t^{R\prime},$$

a premise of the validity argument for the MG variance estimator $\widehat{V}$.

In contrast, as shown by MOPM, lag augmentation is crucial to justify the variance estimator $\widehat{V}_{HC}$. Without lag augmentation, the fact that R_t is not a white noise causes the non-zero serial correlation of the process $R_t\xi_{1t}(h,\mu)$ (*even when* u_t satisfies the full mean independence assumption). This in turn yields $\text{Var}(\sum_{t=p}^{n-h} R_t\xi_{1t}(h,\mu)) \neq \sum_{t=p}^{n-h} \text{E}R_tR_t'\xi_{1t}(h,\mu)^2$, thereby invalidating the variance estimator $\widehat{V}_{HC}$ and consequently affecting the inference based on $\widehat{S}_{HC}$.

5 Simulation experiments

5.1 The designs and methods

In this section we investigate the finite-sample performance of a few inference methods for individual impulse responses $\beta(h)$ based on local projection regression, including those analyzed in the paper. Consider the following univariate data-generating processes (DGPs):

$$\begin{aligned} \text{AR(1)}: &\quad (1-\rho\text{L})y_t = u_t, \\ \text{AR(2)}: &\quad (1-\rho\text{L})(1-0.5\text{L})y_t = u_t, \\ \text{ARMA(1,1)}: &\quad (1-\rho\text{L})y_t = (1+0.5\text{L})u_t, \end{aligned}$$

where the shock u_t is either I.I.D. with the standard normal distribution, or follows a conditionally heteroskedastic process. The combinations of three conditional mean specifications and two shock specifications yield six DGPs in total. For conditionally heteroskedastic shocks, we use the exponential GARCH (EGARCH(1,1)) model. In particular, $u_t = \sigma_t e_t$, where $\ln\sigma_t^2 = -0.23 + 0.95\ln\sigma_{t-1}^2 + 0.25[|e_{t-1}| - (2/\pi)^{1/2}] - 0.3e_{t-1}$, and e_t is an I.I.D. sequence with the standard normal distribution. We then standardize each realization of the path of u_t so that it has unit variance. The parameter choices in the EGARCH model are close to those used in the simulation study of Gonçalves and Kilian (2004, table 4). The autoregressive root ρ is set as 0.0, 0.9, 0.98 or 1, covering persistent and highly persistent data. The number of realizations in simulations is 10,000. We consider two sample sizes $n = 240$ and $n = 1200$, the latter of which is less realistic but is useful to demonstrate the asymptotic behavior of inference procedures. We consider ten integer horizons ranging from 1 to 60. For all DGPs, we use the sieve VAR model, in which the lag order $\widehat{p}_{\text{true}}$ is selected using Akaike information criterion (AIC).

We consider five methods to construct confidence interval for $\beta(h)$ based on the local projection regression (1). The first method (referred to as LALP-HC) is based on the test statistic proposed by MOPM. The method runs the lag-augmented local projection regression (i.e. setting $p = \widehat{p}_{\text{true}} + 1$) and uses the variance estimator $\widehat{V}_{HC}$. The second method (referred to as LALP-HAR) runs the lag-augmented local projection regression, as the first one does, but applies the general HAR approach to inference. To implement, we follow the recommended procedure by Lazarus et al. (2018) to use the equally weighted cosine (EWC) long-run variance estimator and Student's t critical value. This proposal, a competitor to ours following immediately after, provides a natural modification to the MOPM statistic for researchers who would like to draw robust inference of $\beta(h)$ to both highly persistent data and unknown serial correlation in score contributions.

The third method (referred to as LALP-MG) is based on the newly proposed martingale (MG) variance estimator $\widehat{V}$ (and the test statistic $\widehat{S}$), again, after running the lag-augmented local projection regression. The fourth method (referred to as hybrid-MG) is based on a hybrid of the lag-augmented and standard (without lag augmentation) LP regressions, entertaining the controlled lag-order choice flexibility of the MG-standard-error-based inference, discussed in Remark 5. This method uses the MG variance estimator $\widehat{V}$, but in the local projection regression, sets the lag order $p = \widehat{p}_{\text{true}} + 1$ if $\widehat{p}_{\text{true}} = 1$ or $h = 1$, but using $p = \widehat{p}_{\text{true}}$ otherwise; lag augmentation is implemented only when needed. As the last method (referred to as LP-HC), we also consider the standard local projection regression, without lag augmentation (i.e. setting $p = \widehat{p}_{\text{true}}$), coupled with the variance estimator $\widehat{V}_{HC}$. This approach is closely related to the first method listed above, LALP-HC, and is considered in our study for the purpose of sensitivity analysis to the lag order (e.g. shedding light on the question, how would LALP-HC behave if the lag order is under-selected by one?). Note that the hybrid method would not be valid when the HC standard error is used, under the finite-order VAR model. For all methods except LALP-HAR, we use the standard normal critical value. The nominal coverage level is set as 90% in simulations.

5.2 The results

The full set of simulation results on actual coverage and length (adjusted for correct coverage) of confidence intervals is reported in the Supplement (Section S7). In Figures 2 and 3 here we highlight the results on the coverage under AR(2) and ARMA(1,1) data generating processes with $\rho = 0.98$, which provide a snapshot of the findings we summarize below.

LALP-HC v.s. LALP-MG. The results show that for I.I.D. shocks (under which LALP-

HC and LALP-MG are asymptotically equivalent), LALP-MG confidence intervals have better coverage than LALP-HC ones when $n = 240$, especially at long horizons, but the difference almost disappears when $n = 1200$. For EGARCH shocks (which generate serial correlation in score contributions), LALP-MG has notably better coverage than LALP-HC at short, medium and long horizons when $n = 240$, and the difference is even larger and substantial when $n = 1200$. This is true across all three conditional mean designs considered, AR(1), AR(2), and ARMA(1,1). LALP-HC and LALP-MG have overall similar adjusted length. When $n = 240$, in most cases LALP-MG has shorter adjusted length than LALP-HC but the difference is quite small.

LALP-HAR, as an alternative inferential method (to LALP-MG) of accommodating correlated score contributions, does not seem to help mitigate the coverage distortion of LALP-HC under our designs. The method either performs similarly to LALP-HC, and in many cases (especially under EGARCH shocks when we expect LALP-HAR to work), has even worse coverage than LALP-HC. Under ARMA(1,1) model, LALP-HAR has the lowest coverage overall among all methods considered. It has overall similar adjusted length to, and in many cases greater length than, LALP-HC.

Hybrid-MG appears to have the best coverage overall among all methods considered. Compared to LALP-MG, hybrid-MG has notably better coverage under AR(2) model. Under AR(2) model, if the AIC correctly identifies the lag order (which happens for most realizations), the difference between hybrid-MG and LALP-MG is essentially in lag augmentation or not. Not to lag augment leads to better covered (but somewhat wider) confidence intervals. We expect these observations to extend to AR models of first few higher orders. Under AR(1) and ARMA models, the two methods work almost identically, either because of the design or the diminishing effect of lag augmentation when the chosen lag order is large.

Lastly we observe that although LP-HC produces most under-covered confidence intervals among all methods under AR(1) model, the results are much less contrasting under AR(2) model. LP-HC performs almost identically to LALP-HC under the ARMA model. In summary, our simulations show that although the trick of lag augmentation tends to play a less important role in inference as the lag order deviates from the very first few integers (which is probably realistic in applications), the potential improvement generated from using the MG standard error appears to be fairly consistent across different models and horizons.

6 Concluding remarks

Local projection is simple to implement. Recent research shows that the method of LPs has advantage, on top of its implementational simplicity, in drawing *inference* of impulse responses over competing methods, especially regarding uniform validity. The current paper contributes to the literature by showing that in a realistic setting, local projections can actually be more efficient than previously thought. LP is potentially (one of) the most efficient estimators for the impulse response under martingale difference shocks if the controlled lag order diverges. Using a large number of lags is practical in that the finite-order VAR model has been argued to be most plausibly thought as the approximation of the actual data generating processes which are implied by macro models.

In this paper, we study the asymptotic properties and propose novel inference methods for the local projection regression which allow researchers to remain relatively agnostic of persistence levels and form of heteroskedasticity in the data. A potential future research direction can be navigation of the bias and variance trade-off in finite samples when the local projection regression is long, in the presence of many controlled lags of multiple endogenous variables, possibly through some shrinkage-type (machine learning) methods.

Appendix A: Proofs

In this appendix, we provide the proof of Theorem 1. The proofs of Proposition 1 and Theorems 2 and 3 are provided in the Supplement to the paper (Sections S2, S3 and S4).

Proof of Theorem 1. Following Andrews, et al. (2020) and MOPM, we will show $V^{-1/2}[\nu_1'\widehat{\beta}_1(h,\mu) - \nu_1'\beta_1(h,\mu)] \xrightarrow{d} \mathcal{N}(0,1)$, for any sequence $\{a = a(n)\} \in \mathcal{A}$, and any sequences $\{h = h(n)\}$ and $\{p = p(n)\}$ of positive integers satisfying $1 \leq h(n) \leq \overline{h}$, $1 \leq \underline{p} \leq p(n) \leq \overline{p}$ where $\overline{h}/n \to 0$ and $\overline{h}\overline{\mu}^2\overline{p}^2/n \to 0$, and assumptions of Theorem 1. Having this in mind, in what follows we do not write a, $\beta_1(i,\mu), p$ or h explicitly as functions of n, for notational simplicity. For a deterministic function $f(p,h,a)$, write $\sup_{\underline{p}\leq p\leq\overline{p}}\sup_{1\leq h\leq\overline{h}}\sup_{a\in\mathcal{A}} f(p,h,a)$ as $\sup_{p,h,a} f(p,h,a)$ for simplicity. The proof of (10) consists of three parts.

Part I. We first show asymptotic normality for the (mean-zero) score $\mathcal{T}_n = \sum_{t=p}^{n-h} \tau' u_t \xi_{1t}(h,\mu)$, where $\tau = \Sigma^{-1}\nu_1$. [It is given in (20)].

Recall that $w_t = (\sum_{i=1}^{h} u_{t-i}c'_{ti})u_t$ is a martingale difference array, where $c_{ti} = I_{ti}\varphi_{1i}$ with $I_{ti} = \mathbb{I}_{\{p\le t-i\le n-h\}}$. Using the martingale representation (13), we can write

$$\begin{aligned} \mathcal{T}_n = \sum_{t=p}^{n-h} \tau' u_t \xi_{1t}(h,\mu) &\overset{(13)}{=} \sum_{t=p+1}^{n} \tau' w_t \\ &= \underbrace{\sum_{t=p+1}^{p+h-1} \tau' w_t}_{=\mathcal{T}_{n1}} + \underbrace{\sum_{t=p+h}^{n-h+1} \tau' w_t}_{=\mathcal{T}_{n2}} + \underbrace{\sum_{t=n-h+2}^{n} \tau' w_t}_{=\mathcal{T}_{n3}} \qquad (16) \\ &\triangleq \mathcal{T}_{n1} + \mathcal{T}_{n2} + \mathcal{T}_{n3}. \end{aligned}$$

Given the representation above, it is convenient to define the index sets $\mathcal{S} = \{t : p+h \le t \le n-h+1\}$ and $\mathcal{S}^c = \{p+1, \cdots, n\}\backslash\mathcal{S}$. For $t \in \mathcal{S}$, we have $w_t = (\sum_{i=1}^{h} u_{t-i}\varphi'_{1i})u_t$, thus the distribution of w_t does not depend on t.

Among the three terms of (16), $\mathcal{T}_{n2}$ is the dominant one. Lemma MOMT-W(i) shows that $\underline{C}_w \le \pi_1(h,\mu)^{-1}\mathrm{E}(\tau' w_t)^2 \le \overline{C}_w$, for $t \in \mathcal{S}$ and two positive constants $\underline{C}_w$ and $\overline{C}_w$. Since $\mathrm{E}\mathcal{T}_{n2}^2 = (n-2h-p+2)\mathrm{E}(\tau' w_t)^2$, we thus have the asymptotic order of $\mathrm{E}\mathcal{T}_{n2}^2$:

$$\underline{C}_w \le \pi_1(h,\mu)^{-1}(n-2h-p+2)^{-1}\mathrm{E}\mathcal{T}_{n2}^2 \le \overline{C}_w. \qquad (17)$$

For other two terms $\mathcal{T}_{n1}$ and $\mathcal{T}_{n3}$, by Lemma MOMT-W(i) (eqn.(30)),

$$\mathrm{E}\mathcal{T}_{n1}^2 = \sum_{t=p+1}^{p+h-1} \mathrm{E}(\tau' w_t)^2 \le \overline{C}_w(h-1)\pi_1(h,\mu), \qquad (18)$$

$$\mathrm{E}\mathcal{T}_{n3}^2 = \sum_{t=n-h+2}^{n} \mathrm{E}(\tau' w_t)^2 \le \overline{C}_w(h-1)\pi_1(h,\mu). \qquad (19)$$

In view of (17) and (18), $(\mathrm{E}\mathcal{T}_{n2}^2)^{-1}E\mathcal{T}_{n1}^2 \le \underline{C}_w^{-1}\overline{C}_w(n-2h-p+2)^{-1}(h-1) \le \underline{C}_w^{-1}\overline{C}_w(n-2\overline{h}-\overline{p}+2)^{-1}\overline{h} \to 0$, since $(\overline{h}+\overline{p})/n \to 0$. Thus by Chebyshev's inequality, $\lim_{n\to\infty}\sup_{p,h,a}\mathrm{P}((\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}\mathcal{T}_{n1} > M) = 0$, for all $M > 0$. Similarly, the result holds if $\mathcal{T}_{n1}$ is replaced by $\mathcal{T}_{n3}$.

So for any sequences $\{a = a(n)\}$, $\{h = h(n)\}$ and $\{p = p(n)\}$ described above,

$$(\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}\mathcal{T}_n = (\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}\mathcal{T}_{n2} + o_p(1) \xrightarrow{d} \mathcal{N}(0,1), \qquad (20)$$

provided that

$$(\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}\mathcal{T}_{n2} \xrightarrow{d} \mathcal{N}(0,1). \tag{21}$$

The result (21) follows from the central limit theorem for the martingale difference array $\{\tau' w_t : t \in \mathcal{S}\}$, (Davidson, 1994, theorem 24.3), if we can show that

$$\begin{aligned} \lim_{n\to\infty} \sup\nolimits_{p,h,a}(n-2h-p+2)^{-1}[\mathrm{E}(\tau' w_t)^2]^{-2}\mathrm{E}(\tau' w_t)^4 &= 0, \text{ for } t \in \mathcal{S}; & (22) \\ \lim_{n\to\infty} \sup\nolimits_{p,h,a}\mathrm{P}\left((\mathrm{E}\mathcal{T}_{n2}^2)^{-1}\textstyle\sum_{t\in\mathcal{S}}(\tau' w_t)^2 - 1 > M\right) &= 0, \text{ for all } M > 0. & (23) \end{aligned}$$

Consider (22) first. Note that

$$\begin{aligned} & (n-2h-p+2)^{-1}[\mathrm{E}(\tau' w_t)^2]^{-2}\mathrm{E}(\tau' w_t)^4 \\ \le\ & (n-2h-p+2)^{-1}|\mu|_1^2\underline{C}_w^{-2}C_{w4} \quad \text{[by (29) and Lemma MOMT-W(ii)]} \\ \to\ & 0, \ [\text{since } (\overline{h}+\overline{p}+\overline{\mu}^2)/n \to 0] \end{aligned}$$

as $n \to \infty$, thus (22) holds.

Consider (23) now. By Chebyshev's inequality, it suffices to show that

$$\lim_{n\to\infty}\sup_{p,h,a}(\mathrm{E}\mathcal{T}_{n2}^2)^{-2}\mathrm{Var}(\sum_{t\in\mathcal{S}}(\tau' w_t)^2) = 0.$$

Note that

$$(\mathrm{E}\mathcal{T}_{n2}^2)^{-2}\mathrm{Var}(\sum_{t\in\mathcal{S}}(\tau' w_t)^2) = (\mathrm{E}\mathcal{T}_{n2}^2)^{-2}\sum_{t\in\mathcal{S}}\mathrm{Var}((\tau' w_t)^2) + (\mathrm{E}\mathcal{T}_{n2}^2)^{-2}\sum_{\{t,s\}\subset\mathcal{S},t\neq s}\mathrm{Cov}((\tau' w_t)^2,(\tau' w_s)^2).$$

The first term is bounded by

$$(\mathrm{E}\mathcal{T}_{n2}^2)^{-2}\sum_{t\in\mathcal{S}}\mathrm{E}(\tau' w_t)^4 \le (n-2h-p+2)^{-1}[\mathrm{E}(\tau' w_t)^2]^{-2}\mathrm{E}(\tau' w_t)^4 \to 0$$

uniformly in p, h, a, which holds by (22). It thus remains to show

$$(\mathrm{E}\mathcal{T}_{n2}^2)^{-2}\sum_{\{t,s\}\subset\mathcal{S},t\neq s}\mathrm{Cov}((\tau' w_t)^2,(\tau' w_s)^2) \to 0, \tag{24}$$

uniformly in p, h, a. Note that, in view of the invariance of the distribution of w_t for $t \in \mathcal{S}$,

$$\begin{aligned}
&\text{LHS of (24)}\\
= \ & (\mathrm{E}\mathcal{T}_{n2}^2)^{-2} 2 \sum_{\ell=1}^{n-2h-p+1} \sum_{k=1}^{\ell} \mathrm{Cov}((\tau' w_0)^2, (\tau' w_k)^2)\\
= \ & \sum_{\ell=1}^{n-2h-p+1} \varpi_\ell \frac{(n-2h-p+2)(n-2h-p+1)}{\ell} \sum_{k=1}^{\ell} (\mathrm{E}\mathcal{T}_{n2}^2)^{-2} \mathrm{Cov}((\tau' w_0)^2, (\tau' w_k)^2),
\end{aligned}$$

where $\varpi_\ell = 2(n-2h-p+2)^{-1}(n-2h-p+1)^{-1}\ell$ and $\sum_{\ell=1}^{n-2h-p+1} \varpi_\ell = 1$. Thus by Toeplitz Lemma, it suffices to show

$$(n-2h-p+2) \sum_{k=1}^{n-2h-p+1} (\mathrm{E}\mathcal{T}_{n2}^2)^{-2} \mathrm{Cov}((\tau' w_0)^2, (\tau' w_k)^2) \to 0,$$

uniformly in p, h, a, or equivalently, in view of (17),

$$\lim_{n\to\infty} \sup_{p,h,a} (n-2h-p+2)^{-1} \pi_1(h,\mu)^{-2} \sum_{k=1}^{n-2h-p+1} \mathrm{Cov}((\tau' w_0)^2, (\tau' w_k)^2) = 0.$$

The last result holds by Lemma CW. So (24) holds, and then (23) holds. Therefore, (21) is proved.

Part II. We now show that for any sequences $\{a = a(n)\}$, $\{p = p(n)\}$ and $\{h = h(n)\}$ described above,

$$(\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2} \sum_{t=p}^{n-h} \tau' u_t \psi_{1t}(h,\mu) \xrightarrow{d} \mathcal{N}(0,1). \tag{25}$$

To show (25), given the result in Part I (i.e. (20)), we only need to show that for all $M > 0$,

$$\lim_{n\to\infty} \sup_{p,h,a} \mathrm{P}\left(\left| (\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2} \sum_{t=p}^{n-h} \tau' u_t \sum_{\ell=p}^{\infty} \gamma_{1\ell}(h,\mu)' y_{t-\ell} \right| > M \right) = 0. \tag{26}$$

(26) holds provided that, in view of (17),

$$\frac{\mathrm{Var}(\sum_{t=p}^{n-h} \tau' u_t \sum_{\ell=p}^{\infty} \gamma_{1\ell}(h,\mu)' y_{t-\ell})}{\pi_1(h,\mu)(n-2h-p+2)} \to 0, \tag{27}$$

uniformly over p, h, a. In fact, the LHS of (27) is

$$
\begin{aligned}
& \pi_1(h,\mu)^{-1}(n-2h-p+2)^{-1}\sum_{t=p}^{n-h}\mathrm{E}(\tau' u_t)^2[\sum_{\ell=p}^{\infty}\gamma_{1\ell}(h,\mu)y'_{t-\ell}]^2 \\
\leq\ & \pi_1(h,\mu)^{-1}\frac{n-h-p+1}{n-2h-p+2}|\tau|^2[\sum_{\ell=p}^{\infty}\sum_{j=p}^{\infty}|\gamma_{1\ell}(h,\mu)||\gamma_{1j}(h,\mu)|]\mathrm{E}|u_t|^2|y_{t-\ell}||y_{t-j}| \\
& \hfill \text{[by Cauchy inequality]} \\
\leq\ & \pi_1(h,\mu)^{-1}\overline{\pi}(n)\frac{n-h-p+1}{n-2h-p+2}|\tau|^2\left[\textstyle\sum_{\ell=p}^{\infty}|\gamma_{1\ell}(h,\mu)|\right]^2(\mathrm{E}|u_t|^4 K C_{y4})^{1/2} \\
& \hfill \text{[since } \mathrm{E}|y_t|^4 \leq K C_{y4}\overline{\pi}(n)^2, \text{ by Lemma MOMT-Y(i)]} \\
\leq\ & 2(\mathrm{E}|u_t|^4 K C_{y4})^{1/2}|\tau|^2\pi_1(h,\mu)^{-1}\overline{\pi}(n)(1+C_2)\pi_1(h,\mu)\left[\textstyle\sum_{j=1}^{\infty}\min\{j,h+1\}|a_{p-1+j}|\right]^2 \\
& \hfill \text{[by Lemma TAIL(i), and } n \geq 3h+p-3 \text{ as assumed]} \\
=\ & 2(\mathrm{E}|u_t|^4 K C_{y4})^{1/2}|\tau|^2(1+C_2)\cdot\underbrace{\left[\overline{\pi}(n)^{1/2}\textstyle\sum_{j=1}^{\infty}\min\{j,h+1\}|a_{p-1+j}|\right]^2}_{\to 0,\ \text{[weaker than Assumption 2(ii)]}} \\
\to\ & 0,
\end{aligned}
$$

uniformly over p, h, a, where $\overline{\pi}(n) = \max_{1\leq k\leq K}\pi_k(n)$. So (27) holds, and (25) is proved.

Part III. Now we turn to the estimator $\widehat{\beta}_1(h,\mu)$. Writing $\widehat{\Sigma} = (n-p-h+1)^{-1}\sum_{t=p}^{n-h}\widehat{u}_t(h)\widehat{u}_t(h)'$, we have

$$
\begin{aligned}
& (n-p-h+1)(\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}[\nu_1'\widehat{\beta}_1(h,\mu)-\nu_1'\beta_1(h,\mu)] \\
=\ & (\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}\nu_1'\widehat{\Sigma}^{-1}\sum_{t=p}^{n-h}\widehat{u}_t(h)\psi_{1t}(h,\mu)\ \text{[by (9)]} \\
=\ & (\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}\tau'\sum_{t=p}^{n-h}\widehat{u}_t(h)\psi_{1t}(h,\mu)+\text{s.o. [by Lemma SIG]} \\
=\ & (\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}\tau'\sum_{t=p}^{n-h}u_t\psi_{1t}(h,\mu)+o_P(1)\ \text{[by Lemma NEG]} \\
& \xrightarrow{d}\mathcal{N}(0,1),\ \text{[by (25)]}
\end{aligned}
$$

where s.o. means a smaller order term. Note that $V = (n-h-p+1)^{-2}\mathrm{E}\mathcal{T}_n^2$, so

$$\frac{(n-p-h+1)^{-2}\mathrm{E}\mathcal{T}_{n2}^2}{V} = \frac{\mathrm{E}\mathcal{T}_{n2}^2}{\mathrm{Var}(\mathcal{T}_n)} = \frac{\mathrm{E}\mathcal{T}_{n2}^2}{\mathrm{E}\mathcal{T}_{n1}^2 + \mathrm{E}\mathcal{T}_{n2}^2 + \mathrm{E}\mathcal{T}_{n3}^2} \to 1, \tag{28}$$

using (17), (18) and (19). We thus have $V^{-1/2}[\nu_1'\widehat{\beta}_1(h,\mu) - \nu_1'\beta_1(h,\mu)] \xrightarrow{d} \mathcal{N}(0,1)$ for any sequence $\{a = a(n)\}$, $\{p = p(n)\}$ and $\{h = h(n)\}$ described above. Then (10) follows from Pólya's Theorem.

Lastly, to show the bounds in (11),

$$\begin{aligned}
& \pi_1(h,\mu)^{-1}(n-h-p+1)V \\
= \; & \pi_1(h,\mu)^{-1}(n-h-p+1)^{-1}\mathrm{E}\mathcal{T}_n^2 \\
\geq \; & \pi_1(h,\mu)^{-1}(n-h-p+1)^{-1}\mathrm{E}\mathcal{T}_{n2}^2 \geq \frac{n-2h-p+2}{n-h-p+1}\underline{C}_w \quad \text{[by (17)]} \\
\geq \; & 2^{-1}\underline{C}_w. \quad \text{[since } n \geq 3h+p-3 \text{ by assumption]}
\end{aligned}$$

For the upper bound,

$$\begin{aligned}
& \pi_1(h,\mu)^{-1}(n-h-p+1)^{-1}\mathrm{E}\mathcal{T}_n^2 \\
\leq \; & (n-h-p+1)^{-1}[2\overline{C}_w(h-1) + (n-2h-p+2)\overline{C}_w] \quad \text{[by (17), (18), (19)]} \\
= \; & \frac{n-p}{n-h-p+1}\overline{C}_w \leq 2\overline{C}_w. \quad \text{[since } n \geq 3h+p-3 \geq 2h+p-2\text{]}
\end{aligned}$$

Then (11) holds with $\underline{C}_V = 2^{-1}\underline{C}_w$ and $\overline{C}_V = 2\overline{C}_w$. The proof of Theorem 1 is complete. □

The proof of Theorem 1 invokes following lemmas, for which the proofs are provided in the Supplement to the paper (Section S1).

Lemma MART. (i) Let x_t, φ_i and u_t be sequences such that their dimensions are conformable in the product $x_t\varphi_i u_t$. Let $\xi_t(h) = \sum_{i=1}^h \varphi_i u_{t+i}$. Then the following algebraic equality holds:

$$\sum_{t=m}^{M} x_{t-k}\xi_t(h) = \sum_{t=m+1}^{M+h}\Big(\sum_{i=1}^{h}\mathbb{I}_{\{m-k\leq t-k-i\leq M-k\}}x_{t-k-i}\varphi_i\Big)u_t,$$

where $k \geq 0$.

(ii) The equality (13) holds.

The following lemmas hold under the assumptions of Theorem 1.

Lemma MOMT-W. Let $\mathcal{S} = \{t : p + h \leq t \leq n - h + 1\}$ and $\mathcal{S}^c = \{p + 1, \cdots, n\} \backslash \mathcal{S}$. Let $\tau = \Sigma^{-1}\nu_1$.

(i). There exist constants $\underline{C}_w > 0$ and $\overline{C}_w > 0$, such that

$$\begin{aligned} \pi_1(h, \mu)^{-1}\mathrm{E}(\tau' w_t)^2 &\geq \underline{C}_w, \text{ for } t \in \mathcal{S} && (29) \\ \pi_1(h, \mu)^{-1}\mathrm{E}(\tau' w_t)^2 &\leq \overline{C}_w, \text{ for } t \in \mathcal{S} \cup \mathcal{S}^c. && (30) \end{aligned}$$

(ii). For $t \in \mathcal{S}$, $|\mu|_1^{-2}\pi_1(h, \mu)^{-2}\mathrm{E}(\tau' w_t)^4 \leq C_{w4}$, for some constant $C_{w4} > 0$.

Lemma CW.

$$\lim_{n\to\infty} \sup_{1\leq h\leq \bar{h}} \sup_{a\in\mathcal{A}} \left|(n - 2h - p + 2)^{-1}\pi_1(h, \mu)^{-2} \textstyle\sum_{k=1}^{n-2h-p+1} \mathrm{Cov}((\tau' w_0)^2, (\tau' w_k)^2)\right| = 0.$$

Lemma MOMT-Y. Let $\pi_k(n) = \sum_{i=0}^{n-1} |\beta_k(i)|^2$. Then for $k = 1, \cdots, K$ and $t = 1, \cdots, n$,

(i). $\pi_k(n)^{-2}\mathrm{E}y_{kt}^4 \leq C_{y4}$,

(ii). $\mathrm{E}(\Delta y_{kt})^4 \leq C_{\Delta y4}$,

where $\Delta y_t = y_t - y_{t-1}$, and $C_{y4} > 0$ and $C_{\Delta y4} > 0$ are two constants.

Lemma ARP. Write the VAR(∞) model (2) in the following form

$$y_t = \delta_1 y_{t-1} + \sum_{j=1}^{\infty} \delta_{j+1}\Delta y_{t-j} + u_t, \tag{31}$$

where $\delta_1 = \sum_{i=1}^{\infty} a_i$ and $\delta_{j+1} = -\sum_{i=j+1}^{\infty} a_i$ for $j \geq 1$. Let $\{\widehat{\delta}_j(h) : j = 1, \cdots, p-1\}$ be OLS coefficients of the VAR($p - 1$) regression (transformed as in (31)) using data indexed by $t = p, \cdots, n-h$, i.e. regression of y_t on $\{y_{t-1}, \Delta y_{t-1}, \cdots, \Delta y_{t-p+2}\}$. Denote $\delta = (\delta_1, \delta_2, \cdots, \delta_{p-1})$ and $\widehat{\delta}(h) = (\widehat{\delta}_1(h), \widehat{\delta}_2(h), \cdots, \widehat{\delta}_{p-1}(h))$. Then,

$$\lim_{M\to\infty} \lim_{n\to\infty} \sup_{p,h,a} \mathrm{P}\left(\left|p^{-1/2}[\widehat{\delta}(h) - \delta]\Upsilon_n(p-1)\right| > M\right) = 0,$$

where $\Upsilon_n(p-1)$ is defined in Assumption 4.

Lemma SIG. $\lim_{n\to\infty} \sup_{p,h,a} \mathrm{P}\left(\left|(n-p-h+1)^{-1}\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)' - \Sigma\right| > M\right) = 0$, for all $M > 0$.

Lemma TAIL. Let $\theta_{1\ell}(h,\mu)$ and $\gamma_{1\ell}(h,\mu)$, for $\ell \geq 1$, be coefficients in (4) and (8), respectively.

(i). $\sum_{\ell=p}^{\infty} |\gamma_{1\ell}(h,\mu)| \leq [(1+C_2)\pi_1(h,\mu)]^{1/2} \sum_{j=1}^{\infty} \min\{j, h+1\}|a_{p-1+j}|$.

(ii). $\sum_{j=1}^{\infty} j|\theta_{1,p-1+j}(h,\mu)| \leq C_2\pi_1(h,\mu)\sum_{j=1}^{\infty} j|a_{p-1+j}|$, where C_2 is defined in Assumption 1.

Lemma NEG. Let $\mathcal{T}_{n2} = \sum_{t=p+h}^{n-h+1} \tau' w_t$. Then for all $M > 0$,

$$\lim_{n\to\infty} \sup_{p,h,a} \mathrm{P}\left(\left|(\mathrm{E}\mathcal{T}_{n2}^2)^{-1/2}\sum_{t=p}^{n-h} \tau'[\widehat{u}_t(h) - u_t]\psi_{1t}(h,\mu)\right| > M\right) = 0.$$

Lemma SCORE.

(i). $$\lim_{M\to\infty}\lim_{n\to\infty} \sup_{p,h,a} \mathrm{P}\left(\left|(n-h-p+1)^{-1/2}\pi_1(h,\mu)^{-1}\sum_{t=p}^{n-h} \Pi(n)^{-1/2} y_{t-1}\xi_{1t}(h,\mu)\right| > M\right) = 0;$$

(ii). $$\lim_{M\to\infty}\lim_{n\to\infty} \sup_{p,h,a} \mathrm{P}\left(\left|(n-h-p+1)^{-1/2}\pi_1(h,\mu)^{-1}\sum_{t=p}^{n-h} \Delta y_{t-j+1}\xi_{1t}(h,\mu)\right| > M\right) = 0,$$

for $j \geq 2$.

Appendix B: The alternative LP estimator

In Appendix B we present a result on the alternative LP estimator $\check{\beta}(h)$ of the individual response $\beta(h)$ introduced in Section 3.2 for the single-equation model ($K = 1$).

Proposition 2. Suppose the data follow the AR(1) process, $y_t = ay_{t-1} + u_t$, where a is fixed such that $|a| < 1$, and u_t is I.I.D. with zero mean and unit variance. The alternative LP estimator $\check{\beta}(h)$ is defined as the OLS slope coefficient of y_t in the regression of $\check{y}_{t+h} = y_{t+h} - \sum_{i=1}^{h-1} \mathring{\beta}(h-i)\mathring{u}_{t+i}$ on $\{y_t, \cdots, y_{t-p+1}\}$, where $p \geq 2$. The preliminary estimates of $\beta(i)$ and u_t are such that $\mathring{\beta}(i) - \beta(i) = O_P(n^{-1/2})$ for $1 \leq i \leq h-1$, and $\mathring{u}_t$ is obtained by regressing y_t on $\{y_{t-1}, \cdots, y_{t-p_u}\}$, where $p_u \geq h$. Assume that $n^{-1}p^2p_u^2 \to 0$. Then, as

$n \to \infty$, for a given (fixed) $h \geq 2$,

$$n^{1/2}[\check{\beta}(h) - \beta(h)] \xrightarrow{d} \mathcal{N}\left(0, \sum_{i=0}^{h-1} \beta(i)^2\right).$$

Proposition 2 shows that the alternative LP estimator is asymptotically equivalent to LP if a sufficiently large number of lags is included in the regression; see (12). Proposition 2 can be compared with Lusompa (2022, proposition 6) which uses $p = p_u = 1$.

Proof of Proposition 2. As before, let $\widehat{u}_t(h)$ be the projection residual of y_t on $\{y_{t-1}, \cdots, y_{t-p+1}\}$. Then

$$\begin{aligned}
\check{\beta}(h) &= [\sum_{t=p}^{n-h} \widehat{u}_t(h)^2]^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h)\check{y}_{t+h} && \text{[FWL Theorem, since } p \geq 2] \\
&= [\sum_{t=p}^{n-h} \widehat{u}_t(h)^2]^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h)[y_{t+h} - \sum_{i=1}^{h-1} \beta(h-i)u_{t+i}] + A_2 && \text{[add and subtract]} \\
&= \beta(h) + A_1 + A_2, && \text{[OLS algebra]}
\end{aligned}$$

where $A_1 = [\sum_{t=p}^{n-h} \widehat{u}_t(h)^2]^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h) u_{t+h}$ and $A_2 = [\sum_{t=p}^{n-h} \widehat{u}_t(h)^2]^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h)[B_t - \mathring{B}_t]$, with $B_t = \sum_{i=1}^{h-1} \beta(h-i)u_{t+i}$ and $\mathring{B}_t = \sum_{i=1}^{h-1} \mathring{\beta}(h-i)\mathring{u}_{t+i}$. The proof is equivalent to showing that $n^{1/2}(A_1 + A_2) \xrightarrow{d} \mathcal{N}\left(0, \sum_{i=0}^{h-1} \beta(i)^2\right)$.

Define the lower triangular matrix G and the tridiagonal $V_{a,h}$ as

$$\underset{(h-1)\times(h-1)}{G} = \begin{pmatrix} 1 & & & \\ a & \ddots & & \\ \ddots & \ddots & \ddots & \\ a^{h-2} & \ddots & a & 1 \end{pmatrix}, \quad \underset{(h-1)\times(h-1)}{V_{a,h}} = \begin{pmatrix} 1 & -a & & \\ -a & 1+a^2 & \ddots & \\ & \ddots & \ddots & -a \\ & & -a & 1+a^2 \end{pmatrix}.$$

We will show that (in Section S5 of the on-line Supplement)

$$n^{1/2}A_1 = n^{-1/2}\sum_{t=p}^{n-h} u_t u_{t+h} + o_P(1), \tag{32}$$

$$n^{1/2}A_2 = \overline{\beta}' G n^{1/2} \begin{pmatrix} \widehat{a}_1 - a \\ \vdots \\ \widehat{a}_{h-1} - a\mathbb{I}_{\{h=2\}} \end{pmatrix} + o_P(1), \tag{33}$$

where $\overline{\beta}' = (\beta(h-1), \cdots, \beta(1))$, and $\{\widehat{a}_1, \cdots, \widehat{a}_{h-1}, \cdots, \widehat{a}_{p_u}\}$ are OLS slopes in obtaining $\mathring{u}_t$, with $p_u \geq h$. Note that only the first $h-1$ slopes $\{\widehat{a}_1, \cdots, \widehat{a}_{h-1}\}$ matter for the limit distribution of $\check{\beta}(h)$, but at least one more lag is used in the autoregression to obtain these slopes.

Proposition 2 then follows from two convergences $n^{-1/2}\sum_{t=p}^{n-h} u_t u_{t+h} \xrightarrow{d} \mathcal{N}(0,1)$ and

$$n^{1/2} \begin{pmatrix} \widehat{a}_1 - a \\ \vdots \\ \widehat{a}_{h-1} - a\mathbb{I}_{\{h=2\}} \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, V_{a,h}), \tag{34}$$

where two convergences occur jointly, and two limit normal random variables are independent. The convergence in (34) holds from the stationary AR(1) data generating process, and the key assumption $p_u \geq h$. Noting that $GV_{a,h}G' = \mathrm{I}_{h-1}$, the proof is complete. □

References

Andrews, D. W. K., X. Cheng, and P. Guggenberger. (2020). "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *Journal of Econometrics*, 218, 496-531.

Andrews, D. W. K., and P. Guggenberger. (2012). "Asymptotics for LS, GLS, and Feasible GLS Statistics in an AR(1) Model with Conditional Heteroskedasticity," *Journal of Econometrics*, 169, 196-210.

Andrews, D. W. K., and P. Guggenberger. (2014). "A Conditional-Heteroskedasticity-

Robust Confidence Interval for the Autoregressive Parameter," *Review of Economics and Statistics*, 96, 376-381.

Beveridge, S., and C. R. Nelson. (1981). "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics,* 7, 151-74.

Bhansali, R. J. (1997). "Direct Autoregressive Predictors for Multistep Prediction: Order Selection and Performance Relative to the Plug in Predictors," *Statistica Sinica*, 7, 425-449.

Carrasco, M., and X. Chen. (2002). "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18,17-39.

Chamberlain, G. (1987). "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-334.

Chang, P. L, and S. Sakata. (2007). "Estimation of Impulse Response Functions Using Long Autoregression," *Econometrics Journal*, 10, 453-469.

Colacito, R., E. Ghysels, J. Meng, and W. Siwasarit. (2016). "Skewness in Expected Macro Fundamentals and the Predictability of Equity Returns: Evidence and Theory," *Review of Financial Studies*, 29, 2069-2109.

Conrad, J., R. F. Dittmar, and E. Ghysels. (2013). "Ex ante Skewness and Expected Stock Returns," *Journal of Finance*, 68, 85-124.

Cuny, C., and F. Merlevède. (2014). "On Martingale Approximations and the Quenched Weak Invariance Principle," *Annals of Probability*, 42, 760-793.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* London: Oxford University Press.

Deo, S. R. (2000). "Spectral Tests of the Martingale Hypothesis under Conditional Heteroskedasticity," *Journal of Econometrics*, 99, 291-315.

Dufour, J.-M., and E. Renault. (1998). "Short-run and Long-run Causality in Time Series: Theory," *Econometrica*, 66, 1099–1125.

Dufour, J.-M., E. Renault, and D. Pelletier. (2006). "Short Run and Long Run Causality in Time Series: Inference," *Journal of Econometrics*, 132, 337-362.

Dufour, J.-M., E. Renault, and V. Zinde-Walsh. (2021). "Wald Tests when Restrictions are Locally Singular," Working paper, McGill University and University of Warwick.

Gonçalves, S., and L. Kilian. (2004). "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 123, 89-120.

Gonçalves, S., and L. Kilian. (2007). "Asymptotic and Bootstrap Inference for AR(inf) Processes with Conditional Heteroskedasticity," *Econometric Reviews*, 26, 609-641.

Gordin, M. I. (1969). "The Central Limit Theorem for Stationary Processes," Soviet Math. Dokl. 10, 1174-1176.

Hansen, B. E. (1994). "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705-730.

Herbst, E. P., and B. K. Johannsen. (2021). "Bias in Local Projections," Discussion Series 2020-010r1, Federal Reserve Board.

Inoue, A., and L. Kilian. (2002). "Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in Infinite-Order VAR Models," *International Economic Review*, 43, 309-332.

Inoue, A., and L. Kilian. (2020). "The Uniform Validity of Impulse Response Inference in Autoregressions," *Journal of Econometrics*, 215, 450–472.

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

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, 461-87.

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

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, 541–559.

Lewis, R., and G. Reinsel. (1985). "Prediction of Multivariate Time Series of Autoregressive Model Fitting," *Journal of Multivariate Analysis,* 16, 393-411.

Lunsford, K. G. (2020). "Policy Language and Information Effects in the Early Days of Federal Reserve Forward Guidance," *American Economic Review*, 110, 2899–2934.

Lusompa, A. (2022). "Local Projections, Autocorrelation, and Efficiency," Working paper (dated August 31, 2022), Federal Reserve Bank of Kansas City.

Lütkepohl, H. (1990). "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics*, 72, 116-125.

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

Marcellino, M., J. H. Stock, and M. W. Watson (2006). "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics*, 135, 499-526.

Mikusheva, A. (2012). "One-Dimensional Inference in Autoregressive ModelsWith the Potential Presence of a Unit Root," *Econometrica*, 80, 173-212.

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

Montiel Olea, J. L., and M. Plagborg-Møller. (2022). "Corrigendum: Local Projection Inference is Simpler and More Robust Than You Think," Online manuscript, Princeton University.

Nakamura, E., and J. Steinsson. (2018). "Identification in Macroeconomics," Journal of Economic Perspectives, 32, 59-86.

Phillips, P. C. B., and V. Solo. (1992). "Asymptotics for Linear Process," *Annals of Statistics*, 20, 971-1001.

Plagborg-Møller, M., and C. K. Wolf. (2021). "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica,* 89, 955-980.

Wu, W. B. and M. Woodroofe. (2004). "Martingale Approximations for Sums of Stationary Processes," *Annals of Probability*, 32, 1674-1690.

Xu, K.-L. (2020). "Testing for Multiple-Horizon Predictability: Direct Regression Based versus Implication Based," *Review of Financial Studies*, 33, 4403-4443.

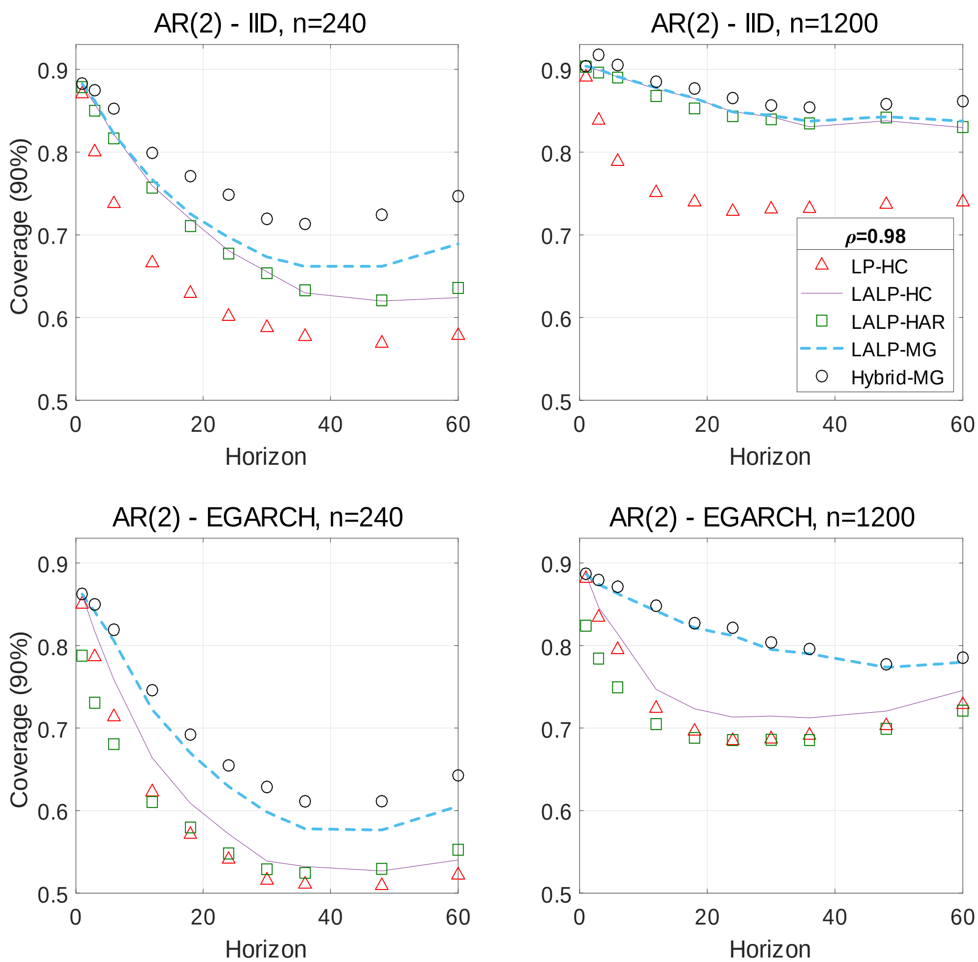

Figure 2: Coverage rates of five 90%-nominal confidence intervals for the impulse response $\beta(h)$ under DGP AR(2) with I.I.D. or EGARCH shocks. The two AR roots are 0.98 and 0.5. The sample size $n \in \{240, 1200\}$.

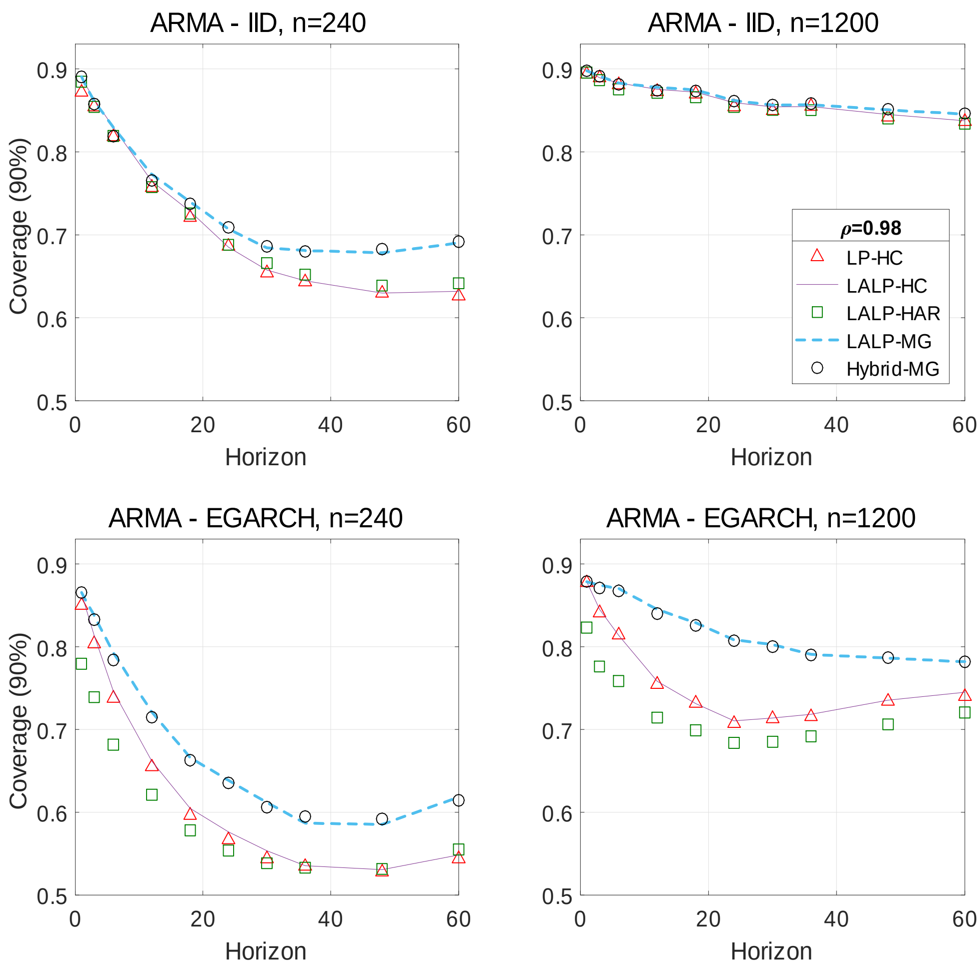

Figure 3: Coverage rates of five 90%-nominal confidence intervals for the impulse response $\beta(h)$ under DGP ARMA(1,1) with I.I.D. or EGARCH shocks. The AR and MA roots are 0.98 and 0.5, respectively. The sample size $n \in \{240, 1200\}$.