Local Projection Based Inference under General Conditions*

Ke-Li Xu†

Indiana University Bloomington

March 31, 2026

Abstract

This paper develops the uniform asymptotic theory for local projection (LP) regression when the true lag order of the model is unknown and potentially infinite. The theory allows for varying degrees of persistence in the data, growing response horizons, and general conditionally heteroskedastic martingale-difference shocks. Based on the theory, we make two main contributions. First, we show that LPs can achieve semiparametric efficiency at a given horizon under classical assumptions on the data, provided that the controlled lag order diverges. Thus the commonly perceived efficiency loss of LPs can become asymptotically negligible with many controls. Second, we propose LP-based inference procedures for (level and cumulated) impulse responses that possess robustness properties not shared by existing methods. Inference methods using two distinct standard errors are considered. The uniform validity for the first method depends on a zero fourth-order cumulant condition on shocks, while that of the second holds more generally for conditionally heteroskedastic martingale-difference

*The author is grateful to the Editor, three anonymous referees, and participants in ES world congress (Seoul), ASSA, Fintech at HK CityU, CFE, ES regional meetings, SETA, Cemmap/SNU workshop, NBER-NSF time series conference, ECMFE workshop, MEG, CESG, Mini-conference at Notre Dame, IAER-DUFE econometrics conference, and departmental seminars at LSE, Nottingham, McGill, Indiana, UC-Irvine, UIUC, McMaster, Kansas, Toronto Metropolitan, Wisconsin-Madison, Purdue, Southern Methodist, Connecticut, HKU, HKUST, Tsinghua, PKU, Remin, and CUFE, especially, Jean-Marie Dufour, Xu Han, Francesca Molinari, and Mikkel Plagborg-Møller, for insightful comments and suggestions. This paper replaces its earlier version first circulated in 2022 under the title "On Local Projection Based Inference".

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

shocks. We propose a bootstrap procedure that improves finite-sample performance and extend the standard error construction to structural responses.

Keywords: Impulse response; local projection; persistence; semiparametric efficiency; uniform inference. (*JEL Classification*: C32.)

1 Introduction

Impulse response analysis is a fundamental tool in applied macroeconomic research. In this paper, we develop inference procedures for impulse responses that are robust to model specification, parameter locations, propagation horizons, and the dependence structure of the shock process. Accordingly, the proposed methods can accommodate key stylized features in macroeconomic data such as stochastic trends, high persistence, comovement, long-range dependence, and volatility clustering.

Our approach is based on the method of local projections (LPs) (Jordà, 2005, 2023, Jordà and Taylor, 2025, Dufour and Renault, 1998). Consider K endogenous variables of interest stacked in the $K \times 1$ vector y_t. The primary interest of this paper is in the responses of the first component, y_{1t}, to all forecast errors after h propagation horizons, where $h \geq 1$. These responses are collected in the $K \times 1$ vector $\beta_1(h)$. The LP method runs the following h-period-lead regression

$$y_{1,t+h} = \varrho_1(h) + \beta_1(h)' y_t + \sum_{\ell=1}^{p-1} \theta_{1\ell}(h)' y_{t-\ell} + \eta_{1t}(h), \tag{1}$$

for a given integer $p \geq 1$, where $\eta_{1t}(h)$ is the regression error and $\varrho_1(h)$ is the intercept. Since we do not assume the shocks to be observed, a sufficiently large number of controlled lags $\{y_{t-\ell} : 1 \leq \ell \leq p-1\}$ should be included to ensure proper identification of $\beta_1(h)$.[1] The reduced-form impulse responses $\beta_1(h)$ are fundamental parameters for constructing structural responses to various economic shocks, typically identified through exclusion restrictions and represented as rotations of the reduced-form responses.

The robust inference we propose is based on a newly developed standard error for the ordinary least squares (OLS) estimator $\widehat{\beta}_1(h)$ from the LP regression (1). The new standard error is constructed by first partialling out control variables and then estimating the variance

[1]The regression (1) makes immediate sense if the data are known to follow the VAR(p) process. Then iterating forward over h periods and focusing on the first equation in the system yields (1). We emphasize that our analysis below does not require such knowledge, so the regression (1) may be misspecified for a given p.

of the martingale-transformed effective regression score. The estimated score variance is simply a sum of squares, and it can be justified even when original score contributions (with or without partialling out) are serially correlated (for $h \geq 2$) in the presence of general conditional heteroskedasticity. In the paper we establish the uniform validity of the proposed inference procedure, which is robust to the aforementioned data features, under the vector autoregression (VAR) model with an unknown and potentially infinite lag order. We extend the standard error construction to recursively identified structural responses and discuss special cases in which their inference does not require long-run variance estimation.

Robust inference of impulse responses potentially generated by a VAR(∞) process is novel. The existing literature —whether based on LP or other estimators—that accommodates 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) has primarily focused on stationary settings and fixed horizons. Allowing the true lag order of the VAR model to remain unrestricted has been emphasized as fundamentally important in the modern empirical macroeconomics 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 relatively large lag order is often recommended; see Kilian and Lütkepohl (2017) and Montiel Olea and Plagborg-Møller (2021). When the true DGP is VAR(∞) (e.g. nondegenerate vector autoregressive and moving average model (VARMA)), the model order is typically required—at least theoretically—to diverge at an appropriate rate for satisfactory approximation quality. While we offer no formal theory for optimally choosing the lag order when it grows with the sample size, our analysis suggests an important implication. When the model order diverges, the LP estimator of the impulse response at a given horizon can attain semiparametric efficiency under classical assumptions on the DGP (stationarity and homoskedastic shocks). Our result extends Plagborg-Møller and Wolf (2021), who show that iterative VAR-implied and LP estimators share the same population estimand and probability limit under a VAR(∞) model, by establishing the equivalence of their asymptotic distributions. The equivalence result motivates the use of LP for inference, as the efficiency loss relative to the VAR estimator is asymptotically negligible, contrary to the conventional view that LP is inefficient under a finite-order VAR specification.[2]

This paper's contributions are summarized as follows:

1. We derive the asymptotic Gaussian theory for the LP estimator of $\beta_1(h)$ in the re-

[2] As Breitung and Brüggemann (2023, p.1321) note, "$\cdots$, despite being potentially less efficient than iterated response estimators, local projections are nowadays a popular tool in empirical economics".

gression (1), which is valid uniformly over the parameter space of the VAR(∞) model that allows the data to be weakly dependent, contain unit roots, and exhibit cointegration of unknown form. The uniform theory also accommodates general conditional heteroskedasticity and growing forecast horizons.

2. We propose a new standard error for the LP estimator, on the basis of which we construct uniform inference procedures for $\beta_1(h)$.

3. We prove that the LP estimator is asymptotically as efficient as the VAR estimator for the impulse response at a given horizon if the lag order diverges, and the data are stationary with homoskedastic martingale-difference innovations, regardless of whether the data are generated by a finite-order VAR or VAR(∞) model.

4. We provide a bootstrap implementation that further enhances finite-sample performance of existing and new inference methods, along with empirically calibrated simulations supporting the theoretical results.

5. We extend the proposed standard error and inference procedures to other types of impulse responses, including cumulative responses and structural responses identified through short-run zero restrictions.

Recent literature. Montiel Olea and Plagborg-Møller (2021; hereafter MOPM)—which clearly motivates the current study—focuses on persistence-robust inference for local projection regression under the *finite-order* VAR model. Another fundamental assumption underlying their uniformity results concerns conditionally mean-independent shocks. As the authors note, "... 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). We relax this assumption. The present paper therefore contributes to the line of work in several ways: it allows for more general data generating processes (with potentially infinite lags), more flexible control lag orders (which may diverge) in the LP regression, a broader parameter space for uniform inference (accommodating unknown cointegration in the system), general conditionally heteroskedastic martingale-difference shocks, and a new standard error construction.

Our paper complements the literature on *persistence-robust* VAR-implied inference for impulse responses (Mikusheva, 2012, Inoue and Kilian, 2020, and references therein). These methods, typically developed for finite-order univariate AR models, infer impulse responses as functions of estimated model parameters using the delta method or likelihood-based inference. Our approach offers several advantages over traditional model-implied methods: it delivers uniform validity over a broader class of models and parameter spaces, accommodates growing forecast horizons in a general manner, allows for more general shock processes, and remains readily extendable—both theoretically and computationally—to multivariate settings.

For example, Mikusheva (2012) studies a univariate model and extends it to VARs but allows a unit root in only one specific endogenous variable. Her uniformity results permit heteroskedastic shocks only under local-to-unity persistence. Inoue and Kilian (2020), under a finite-order univariate AR with independent innovations, impose a rank condition (their assumption B) that restricts the parameter space (see also Dufour, et al., 2025), which we do not. Their analysis mainly covers fixed horizons, allowing growing horizons only in the local-to-unity case.[3]

In the LP context, Lusompa (2022) proposes an alternative estimator that exploits serial correlation in the LP regression error, improving efficiency under a finite-order model (see Section 2 and on-line Appendix D for why this gain does not extend to infinite-order settings). Importantly, his inference relies on substantially stronger assumptions—stationarity, fixed horizons, and a more restrictive lag-order growth rate—so its uniform validity under the broader conditions considered here remains unclear.

In a more recent paper than the current one, Montiel Olea, et al. (2024) take a different approach to dynamic misspecification. In a stationary, homoskedastic, and fixed-horizon setting, they identify the degree of misspecification for which LP inference—ignoring the misspecification—remains asymptotically valid, whereas the same type of VAR inference does not, thereby further highlighting the robustness of local projections. Unlike the present paper, they do not consider a general VAR(∞) model and rely on the MOPM method for inference.

Recent research on LP has been particularly active. Li, et al. (2024) conduct extensive simulations comparing LP and other impulse response estimators in settings with relatively small lag orders. Dufour and Wang (2024) consider the inference of all coefficients in the LP regression (1)—which they term as generalized impulse responses—not only $\beta_1(h)$. Methods addressing a large number of control variables in LP regressions are studied by Adamek, et al. (2024), Cha (2024), and Dinh et al. (2024). Nonlinearity and its implications on linear VAR models and linear methods such local projections have been recently examined by Gonçalves, et al. (2024), Inoue, et al. (2024), and Kolesár and Plagborg-Møller (2024). The Bayesian formulation of the LP framework has been analyzed by Ferreira, et al. (2025)

[3]Earlier studies on inference of impulse responses based on the VAR recursive estimator for (nearly) integrated time series include Phillips (1998), Wright (2000), Gospodinov (2004), Pesavento and Rossi (2007), among many others.

and Huber, et al. (2025).

Organization of the paper. We first provide a practical summary of our main results in Section 2 and present finite sample evidence for asymptotic and bootstrap inferences in Section 3. The formal asymptotic arguments, assumptions and general framework are detailed in Section 4. We conclude in Section 5 and discuss possible extensions for future research. Technical details, proofs, and additional results are provided in the on-line Supplement to the paper (Appendices A-G, S1-S3).

2 Summary of main results

A motivating AR(1) example and martingale score. Our analysis is based on a key tool of representing the LP regression score as a martingale process. To motivate this idea, consider a simple case where the scalar time series y_t follows a stationary AR(1), $y_t = ay_{t-1} + u_t$, with u_t a martingale difference sequence (MDS) with respect to the natural filtration (recording the past) and $|a| < 1$. The LP regression at horizon $h = 2$ is

$$y_{t+2} = \beta(2)y_t + \xi_t(2), \tag{2}$$

where $\xi_t(2)$ is the regression error. We omit the intercept for illustrative purposes. In this simple model, $\beta(0) = 1$, $\beta(1) = a$ and $\beta(2) = a^2$. Using one iteration, we can see that the error satisfies $\xi_t(2) = \beta(1)u_{t+1} + u_{t+2}$. Given the data $\{y_t,\ 1 \leq t \leq n\}$, the OLS of (2) gives the LP estimator $\widehat{\beta}(2)$ which takes the form $\widehat{\beta}(2) = (\sum_{t=1}^{n-2} y_t^2)^{-1} \sum_{t=1}^{n-2} y_t y_{t+2} = \beta(2) + (\sum_{t=1}^{n-2} y_t^2)^{-1} \sum_{t=1}^{n-2} y_t \xi_t(2)$. The main complication in inference is to estimate the variance of the LP score $\sum_{t=1}^{n-2} y_t \xi_t(2)$, which is usually done in practice using a generic method, e.g. a Newey-West-type long-run variance estimator for a chosen kernel function and truncation parameter. We propose an alternative method, which adapts to the LP setting. The method is based on the following representation of the LP score,

$$\begin{aligned} & \sum_{t=1}^{n-2} y_t[\beta(1)u_{t+1} + u_{t+2}] \\ = \quad & \beta(1)y_1u_2 + \sum_{t=3}^{n-1} \Big[\beta(1)y_{t-1} + y_{t-2}\Big] u_t + y_{n-2}u_n \triangleq \sum_{t=2}^{n} w_t, \end{aligned} \tag{3}$$

where $w_2 = \beta(1)y_1u_2$, $w_t = \Big[\beta(1)y_{t-1} + y_{t-2}\Big] u_t$, for $3 \leq t \leq n-1$, and $w_n = y_{n-2}u_n$. The representation (3) is obtained by re-assembling the sum via collecting terms according to u_t, instead of collecting terms according to y_t in the original sum. In its original form, the score is a sum of $n-2$ terms $y_t\xi_t(2)$ which are serially correlated, and after re-arranging the sum, the score becomes a sum of $n-1$ serially uncorrelated terms (which we call the transformed score contributions w_t). In fact, w_t is an MDS, by the MDS property of u_t, so the score now written as $\sum_{t=2}^{n} w_t$ is a martingale.

The advantage of expressing the LP regression score in martingale form is that the variance of the score can now be written as a sum of variances of martingale difference terms, which is much simpler to analyze. This representation forms the foundation of our asymptotic analysis and the construction of the standard errors throughout the paper.

A realistic setting. For a vector time series y_t, the LP regression (1) at horizon h in general includes control lags and an intercept, and the exact lag order of the VAR data-generating process (and therefore of the LP regression) may be unknown. To estimate $\beta_1(h)$, the impulse response vector of $y_{1,t+h}$, the Frisch-Waugh-Lovell theorem allows us to focus on the effective regressor, denoted by $\widehat{u}_t(h)$, the partialled-out residual in the regression (1)—that is, the residual from regressing the focal regressor y_t on controlled lags $y_{t-1}, \cdots, y_{t-p+1}$ and a constant. The OLS estimator can then be written as $\widehat{\beta}_1(h) = [\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)']^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h) y_{1,t+h}$. Let R_t (population linear projection residual) be the population analog of $\widehat{u}_t(h)$. In this general setting, what contributes to the estimator variance is the *effective* score—the score relevant only to the impulse response parameter $\beta_1(h)$ and unaffected by the misspecification bias (assumed to be of smaller order). The effective score is given by $\sum_{t=p}^{n-h} R_t\xi_{1t}(h)$, where $\xi_{1t}(h) = \sum_{i=1}^{h} \beta_1(h-i)' u_{t+i}$.

Extending the algebra in (3), we rewrite the effective score as follows:

$$\sum_{t=p}^{n-h} R_t \xi_{1t}(h) = \sum_{t=p+1}^{n} w_t, \tag{4}$$

where

$$w_t = \left[\sum_{i=1}^{h} \mathbb{I}_{\{p\le t-i\le n-h\}} R_{t-i}\beta_1(h-i)'\right] u_t.$$

Note that after this rearrangement, each summand w_t depends only on the current (u_t) and past information (R_{t-i}, $1 \le i \le h$, with coefficients being earlier impulse responses $\beta_1(0), \cdots, \beta_1(h-1)$); the original score contributions $R_t\xi_{1t}(h)$, by contrast, *also* involve future innovations. The indicator functions[4] $\mathbb{I}_{\{\cdot\}}$ in w_t makes the expression slightly untidy, but they are necessary for the equality (4) to hold exactly (everywhere)—no approximation

[4]In definition, $\mathbb{I}_{\{event\}} = 1$ if the event is true, and zero otherwise.

of any sort is needed. These indicators reflect that, except for the $h-1$ earliest and $h-1$ latest t indices, w_t is sum of h terms. See (3) for the case $h=2$. Because R_t depends only on $\{u_s, s \leq t\}$ (the current and past innovations), and u_t is an MDS, each w_t is again an MDS. Hence (4) gives a martingale representation of the (effective) score for the general LP regression (1). The representation (4) is the fundamental building block of our asymptotic analysis; see Remark 5 for its relation to the martingale approximation techniques in the classical time-series literature.

In this realistic setting, since we do not assume the shocks to be observed, the inclusion of controlled lags is crucial for identification. If the true DGP has p_{true} lags, in general $p_{\text{true}}-1$ controlled lags are required in the LP regression (1) for the leading slope coefficient $\beta_1(h)$ to have the causal response interpretation. To allow for the possibility that p_{true} is very large (potentially infinite, as in VARMA models), we rely on sieve-type asymptotics (large$-p$) to control the misspecification bias.[5] We emphasize that we allow the true DGP to be a VAR process of infinite order, with the finite-order VAR treated as a special case.

An additional profit of owning controlled lags (i.e. $p \geq 2$) in regression (1) is that the residualized regressor R_t becomes substantially less persistent than the original (focal) regressor y_t. If y_t is integrated of order one, R_t becomes integrated of order zero (i.e. weakly dependent). This property opens the door for persistence-robust inference. If controlled lags are sufficiently many (p is large enough, more on this below), R_t recovers the VAR innovation, i.e. $R_t = u_t$.

We extend the analysis above for level impulse responses to cumulative responses, $\sum_{j=1}^{h} \beta_1(j)$, for which the LP regression (1) must be modified by replacing the dependent variable by the cumulative outcome. After this modification, the formula for the martingale representation (4) of the LP regression score continues to hold, provided that we replace the weight $\beta_1(h-i)$ with its cumulative counterpart, $\sum_{j=i}^{h} \beta_1(j-i)$, for $1 \leq i \leq h$.

Standard errors. The martingale representation (4) simplifies the calculation of the (asymptotic) variance, and also motivates a new variance matrix estimator for the OLS estimator $\widehat{\beta}_1(h)$. Although our construction allows for joint inference across shocks and response variables (see (21) for the general form of the construction), for simplicity we focus here on a scalar response $\nu_1'\beta_1(h)$, where ν_1 is a known, nonzero $K \times 1$ vector. For a sample

[5] To reflect the potential misspecification, in our notation we use $\xi_{1t}(h)$ to denote the ideal LP regression error under correct specification. In the actual LP regression (1), the error $\eta_{1t}(h)$ equals $\xi_{1t}(h)$ plus the potential misspecification error (for given p). In the case $p \geq p_{\text{true}}$, we have $\eta_{1t}(h) = \xi_{1t}(h)$.

of observations of y_t, $1 \le t \le n$, the new variance estimator for $\nu_1' \widehat{\beta}_1(h)$ takes the form

$$\widehat{V} = \nu_1' \left[\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)'\right]^{-1} \left(\sum_{t=p+1}^{n} \widehat{w}_t \widehat{w}_t'\right) \left[\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)'\right]^{-1} \nu_1. \tag{5}$$

In this expression, $\widehat{w}_t = [\sum_{i=1}^{h} \mathbb{I}_{\{p \le t-i \le n-h\}} \widehat{u}_{t-i}(h) \widetilde{\beta}_1(h-i)'] \widetilde{u}_t$, where $\{\widetilde{\beta}_1(i),\, 0 \le i \le h-1\}$ trace out impulse responses at earlier horizons (with $\widetilde{\beta}_1(0)$ being the first column of the identity matrix I_K), and $\widetilde{u}_t$ can be the horizon-one LP regression residual.[6] For the choice $\widetilde{\beta}_1(i)$, one can use the VAR estimator (or the convenient LP estimator $\widehat{\beta}_1(i)$); see Assumption 7 for the general conditions that $\widetilde{\beta}_1(i)$ and $\widetilde{u}_t$ must satisfy for the consistency of $\widehat{V}$. Because of its construction, $\widehat{V}$ is referred to as the MG (martingale) variance estimator.

We also consider the state-of-the-art Eicker-Huber-White sandwich HC variance estimator (popularized by MOPM) for regression (1), specialized for $\nu_1' \widehat{\beta}_1(h)$:

$$\widehat{V}_{HC} = \nu_1' \left[\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)'\right]^{-1} \left[\sum_{t=p}^{n-h} \widehat{\eta}_{1t}(h)^2 \widehat{u}_t(h)\widehat{u}_t(h)'\right] \left[\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)'\right]^{-1} \nu_1, \tag{6}$$

where $\widehat{\eta}_{1t}(h)$ is the OLS residual from (1).[7] We emphasize that $\widehat{V}_{HC}$ is implemented exactly as in MOPM (except that the intercept is included in the regression), although we use slightly different notation, $\widehat{\eta}_{1t}(h)$, to indicate that the regression may be misspecified. The variance estimators $\widehat{V}_{HC}$ and $\widehat{V}$ differ only in the middle component, which reflects how the score variance is estimated. The validity of $\widehat{V}_{HC}$ at multiple horizons ($h \ge 2$) is nontrivial, since it ignores serial correlation in the regression error; see (2) for example. However, with sufficiently many controlled lags (none are present in (2)) and under suitable restrictions on higher-order dependence in the innovations, $\widehat{V}_{HC}$ can consistently estimate the variance of $\nu_1' \widehat{\beta}_1(h)$. Our martingale-based treatment of the LP score extends MOPM's analysis to a broader and more general framework.

Both the MG and HC variance estimators, $\widehat{V}$ and $\widehat{V}_{HC}$, extend naturally to the LP *cumulative* response estimator after appropriate modification; see Section 4 for details.

Theorems 2 and 3 in Section 4 present the conditions that ensure the uniform asymptotic normality of the t-statistics (denoted as $\widehat{S}_{HC}$ and $\widehat{S}$) based on the variance estimators $\widehat{V}_{HC}$ and $\widehat{V}$, respectively. These results provide formal justification for constructing confidence intervals, with implementation details discussed later in this section. In what follows, we

[6] The difference between $\widehat{u}_t(1)$ and $\widetilde{u}_t$ is that, $\widehat{u}_t(1)$ is the VAR$(p-1)$ regression residual of y_t, while $\widetilde{u}_t$ is the VAR(p) regression residual, both using the regression data $\{y_t : t = p+1, \cdots, n\}$.

[7] See applications of $\widehat{V}_{HC}$ in recent studies by Metcalf and Stock (2020), Obstfeld and Zhou (2022), Känzig (2023), Cloyne, Dimsdale and Hürtgen (2025), and Ringo (2025), among others.

first discuss the implications of our theory for empirical applications, and then examine the implications of using different constructions of the standard error.

Robustness. Developed under general conditions, our asymptotic theory ensures valid and robust inference across diverse settings. Robustness in our framework is fourfold.

First, given a VAR model, the inference is robust to the locations of the true VAR slopes (which may be infinitely many) in the parameter space, as defined in Section 4.1. In particular, as long as the integration order is no greater than one, the framework allows any number (between 0 and K) of integrated series in the system y_t, as well as any number of unit roots (which may be fewer than the number of integrated series if cointegration exists). The key point is that, with controlled lags, the residualized regressors are stationary even when the data exhibit unit roots. Because our inference results hold uniformly over the closed parameter space, they also accommodate arbitrarily small neighborhoods of specific points—typically modeled as drifting sequences—to capture near-unit-root or near-cointegration behavior. Our approach allows researchers to work directly with macroeconomic variables in levels, contrasting with the indirect approach of differencing (commonly used for near-unit-root processes). The latter approach can lead to efficiency loss when unit roots are absent and also rules out cointegration when the differenced data are modeled as a VAR process.

Our framework encompasses the conventional stationary case as a special instance. Beyond persistence robustness, our inference procedure does not rely on parameter values satisfying the rank condition commonly imposed in VAR-based inference to ensure nonsingularity of the variance matrix (see, for example, Inoue and Kilian, 2020; Dufour et al., 2025).

Second, the inference is robust to higher-order dependence of the shocks[8]—particularly to second-order dependence, such as conditional heteroskedasticity or volatility clustering—as long as certain regularity (mixing) conditions hold, especially when the MG variance estimator is employed.[9] The intuition here is similar to that of the sandwich-form variance estimator of White (1980), irrespective of horizons, thanks to the martingale re-arranged LP score. As shown later, higher-order temporal dependence can invalidate simple LP inference at multiple horizons.

Third, our inference is robust to diverging horizons—a property generally difficult to achieve for persistence-robust VAR inference but feasible for linear estimators like LP. We therefore extend MOPM's arguments to this broader setting. However, LP estimators can

[8] Shocks are assumed to be serially uncorrelated, thus not dependent in the first order.

[9] Conditional heteroskedasticity has been recognized as a key stylized fact in macroeconomic data since, at least, Engle's (1982) ARCH model, and has recently been exploited for identifying structural VARs (Lewis, 2021).

still exhibit substantial finite-sample bias at medium and long horizons, especially with persistent data, motivating our novel bias-aware bootstrap method within the LP framework.

Fourth, our inference is robust to the model order of the true DGP—whether finite or infinite. This robustness requires the lag order in action (i.e. p) to grow with the sample size, implying that in practice, a relatively large number of controlled lags should be included. Embedding the LP regression in a large$-p$ asymptotic framework, we show that as the sample size increases, the LP estimator recovers the true response and the resulting confidence intervals achieve correct coverage. We elaborate on this below.

The large$-p$ framework. The large$-p$ asymptotics, a key component of our misspecification robust framework, is relevant for applications where researchers choose a conservative lag order (large p) either for specification safety or to allow sufficient time for economic variables to respond.[10] Lag orders chosen in this way are often substantially larger than those selected by data-driven criteria such as Akaike information criterion (AIC).[11] Nakamura and Steinsson (2018, p.81) highlighted the difficulty of approximating a macro system with finite-order VAR models, motivating the use of a large lag order. As mentioned earlier, including a sufficient number of lags is essential to achieve identification—or "invertibility," as commonly referred to in the structural VAR literature.

Moreover, it is common for empirical researchers to examine a range of lag orders for sensitivity checks. Although a formal procedure to determine the unique "optimal" lag order (in a proper sense) in a robustness setting like ours is still lacking in the large$-p$ framework, our theory provides guidance on a rough but informative range of lag orders that practitioners can explore for robust inference. In principle, the lag order should be large enough to ensure that potential misspecification bias is negligible, but not so large that model variance becomes unmanageable. In contrast, the finite$-p$ framework is silent on the maximum lag order suitable for sensible inference.

The large$-p$ framework aligns naturally with the lag-augmentation idea for the LP in the finite$-p$ framework, advocated by MOPM (see also Breitung and Brüggemann, 2023).[12] In the finite$-p$ framework, an extra lag is used for the partialling-out step to recover the

[10] A large lag order has been recommended for VAR method (Kilian and Lütkepohl, 2017, chapter 2.6, pages 63-65; see also Kilian, 1998, Inoue and Kilian, 2002), and this recommendation has been partly inherited in the LP literature. For example, Romer and Romer (2004) used 24 lags to control for the usual dynamics of macro variables. Kilian used $p = 24$ in his trivariate VAR setting. Känzig (2021, appendix A.2) use 12, 18, 24 lags in his LP specification.

[11] For Kilian (2009) data, various information criteria select a VAR lag order of 2 or 3.

[12] Under finite-order VAR model, Dufour et al. (2006) also discussed the robustness of the lag-augmented regression to persistent data and suggested using a Newey-West type standard error.

shock of interest, whereas in the large$-p$ framework, a similar goal is achieved by increasing the number of controlled lags asymptotically to capture all relevant dynamics.

On the other hand, the large$-p$ framework somewhat mitigates the tension between adjacent lag orders, once p is large, and instead focuses on a range. Mathematically, this is justified by the heuristics that $p-1 \to \infty$ ($p-1$ is the number of lags in the partialling-out step) if $p \to \infty$. Once $p \to \infty$, the requirements (in the form of upper and lower bounds) for the increasing rate for p (without lag augmentation), as a function of the sample size, is equivalent to the same conditions applied to $p+1$ (with lag augmentation). For a sufficiently large p, we can reasonably expect the last few LP regression coefficients to be rather small, so estimation and inference results for the leading (lag-zero) coefficient $\beta_1(h)$, based on adding an extra *remote* lag or not, are likely to be very similar.

To illustrate the diminishing returns of lag augmentation, in the simulations reported in Section 3 where the DGP is known to be VAR(p_{true}), lag augmentation can dramatically affect inference based on the HC standard error when $p_{\text{true}} = 1$, but the effect diminishes substantially when $p_{\text{true}} \geq 2$.

Pros and cons of different standard errors. Different structures of $\widehat{V}$ and $\widehat{V}_{HC}$ in (5) and (6) lead to important differences in inference construction. While remarkably simple, the variance estimator $\widehat{V}_{HC}$ is consistent under the crucial condition that score contributions $R_t\xi_{1t}(h)$ are uncorrelated across t. This condition is typically justified by first recovering the VAR innovation u_t through R_t, and then imposing restrictions on u_t to rule out serial correlation in the score contributions.

This justification process for $\widehat{V}_{HC}$ leads to the following advantages of its alternative, $\widehat{V}$. First, while the uncorrelatedness of score contributions for $h = 1$ is implied by the MDS property of u_t, it is not an innocuous assumption for $h \geq 2$. In fact, it typically imposes restrictions on the higher-order dependence of u_t (e.g. via the zero fourth-order cumulant condition in Assumption 6), which in turn rules out certain forms of conditional heteroskedasticity; see Section 4.3 for examples. One way to impose these restrictions is through the convenient assumption of "conditional mean independence", $\mathrm{E}(u_t|u_s, s \neq t) = 0$, almost surely, as used by MOPM. In contrast, the validity of $\widehat{V}$ does not require such restrictions on higher-order dependence of the innovations and is therefore more robust, for example, to general conditional heteroskedasticity.

Second, $\widehat{V}$-based inference is in general more robust to lag choices. Under a finite-order VAR(p) model with known order, $\widehat{V}$ has two additional advantages compared with $\widehat{V}_{HC}$. Lag augmentation for the LP regression is not required for persistence-robust inference based on $\widehat{V}$ once $p \geq 2$ (see Remark 8). Moreover, if stationarity is additionally imposed, lag augmentation is unnecessary for valid $\widehat{V}$-based inference for all $p \geq 1$. In simulations, we find that $\widehat{V}$-based inference is more robust to a range of lag choices, including large ones.

Despite these differences, both variance estimators $\widehat{V}_{HC}$ and $\widehat{V}$—each estimating the long-run variance matrix as a sum of squares (or outer products)—are straightforward to compute. This simplicity makes Newey-West type variance estimators unnecessary for inference on reduced-form impulse responses (and certain types of structural responses), at least in our settings (especially when $\widehat{V}$ is used).

Implementation.

ALGORITHM 1. (Asymptotic confidence interval for $\nu_1'\beta_1(h)$)

1. For a given lag order p (see the following paragraph for recommendations) and horizon h, run regression (1) to obtain the OLS estimator $\widehat{\beta}_1(h)$.

2. Compute the variance estimator $\widehat{V}$, using formula (5). One can also compute $\widehat{V}_{HC}$, using either the standard regression package output or equivalently formula (6).

3. The $100(1-\tau)\%$-level confidence interval (CI) is given by $\nu_1'\widehat{\beta}_1(h) \pm \widehat{V}^{1/2}\mathfrak{q}(1-\tau/2)$, where $\mathfrak{q}(1-\tau/2)$ is the $(1-\tau/2)$-th quantile of the standard normal distribution. Alternatively, one can use the standard error $\widehat{V}_{HC}^{1/2}$ in place of $\widehat{V}^{1/2}$.

For the lag order choice p, while a formal theory is yet to be developed, we draw insight from the VAR literature, which focuses on impulse responses and their inference rather than identifying the true model; Kilian and Lütkepohl (2017, chapter 2.6, pages 63-65). A practical method is to consider a range of lags chosen ex ante based on the economic context (and possibly the number of macro variables). A sufficiently long lag should be included to capture delayed responses—for instance, a starting point can be 12 for monthly data and 4 for quarterly data (one year worth of lagged values; Nakamura and Steinsson, 2018, p. 80), and potentially up to twice these numbers. Consistent inference results across a range of lags are especially appealing.

It is also common to use the same lag order across all horizons considered (Jordà and Taylor, 2025, p.65) for ease of interpretation and for the support provided by the identification analysis.[13] This constant lag choice across horizons in fact provides a robust argument for using the LP: if the true horizon-one LP involves p_{true} lags, then horizon-h LP requires at most

[13] It might be tempting to allow different lag orders across horizons. However, if an information criterion is applied separately to the LP regression for each horizon h, the procedure is likely to select a lag order that is too small for large h (see Brugnolini, 2018, for simulation evidence), possibly due to small remote LP coefficients and large residual variance. This may lead to misspecification bias and low CI coverage.

p_{true} lags for identification of impulse responses.[14] The information-criterion-based method (in particular, AIC) applied to the horizon-one LP regression can be used as a guidance, e.g. to adjust the range. As a side note, using a single lag order chosen in a data-dependent way (e.g. via certain information criterion) requires caution. The order such selected is often too small to allow the response delay. Also, the post-selection inference (e.g. Leeb and Pötscher, 2005) remains uninvestigated in the LP setting. Our asymptotic theory provides justification for a range of lag orders. Although it is difficult to pin down exact lower and upper bounds, the theory shows that there is no asymptotic cost to using a relatively large lag order (within the range).

Semiparametric efficiency of local projections. An important implication of our theory is that when p diverges (at an appropriate rate), the LP estimator is as efficient as the conventional VAR iterative estimator of the impulse response at a given horizon under classical assumptions of stationarity (as in Plagborg-Møller and Wolf, 2021) and homoskedastic MDS shocks ($\mathrm{E}(u_t u_t'|u_s, s \leq t-1) = \Sigma$, a.s.). Under these assumptions, our theory (Theorem 1 in Section 4) implies

$$n^{1/2} \left(\nu_1'[\textstyle\sum_{i=0}^{h-1} \beta_1(i)'\Sigma\beta_1(i)]\Sigma^{-1}\nu_1\right)^{-1/2} [\nu_1'\widehat{\beta}_1(h) - \nu_1'\beta_1(h)] \xrightarrow{d} \mathcal{N}(0,1). \tag{7}$$

For a scalar series, $n^{1/2}[\sum_{i=0}^{h-1} \beta(i)^2]^{-1/2}[\widehat{\beta}(h) - \beta(h)] \xrightarrow{d} \mathcal{N}(0,1)$. In fact, this asymptotic distribution is shared by the VAR iterative estimator for given h under the same set of classical assumptions as p diverges, by a slight extension of the argument in Lütkepohl (1990) (see also Lütkepohl, 2005, eqn. (15.4.1)).[15]

Thus LP possesses the same optimality properties as the VAR-implied estimator. In particular, under the classical assumptions mentioned above, LP (VAR, as well) reaches the asymptotic efficiency bound under the semiparametric conditional moment condition model $\mathrm{E}(y_t - \sum_{j=1}^{\infty} a_j y_{t-j}|y_{t-s}, s \geq 1) = 0$. Such bounds are characterized in Chamberlain (1987) and extended here through a sieve approximation. If one further assumes Gaussianity, the

[14] In other words, the LP focuses on the responses of a particular macro variable and therefore does not require the entire VAR system to be correctly specified. For this reason, and given the robustness results recently established by Montiel Olea et al. (2024), we do not recommend using an excessively long lag order for the LP—it should be bounded by the lag length a researcher would use for the VAR, if chosen conservatively.

[15] We offer two comments here. First, the sieve asympotics for the VAR estimator rely on the assumption $h \leq p$, as noted in Lütkepohl (1990). See Ballarin (2021) for a recent cautionary note. In contrast, this assumption is not needed for the sieve asympotics of the LP estimator. Second, the asymptotic theory we obtain for the LP estimator (Theorem 1) holds under much weaker assumptions than those stated here for the efficiency claim.

LP is asymptotically Cramér-Rao efficient.

The optimality of LP contrasts with the well-known result that the direct regression is less efficient than the iterative (VAR-implied) estimator under a *finite-order* VAR model; see e.g. Bhansali (1997), Schorfheide (2005), Marcellino, et al. (2006), Xu (2020a), among others, in slightly different contexts. Based on the argument that the VAR estimator is more efficient than LP under a finite-order VAR model with homoskedastic errors, Stock and Watson (2018) proposed a Hausman test for invertibility. The intuition behind this efficiency gain is analogous to that of a dimension-reduction factor model; all impulse responses are functions of a relatively small number of common parameters (VAR slopes). Imposing such functional relation—if correctly specified—yields an efficient estimator. However, the efficiency gain generated from such extraction through a parsimonious model dissipates as the model dimension increases, ultimately leading to our efficiency equivalence result.

Focusing on identification and first-order sample equivalence (i.e. the difference between LP and VAR estimators converging to zero), Plagborg-Møller and Wolf (2021, section 2.5) conjectured that LP and VAR could be equally efficient but did not provide formal analysis.[16]

In on-line Appendix E, we report some simulation results comparing the finite-sample behavior for LP and VAR estimators across a range of lag order choices. The design is based on an empirically fitted trivariate VARMA model (one of the designs used in Section 3). While LP and VAR estimators exhibit markedly different mean square errors (MSEs) for small lag orders—with either estimator potentially dominating depending on the response variable and shock—the MSE difference generally disappears for large lag orders, corroborating our theoretical results.

Lusompa (2022) proposes an alternative LP estimator that exploits the serial correlation structure of the LP regression error.[17] In on-line Appendix D (Proposition 1) we show that, in a simple setting, this alternative LP estimator has the same asymptotic distribution as the standard LP estimator (as in (7)) at a given horizon, provided that a sufficiently large lag order is used. This result contrasts with Lusompa (2022, proposition 6), which shows an efficiency gain for the alternative LP estimator under knowledge of the true (finite) model order. Although the asymptotic theory under more general conditions warrants further investigation, our efficiency analysis suggests that, under the VAR(∞) model, the alternative estimator is unlikely to be more efficient than the standard OLS estimator under the classical

[16] Probably influenced by the inefficiency result on LP under finite-order VAR model, the literature has expressed mixed conjectures regarding the efficiency of LP under the *infinite-order model*. Lusompa (2022, footnote 5) notes that "... in the infinite lag case ... most people would probably assume this (LP is less efficient than VAR-implied estimator)".

[17] See also Breitung and Brüggemann (2023, section IV) for a similar estimator.

assumptions above, unless additional structural restrictions are imposed.

We emphasize that our asymptotic distributional equivalence result between LP and VAR concerns the behavior as the model order increases. For a small p and in finite samples, the bias-variance trade-off between LP and VAR remains non-trivial and has been explored in recent studies (Li, et al., 2024, Ferreira, et al., 2025, among others).

3 The bootstrap and simulation experiments

In finite samples, we suggest the following bootstrap implementation.

ALGORITHM 2 (Re-centered bootstrap CIs for $\nu_1'\beta_1(h)$).

Step 1 (Bootstrap samples). Given the data $\{y_t, t = 1, \cdots, n\}$ and the lag order p chosen by the empirical researcher, an OLS fit of the VAR(p) model (with no intercept) to $y_t - \widehat{\varrho}^y$, where $\widehat{\varrho}^y = n^{-1}\sum_{t=1}^{n} y_t$, yields slope estimates $\overline{a}_j$ and residuals $\overline{u}_t$. For $t = p+1, \cdots, n$, let $u_t^* = [\overline{u}_t - (n-p)^{-1}\sum_{t=p+1}^{n}\overline{u}_t]v_t$, where v_t is a sequence of I.I.D. zero-mean unit-variance random variables (we use standard normal variables), producing bootstrap randomness. Generate $Y_t^* = \sum_{j=1}^{p}\overline{a}_j Y_{t-j}^* + u_t^*$, where the starting values can be $Y_t^* = Y_t$, for $t = 1, \cdots, p$. A bootstrap sample is then obtained as $\{y_t^*, t = 1, \cdots, n\}$, where $y_t^* = \widehat{\varrho}^y + Y_t^*$.

Step 2 (Bias adjustment term). Using the original sample $\{y_t\}$, compute $\widehat{\beta}_1(h)$, $\widehat{V}$ and $\widehat{\text{bias}}$. Here $\widehat{\beta}_1(h)$ is the OLS of LP regression (1), and $\widehat{V}$ follows from ALGORITHM 1. $\widehat{\text{bias}}$ is defined as the average of $\nu_1'\widehat{\beta}_1^{\triangle}(h) - \nu_1'\widehat{\beta}_1^{VAR}(h)$ over B_{in} bootstrap samples, in each of which $\widehat{\beta}_1^{\triangle}(h)$ is computed for the bootstrap sample $\{y_t^{\triangle}\}$, generated by resampling the original sample $\{y_t\}$ following Step 1. Here $\widehat{\beta}_1^{VAR}(h)$ is the iterative VAR estimator of $\beta_1(h)$ based on VAR slopes $\{\overline{a}_1, \cdots, \overline{a}_p\}$.

Step 3 (Bootstrap test statistic). For a bootstrap sample $\{y_t^*\}$, construct the test statistic as following:

$$\widehat{S}^{*,bc} = \widehat{V}^{*-1/2}[\nu_1'\widehat{\beta}_1^*(h) - \widehat{\text{bias}}^* - \nu_1'\widehat{\beta}_1^{VAR}(h)]. \tag{8}$$

Here $\widehat{\beta}_1^*(h)$ and $\widehat{V}^*$are computed similarly as $\widehat{\beta}_1(h)$ and $\widehat{V}$ except that the bootstrap sample is used in place of the original sample. The estimated bias term $\widehat{\text{bias}}^*$is the average of $\nu_1'\widehat{\beta}_1^{*\triangle}(h) - \nu_1'\widehat{\beta}_1^{VAR,*}(h)$ over B_{in} bootstrap samples, in each of which $\widehat{\beta}_1^{*\triangle}(h)$ is computed for the bootstrap sample $\{y_t^{*\triangle}\}$, generated by resampling (B_{in} times) the bootstrap sample $\{y_t^*\}$ according to Step 1. $\widehat{\beta}_1^{VAR,*}(h)$ is the iterative VAR estimator of $\beta_1(h)$ based on the bootstrap sample $\{y_t^*\}$.

Step 4 (Confidence interval). Putting together the results in Steps 2 and 3, the $100(1-$

τ)%-level CI is constructed as

$$\nu_1' \widehat{\beta}_1(h) - \widehat{\text{bias}} \pm \widehat{V}^{1/2} \mathfrak{q}^*_{sym}(1-\tau), \tag{9}$$

where $\mathfrak{q}^*_{sym}(1-\tau)$ is the $(1-\tau)$-th quantile of B_{out} replications of $|\widehat{S}^{*,bc}|$ in (8).

Algorithm 2 is written for the variance estimator $\widehat{V}$ (and the statistic $\widehat{S}$). Corresponding bootstrap CIs can be obtained when $\widehat{V}_{HC}$ and $\widehat{S}_{HC}$ are used instead.[18] Algorithm 2 modifies the recursive residual-based bootstrap proposed by MOPM, which we refer to as the unadjusted bootstrap.[19] The re-centering step implements a bias adjustment of the point estimator, achieved by an inner-loop bootstrap following the same bootstrap.[20] The motivation for this direct adjustment[21] is that finite-sample bias appears to be the dominant factor distorting inference, especially for persistent data and for medium to long horizons. Bias adjustment alters (typically increases) the finite-sample variance of the point estimator, which is properly accounted for by the outer-loop bootstrap that generates adjusted critical values (Step 3).[22] For the re-centered statistic, bootstrap critical values are constructed symmetrically rather than as equal-tailed intervals (as in the unadjusted bootstrap).[23]

In implementation, the inner-loop bootstrap only computes the point estimator (not the standard error), and the bias adjustment term—being an average—can be estimated with good precision using a modest number of bootstrap replications, keeping the additional computational cost manageable. In the remainder of this section, we evaluate the finite-sample performance of both asymptotic and bootstrap inference methods for level impulse

[18] The uniform validity of the recursive bootstrap scheme applied to LP regression with the HC standard error and the associated CI coverage properties are recently studied by Velez (2024) under the AR model. Combined with the steps we provide in the proofs of Theorems 1–3, these arguments can be extended to prove bootstrap validity in our more general setting.

[19] For completeness, the unadjusted bootstrap is provided in Algorithm 3 in the on-line Appendix G, which extends the bootstrap of MOPM by allowing for non-zero data means.

[20] The bootstrap-based bias correction in constructing inference has been used in various contexts, e.g. Kilian (1998) and Gupta and Seo (2023), but with some important differences in implementation from ours.

[21] Although our focus is on inference, the bias-adjusted point estimator itself is also practically useful. It complements the proposal by Herbst and Johannsen (2024), whose analytical bias correction assumes directly observed shocks (which we do not). See Mei, et al. (2023) for alternative bias correction methods in panel LP regressions.

[22] MOPM also considered bias adjustment (as an optional step) but in a different manner. Their approach (analytically) bias-adjusts the VAR slopes in the bootstrap DGP, while we bias-adjust the test statistic and use unadjusted VAR slopes in the bootstrap DGP. In our simulation experiments, the latter approach provides more substantial improvement.

[23] The literature provides both theoretical and numerical support for symmetric critical values when the test statistic is nearly unbiased (as in our setting, owing to the bias correction); see Hall (1988), Hall and Horowitz (1996), Andrews (2002), Gonçalves and Kilian (2004), and Romano and Wolf (2006).

responses $\beta(h)$ based on the local projection regression studied in this paper.

Designs. To make the simulation designs empirically relevant, we fit three trivariate models to the monthly data used in Kilian (2009)—VAR(1), VAR(2) and VARMA(2,1) (all including anintercept)—and use them as our data generating processes (DGPs).[24]

The estimated coefficient matrices are used to compute the true impulse responses implied by the three mean models. The shocks are generated from a three-dimensional, time-invariant conditional-correlation Gaussian vector GARCH(1,1) model, with parameters and the correlation matrix obtained by fitting the vector-GARCH model to the residuals from each mean model. In the Kilian data, the individual first-order autocorrelation coefficients of the three series are -0.091, 0.971, and 0.984, respectively. Thus, the first series (growth rate of world oil production) exhibits only minor serial correlation, whereas the latter two (global real activity and the real price of oil) display strong serial persistence. These persistence properties, along with cross-sectional correlations and conditional heteroskedasticity, are captured in our DGPs.

We set the sample size to $n = 240$ and consider integer horizons ranging from 1 to 60. The number of simulation replications in simulations is 10,000 for the asymptotic CIs and 2000 for the bootstrap CIs. For the bootstrap CIs, we draw $B_{out} = 1000$ bootstrap replications in the outer loop to compute critical values and $B_{in} = 50$ replications in the inner loop to compute the bias adjustment term (for the re-centered bootstrap CI).[25] The nominal confidence level for all CIs is 90%.

For the lag order, we consider $p \in \{1, 2, 12\}$ for the VAR(1) DGP. The choice $p = 12$ represents a common and conservative specification for monthly data, without requiring knowledge of the true model order. We deliberately include $p = 1$ and $p = 2$, corresponding respectively to the LP with the true lag order and and the lag-augmented LP (the true order plus one), to examine the effect of lag augmentation. Similarly,we consider $p \in \{2, 3, 12\}$ for the VAR(2) DGP. For the VARMA(2,1) DGP, since no finite true lag order exists, we use $p \in \{2, 6, 12\}$ to illustrate performance across a range of lag choices.

Methods. We focus on the responses of the third variable to all three shocks. For these responses, we consider four methods for constructing confidence intervals: the asymptotic

[24]The three models are employed as simulation designs to generate different patterns of temporal dependence. We do not take a position on which model provides the best fit to the Kilian data.

[25]The choice of $B_{in} = 50$ is admittedly small for computational feasibility. However, in our settings, the MSE of the bias-adjusted point estimator does not change much as B_{in} increases. In empirical applications, where computational burden is less of a concern than in simulation studies, researchers can use larger values of B_{in} (and B_{out}) if desired.

Table 1: Asymptotic confidence intervals (CIs) under the DGP VAR(1). [Equation 3, $n = 240$.] [The standard error used is $\widehat{V}_{HC}^{1/2}$ (HC) or $\widehat{V}^{1/2}$ (MG).]

	Coverage (90%)						Median Length					
	HC CI			MG CI			HC CI			MG CI		
$h \diagdown p$	1	2	12	1	2	12	1	2	12	1	2	12
						Shock 1						
1	.892	.880	.855	.892	.880	.855	.0168	.0169	.0172	.0168	.0169	.0172
6	.914	.888	.839	.886	.885	.847	.0379	.0386	.0387	.0345	.0381	.0398
12	.914	.885	.818	.878	.882	.839	.0499	.0506	.0505	.0444	.0493	.0534
24	.922	.888	.806	.880	.885	.864	.0635	.0637	.0629	.0547	.0613	.0697
36	.917	.888	.795	.879	.878	.871	.0710	.0712	.0692	.0599	.0671	.0787
48	.915	.887	.791	.874	.875	.872	.0752	.0761	.0720	.0626	.0701	.0832
60	.914	.892	.770	.868	.875	.866	.0774	.0783	.0731	.0633	.0708	.0865
						Shock 2						
1	.834	.891	.865	.834	.891	.865	.0009	.0034	.0035	.0009	.0034	.0035
6	.425	.871	.842	.841	.880	.848	.0021	.0078	.0079	.0050	.0079	.0081
12	.318	.844	.797	.840	.870	.825	.0027	.0102	.0104	.0091	.0107	.0109
24	.242	.773	.721	.817	.809	.769	.0034	.0131	.0129	.0153	.0143	.0147
36	.204	.691	.613	.792	.738	.702	.0038	.0147	.0142	.0197	.0168	.0173
48	.179	.620	.532	.758	.677	.630	.0040	.0157	.0148	.0227	.0183	.0189
60	.163	.564	.471	.726	.630	.584	.0042	.0161	.0149	.0245	.0192	.0198
						Shock 3						
1	.774	.873	.830	.774	.873	.830	.043	.286	.275	.043	.286	.275
6	.343	.807	.754	.743	.807	.760	.095	.514	.510	.233	.515	.518
12	.237	.740	.676	.720	.749	.697	.123	.650	.643	.429	.661	.677
24	.169	.659	.597	.698	.691	.657	.156	.801	.789	.734	.845	.892
36	.136	.647	.585	.693	.695	.671	.176	.892	.860	.958	.962	1.033
48	.127	.670	.605	.714	.720	.723	.189	.943	.899	1.131	1.026	1.118
60	.127	.691	.627	.739	.763	.762	.198	.970	.907	1.256	1.071	1.159

CIs (ALGORITHM 1) and the re-centered bootstrap CIs (ALGORITHM 2), each implemented with either the variance estimator $\widehat{V}_{HC}$ or $\widehat{V}$. These are referred to as the HC CI, MG CI, bootstrap HC CI, and bootstrap MG CI, respectively. Simulation results are reported in Tables 1 and 2 for the asymptotic CIs under DGPs VAR(1) and VAR(2), and in Tables 3 and 4 for the bootstrap CIs. To save space, corresponding results for the VARMA(2,1) DGP are presented in on-line Appendix F. To assess the impact of bias adjustment on the bootstrap CIs, we also implement a simpler version without bias adjustment—the unadjusted bootstrap CIs (ALGORITHM 3, on-line Appendix G)—with results for all three DGPs reported in on-line Appendix G.

Results: Asymptotic methods. We first present results when the DGP is VAR(1). Because the data are highly persistent, researchers may choose between the stationary frame-

Table 2: Asymptotic confidence intervals (CIs) under the DGP VAR(2). [Equation 3, $n = 240$.] [The standard error used is $\widehat{V}_{HC}^{1/2}$ (HC) or $\widehat{V}^{1/2}$ (MG).]

	Coverage (90%)						Median Length					
	HC CI			MG CI			HC CI			MG CI		
$h \setminus p$	2	3	12	2	3	12	2	3	12	2	3	12
						Shock 1						
1	.890	.880	.869	.890	.880	.869	.0168	.0169	.0173	.0168	.0169	.0173
6	.907	.867	.836	.885	.869	.848	.0567	.0569	.0577	.0523	.0565	.0586
12	.910	.871	.821	.884	.871	.843	.0765	.0764	.0765	.0679	.0747	.0810
24	.911	.875	.797	.873	.873	.853	.0928	.0937	.0927	.0802	.0882	.1006
36	.915	.872	.792	.866	.873	.859	.0993	.0996	.0980	.0834	.0937	.1095
48	.905	.871	.766	.870	.871	.862	.1013	.1017	.0985	.0857	.0951	.1133
60	.915	.863	.773	.875	.867	.862	.1033	.1032	.0988	.0865	.0974	.1172
						Shock 2						
1	.891	.885	.863	.891	.885	.863	.0033	.0034	.0035	.0033	.0034	.0035
6	.800	.859	.844	.893	.863	.853	.0109	.0116	.0117	.0136	.0117	.0119
12	.779	.845	.811	.873	.860	.835	.0148	.0157	.0157	.0183	.0161	.0165
24	.755	.797	.751	.842	.834	.815	.0181	.0191	.0189	.0221	.0205	.0212
36	.722	.759	.703	.814	.799	.787	.0194	.0206	.0199	.0236	.0225	.0239
48	.711	.746	.678	.801	.795	.784	.0197	.0210	.0203	.0241	.0236	.0252
60	.714	.745	.675	.816	.806	.804	.0199	.0212	.0204	.0247	.0242	.0262
						Shock 3						
1	.869	.861	.820	.869	.861	.820	.254	.284	.274	.254	.284	.274
6	.694	.788	.737	.819	.785	.742	.692	.774	.764	.928	.775	.772
12	.639	.727	.689	.782	.744	.715	.885	.979	.975	1.225	1.011	1.026
24	.622	.683	.626	.762	.718	.693	1.051	1.170	1.150	1.455	1.248	1.312
36	.640	.689	.623	.788	.738	.718	1.123	1.254	1.214	1.535	1.366	1.465
48	.686	.719	.666	.832	.788	.783	1.143	1.278	1.239	1.561	1.421	1.548
60	.696	.766	.711	.859	.836	.841	1.153	1.279	1.238	1.597	1.463	1.601

work and the robust framework, which treats potential unit roots as part of the truth. Under the specification with $p = 1$, the MG CI is asymptotically valid within the stationarity framework, whereas the HC CI is generally not. The invalidity of the HC CI arises from serial correlation—particularly for the responses to shocks 2 and 3, which are recovered from highly persistent macro variables. Under the specification $p \geq 2$, both MG CI and HC CI become asymptotically valid under either the stationary or robust framework. These theoretical properties are supported by the simulation results reported in Table 1. Specifically, when $p = 1$, the HC CI exhibits severe undercoverage for responses to shocks 2 and 3, with coverage rates as low as 12.7% for a nominal 90% CI. In contrast, the MG CI achieves much more reasonable coverage across all shocks. When $p = 2$, the coverage of HC CI improves

Table 3: Re-centered bootstrap confidence intervals (CIs) under the DGP VAR(1). [Equation 3, $n = 240$.] [The standard error used is $\widehat{V}_{HC}^{1/2}$ (HC) or $\widehat{V}^{1/2}$ (MG).]

	Coverage (90%)						Median Length					
	HC CI			MG CI			HC CI			MG CI		
$h \diagdown p$	1	2	12	1	2	12	1	2	12	1	2	12
						Shock 1						
1	.912	.917	.915	.912	.917	.915	.0174	.0180	.0193	.0174	.0180	.0193
6	.906	.910	.889	.906	.895	.892	.0369	.0416	.0472	.0359	.0404	.0464
12	.925	.905	.895	.907	.899	.880	.0473	.0548	.0652	.0467	.0528	.0635
24	.909	.906	.910	.905	.901	.921	.0591	.0699	.0843	.0566	.0663	.0811
36	.906	.912	.897	.911	.905	.899	.0662	.0782	.0937	.0634	.0724	.0901
48	.908	.918	.889	.905	.895	.899	.0712	.0817	.1011	.0658	.0759	.0963
60	.910	.921	.900	.903	.918	.896	.0733	.0846	.1033	.0655	.0771	.0949
						Shock 2						
1	.876	.902	.896	.876	.902	.896	.0010	.0036	.0040	.0010	.0036	.0040
6	.861	.894	.915	.869	.879	.908	.0055	.0088	.0098	.0056	.0085	.0097
12	.852	.900	.888	.863	.893	.885	.0099	.0125	.0139	.0105	.0121	.0135
24	.841	.893	.874	.873	.893	.862	.0173	.0180	.0190	.0184	.0177	.0190
36	.808	.844	.846	.850	.839	.844	.0226	.0223	.0235	.0247	.0221	.0230
48	.790	.792	.795	.835	.783	.803	.0258	.0245	.0256	.0287	.0246	.0252
60	.767	.743	.754	.809	.739	.761	.0272	.0251	.0263	.0312	.0251	.0260
						Shock 3						
1	.873	.909	.888	.873	.909	.888	.051	.314	.322	.051	.314	.322
6	.869	.902	.891	.873	.898	.884	.281	.615	.658	.289	.582	.646
12	.854	.900	.870	.873	.889	.863	.514	.830	.900	.535	.772	.875
24	.832	.875	.845	.869	.863	.843	.820	1.103	1.156	.886	1.041	1.132
36	.811	.846	.814	.850	.844	.819	1.007	1.241	1.291	1.098	1.176	1.289
48	.792	.858	.833	.853	.850	.835	1.147	1.311	1.401	1.225	1.237	1.349
60	.810	.875	.845	.852	.865	.859	1.253	1.326	1.416	1.337	1.254	1.386

dramatically for shocks 2 and 3,[26] although it generally remains below that of the MG CI, particularly at longer horizons.

Using the agnostic lag order $p = 12$ is not without cost in finite samples. The most pronounced cost arises for the HC CI when the DGP is VAR(1); its coverage can fall by more than 12 percentage points compared to that for $p = 2$. In contrast, the cost of using the agnostic lag order is generally milder for the MG CI—sometimes substantially so—for instance, in the coverage of responses to shock 1 at medium and long horizons.

Regarding efficiency, note that both MG and HC CIs are centered on the same estimator (OLS) and use the same Gaussian critical values. Therefore, the superior coverage of the

[26] These findings are consistent with earlier simulation results in the literature (e.g., MOPM, under the scalar AR(1) model) that document similar behavior of HC standard errors in lag-augmented LP regressions.

Table 4: Re-centered bootstrap confidence intervals (CIs) under the DGP VAR(2). [Equation 3, $n = 240$.] [The standard error used is $\widehat{V}_{HC}^{1/2}$ (HC) or $\widehat{V}^{1/2}$ (MG).]

	Coverage (90%)						Median Length					
	HC CI			MG CI			HC CI			MG CI		
$h \diagdown p$	2	3	12	2	3	12	2	3	12	2	3	12
						Shock 1						
1	.916	.903	.900	.916	.903	.900	.0176	.0177	.0198	.0176	.0177	.0198
6	.923	.915	.894	.920	.901	.894	.0577	.0635	.0716	.0552	.0614	.0708
12	.915	.918	.902	.901	.908	.896	.0763	.0853	.1033	.0722	.0813	.1004
24	.918	.922	.900	.913	.902	.899	.0917	.1048	.1265	.0839	.0977	.1201
36	.898	.910	.895	.876	.910	.885	.0989	.1111	.1344	.0910	.1022	.1259
48	.919	.920	.906	.908	.897	.898	.1015	.1158	.1375	.0921	.1049	.1275
60	.912	.907	.908	.897	.903	.909	.1040	.1172	.1394	.0930	.1075	.1280
						Shock 2						
1	.915	.903	.904	.915	.903	.904	.0034	.0037	.0040	.0034	.0037	.0040
6	.925	.915	.896	.918	.903	.884	.0147	.0135	.0143	.0145	.0131	.0142
12	.898	.898	.873	.897	.894	.865	.0205	.0197	.0209	.0204	.0188	.0202
24	.887	.873	.862	.886	.856	.860	.0259	.0265	.0279	.0255	.0252	.0274
36	.865	.857	.851	.850	.856	.859	.0282	.0292	.0303	.0270	.0275	.0300
48	.870	.858	.857	.859	.851	.851	.0285	.0293	.0307	.0271	.0281	.0305
60	.861	.877	.869	.857	.870	.867	.0282	.0289	.0307	.0270	.0270	.0301
						Shock 3						
1	.899	.908	.891	.899	.908	.891	.279	.315	.326	.279	.315	.326
6	.886	.901	.886	.876	.897	.871	1.079	.958	1.005	1.047	.919	.961
12	.883	.907	.878	.868	.890	.872	1.474	1.324	1.391	1.420	1.253	1.334
24	.875	.857	.847	.878	.853	.840	1.747	1.660	1.747	1.709	1.559	1.672
36	.868	.867	.843	.875	.871	.852	1.795	1.680	1.836	1.735	1.590	1.786
48	.887	.881	.871	.898	.895	.886	1.742	1.643	1.816	1.707	1.560	1.763
60	.894	.875	.896	.897	.875	.878	1.760	1.649	1.798	1.678	1.540	1.739

MG CI relative to the HC CI can be attributed to the MG standard error, which appears to more accurately capture finite-sample estimation uncertainty. Consequently, the MG CI tends to be longer than the HC CI. When the two CIs achieve comparable coverage, their lengths are also similar.

The results for the VAR(2) DGP differ strikingly from those for the VAR(1) DGP, even though both DGPs are fitted to the same dataset as in our designs. Under the VAR(2) DGP, the HC CI exhibits substantially better *overall* coverage than under the VAR(1) DGP.

Interestingly, comparing $p = 2$ (the true lag order) and $p = 3$ (the true lag order plus one) shows that lag augmentation produces far less contrast than it does for the VAR(1) DGP. We conjecture that this pattern extends to higher-order VAR DGPs, consistent with our theoretical implication that lag augmentation becomes less crucial as the lag order increases.[27] As in the VAR(1) case, under the VAR(2) DGP the MG CI delivers similar or markedly better coverage than the HC CI across all shocks, lag orders, and horizons. Note that under the VAR(2) DGP, the MG CI is asymptotically valid for all lag orders considered (see Remark 8), whereas the HC CI lacks asymptotic justification when $p = 2$ under both the stationary and robust frameworks.

When the DGP is VARMA(2,1), among the lag orders considered, the HC CI performs best at $p = 6$, while the MG CI favors a smaller lag order, $p = 2$. Using $p = 12$ slightly reduces the coverage for MG CI, with a more pronounced reduction for the HC CI. For shock 1, the HC CI tends to over-cover when $p = 2$. Similar to the earlier VAR(1) and VAR(2) results, across all shocks, lag orders, and horizons, the MG CI provides more accurate and more robust coverage (with respect to lag order choice) than the HC CI.

Results: Bootstrap methods. In our simulations, the asymptotic CIs achieve satisfactory coverage for shock 1—particularly the MG CIs across all lag orders considered (and the HC CIs for certain lag choices). However, coverage is generally unsatisfactory, even for MG CIs, for shocks 2 and 3, especially at medium and long horizons.

The (re-centered) bootstrap procedure improves coverage for both types of CIs, often substantially so when the asymptotic CIs perform poorly, as shown in Tables 3, 4 (and Table F2 in on-line Appendix F). For instance, for $h \leq 24$, coverage rarely falls below 85% for either CI under any of the three DGPs. The improvement provided by the bootstrap is especially pronounced for the HC CI.

When comparing the two bootstrap CIs, the bootstrap MG CI generally delivers more stable coverage than the bootstrap HC CI, although the difference between them is much smaller than that observed for their asymptotic counterparts. Some over-coverage remains for the bootstrap HC CI for shock 1 under the VARMA(2,1) DGP, whereas the bootstrap MG CI nearly attains nominal coverage. In cases where the two bootstrap CIs exhibit comparable coverage, the bootstrap MG CI is typically noticeably shorter.

The bias adjustment to the test statistic plays an important role in constructing bootstrap CIs. Broadly across horizons, the bias adjustment improves coverage accuracy—sometimes substantially, by nearly 10 percentage points—relative to the unadjusted bootstrap CI. For example, under the VARMA(2,1) design, Figure 1 illustrates the improved performance by plotting the coverage rates of unadjusted and re-centered bootstrap CIs (using the MG standard error) when the lag order $p = 12$ is employed. For reference, the figure also includes the coverage rates of the asymptotic MG CIs. With the practical sample size and highly

[27] Our simulation experience suggests that the case $p = 1$ is somewhat unique—regardless of the DGP—often yielding results that become much less pronounced once larger lag orders are used.

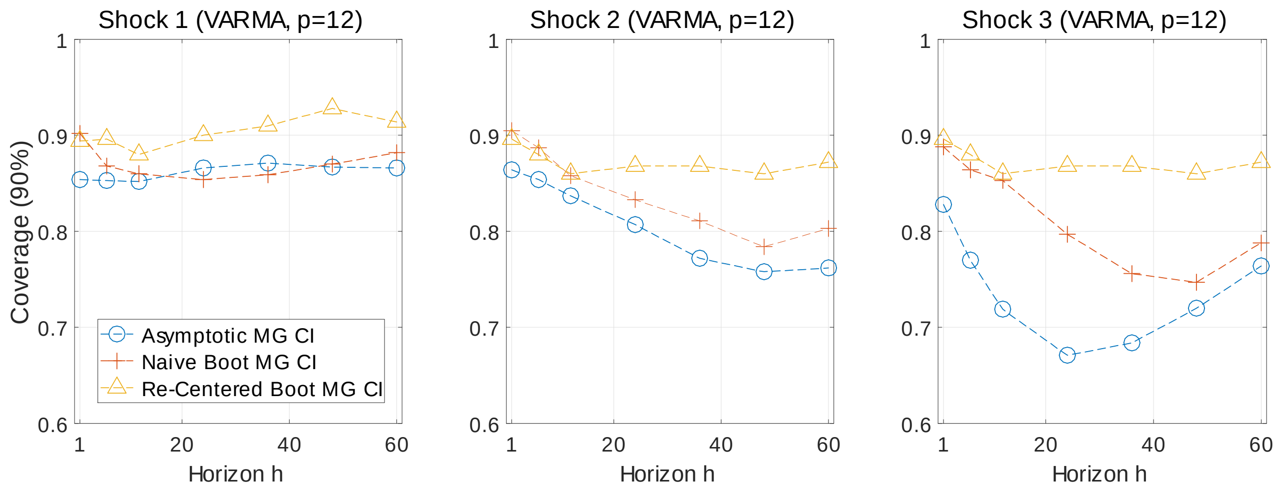

Figure 1: Coverage rates for asymptotic and (unadjusted and recentered) bootstrap CIs with the MG standard error. The nominal coverage rate is 90%. The DGP is VARMA(2,1), and $p = 12$.

persistent data, the re-centered bootstrap CIs achieve actual coverage rates that are close to the nominal level across a wide range of horizons. The re-centered CI is in general (only slightly, in many cases) wider than the unadjusted bootstrap CI.

4 Asymptotic theory

We provide the formal theory in this section. Section 4.1 introduces the inferential framework and key assumptions on the data-generating process, parameter space, and approximation model. Section 4.2 develops the uniform asymptotic theory for the LP estimator and discusses a few implications. Section 4.3 establishes the validity of inference based on two standard errors under alternative assumptions, each implying different forms of robustness. Neither variance estimator requires kernel-based long-run variance estimation.

Notation. 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, etc.) to denote a positive constant that 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 symmetric positive semi-definite matrix D, denote $\lambda_{\min}(D)$ and $\lambda_{\max}(D)$ as its smallest and largest (in magnitude) eigenvalues. Denote $\text{diag}(x_1, \cdots, x_K)$ as the block-diagonal matrix with blocks $x_1, \cdots, x_K$ (not necessarily of the same dimension) on the diagonal. We write $\otimes$ for the Kronecker product of matrices and $\text{vec}(D)$ for the column-wise vectorization of the matrix D. We write the $K \times K$ identity matrix as I_K.

4.1 Data generating process and parameter space

Suppose that the data-generating process (DGP) for the observed series y_t follows the K-dimensional vector autoregression of infinite order (VAR(∞)), $y_t = \varrho^y + Y_t$, where

$$\underset{K\times 1}{Y_t} = \sum_{j=1}^{\infty} a_j Y_{t-j} + u_t, \tag{10}$$

with u_t a serially uncorrelated shock process. Initial conditions are $Y_t = 0$, for $t \leq 0$. Assume $\varrho^y = 0$ without loss of generality. We can conveniently write (10) as $a(\mathrm{L})Y_t = u_t$, where $a(\mathrm{L}) = \mathrm{I}_K - \sum_{j=1}^{\infty} a_j \mathrm{L}^j$, with L being the lag operator such that $\mathrm{L}Y_t = Y_{t-1}$.

Write the DGP in the regression form:

$$y_t = \varrho_0 + \sum_{j=1}^{\infty} a_j y_{t-j} + u_t, \tag{11}$$

where $\varrho_0 = (\mathrm{I}_K - \sum_{j=1}^{\infty} a_j)\varrho^y$. The interest is in the horizon-h forecast-error impulse response matrix $\beta(h)$, for $h \geq 1$, which is (as usual) recursively defined from VAR coefficients. Structural responses obtained through recursive identification will be discussed at the end of this section. Equivalently, $\beta(h)$ is identified in the following LP(∞) form (local projection with infinite lags),

$$y_{t+h} = \varrho(h) + \beta(h)y_t + \sum_{\ell=1}^{\infty} \theta_\ell(h) y_{t-\ell} + \xi_t(h), \tag{12}$$

where $\xi_t(h) = \sum_{i=1}^{h} \beta(h-i)u_{t+i}$. The form (12) is obtained by recursive substitution in (11). Write $\beta(h) = [\beta_1(h), \cdots, \beta_K(h)]'$, then $\beta_1(h)$ contains the regression coefficient on y_t for the first entry $y_{1,t+h}$ of the response variable y_{t+h} in the system (12). Let $\beta(0) = \mathrm{I}_K$.

Under stationarity, $\beta(h)$ coincides with the lag-h coefficient matrix of the vector moving average (VMA(∞)) form of y_t. Specifying the DGP in the VAR form (instead of the VMA form) has the advantage of accommodating both stationary and integrated macro variables.

Focusing on the responses of y_{1t} (to the K shocks in u_t), the *parameter of interest* in this section 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 level responses and cumulative responses, corresponding to μ set as $\mu_{\mathrm{IR}} = (0, \cdots, 0, 1)'$ and $\mu_{\mathrm{CIR}} = (1, \cdots, 1)'$, respectively. To reconcile with earlier notation, we write $\beta_1(h, \mu_{\mathrm{IR}})$ simply as $\beta_1(h)$. In what follows we use this notational convention by implicitly imposing $\mu = \mu_{\mathrm{IR}}$

whenever we drop the argument μ, e.g. in $\pi_1(h,\mu)$, $y_{1t}(h,\mu)$, etc.

We first define the parameter space $\mathcal{A}$ of coefficients in the model (10), where $\mathcal{A} = \{a \in \mathbb{R}^\infty : a = \text{vec}(a_1, a_2, \cdots)\}$ is a subset of $\mathbb{R}^\infty$. Define the scalar quantity $\pi_k(h,\mu)$, which determines the convergence rate, by $\pi_k(h,\mu) = \sum_{i=1}^h |\varphi_{ki}|^2$, 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 (13) introduced below. Restrictions on $\mathcal{A}$ are imposed through conditions on impulse response matrices, as stated in Assumption 1.[28]

Assumption 1. For $k = 1, \cdots, K$ and a given μ, (i) $\sup_{a\in\mathcal{A}} \sup_{i\geq 0} |\beta_k(i)| \leq C_1$; (ii) $\sup_{a\in\mathcal{A}} \sup_{h\geq 1} [\pi_k(h,\mu)]^{-1} \sum_{i=1}^h |\varphi_{ki}| \leq C_2$; (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 condition. For illustration, if the DGP follows a scalar AR(1), the assumption requires $|a_1| \leq 1$. For an AR(2) process, the assumption can be satisfied by the 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 some constant $\varepsilon \in (0, 1]$. Assumption 1 thus rules out explosive roots (which would violate (i) and (ii)) and integration of order greated than one (which would violate (iii)) for the data y_t.

From a technical perspective, the conditions in Assumption 1 are especially useful for establishing uniform moment bounds in the asymptotic analysis.

Define $y_{1t}(h,\mu) = \sum_{j=1}^h \mu_j y_{1,t+j}$. Similarly to (12), and focusing on responses of the first entry of y_t (without loss of generality), we can write

$$y_{1t}(h,\mu) = \varrho_1(h,\mu) + \beta_1(h,\mu)' y_t + \sum_{\ell=1}^{\infty} \theta_{1\ell}(h,\mu)' y_{t-\ell} + \xi_{1t}(h,\mu), \tag{13}$$

where $\xi_{1t}(h,\mu) = \sum_{i=1}^h \varphi_{1i}' u_{t+i}$ is the LP(∞) regression error, with φ_{1i} being defined earlier. To estimate $\beta_1(h,\mu)$, we run the truncated regression

$$y_{1t}(h,\mu) = \varrho_1(h,\mu) + \beta_1(h,\mu)' y_t + \sum_{\ell=1}^{p-1} \theta_{1\ell}(h,\mu)' y_{t-\ell} + \eta_{1t}(h,\mu), \tag{14}$$

for a chosen model order p. The LP(p) regression error $\eta_{1t}(h,\mu)$ in (14) therefore has the form $\eta_{1t}(h,\mu) = \sum_{\ell=p}^{\infty} \theta_{1\ell}(h,\mu)' y_{t-\ell} + \xi_{1t}(h,\mu)$. Although we focus here on reduced-form impulse

[28] Although we allow drifting sequences in the parameter space and permit the horizon and model order to grow with n, for notational simplicity we do not write a_i, p, or h explicitly as functions of n.

responses, a non-random rotation of $\beta_1(h,\mu)$ can be given a structural interpretation; see Remark 11.

Our inferential framework is valid for a range of model orders p, such that $\underline{p} \le p \le \overline{p}$, where $\underline{p}$ and $\overline{p}$ are two positive integers so that Assumption 2 below holds.[29] Besides controlling bounds on p, Assumption 2 further restricts the coefficient parameter space. While Assumption 1 imposes conditions on all a_j, Assumption 2 concerns only the tail coefficients a_j, for $j \ge \underline{p}$. For the horizon h in (14), we consider the range $1 \le h \le \overline{h}$, where the upper bound $\overline{h}$ may grow with the sample size. Denote $\overline{\mu} = \sup_{1\le h\le \overline{h}} |\mu|_1$, where $|\mu|_1 = \sum_{j=1}^{h} |\mu_j|$.

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 an inherent tension between the ranges of model orders and response horizons; the upper bound on p becomes tighter as a wider range of horizons is allowed. Assumption 2(ii) regulates the degree of dynamic misspecification induced by truncation. Given such an upper bound $\overline{p}$, Assumption 2(ii) restricts the lower bound $\underline{p}$; $\underline{p}$ should be sufficiently large so that the tail 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 (non-local) misspecification bias vanishes.[30] Assumption 2(ii) may appear strong, but it can be substantially relaxed if we restrict attention to stationary models (Lewis and Reinsel, 1985, Gonçalves and Kilian, 2007).

If the data y_t follow a finite-order VAR(p_{true}) model with $a_{p_{\text{true}}} \ne 0$ and $a_j = 0$ for $j \ge p_{\text{true}}+1$, Assumption 2(ii) is trivially satisfied whenever $\underline{p} \ge p_{\text{true}}+1$. In other words, the lag order need not diverge if the true model is finite-order VAR. We can also consider a locally misspecified finite-order VAR process, typically modeled by allowing the MA coefficients to shrink toward zero, as in Montiel Olea, et al. (2024). If the local misspecification is only mild (the MA coefficients shrink to zero sufficiently fast), Assumption 2(ii) remains valid when $\underline{p} \ge p_{\text{true}} + 1$, like in the case of correct specification. At the other end, if the local misspecification is more pronounced (though not as severe as global misspecification in the nondegenerate VARMA model), $\underline{p}$ must diverge to eliminate the misspecification bias.[31] We

[29] Onatski and Uhlig (2012) warn of a potential problem with the estimated autoregressive roots when p is large. The problem does not apply here since our estimation and inference for impulse responses do not rely on estimating autoregressive roots.

[30] Assumption 2(ii) is stated for clarity of interpretation rather than being the weakest possible; see the weaker Assumption 2B in the Supplement which is sufficient for the asymptotic results developed below.

[31] In this paper we do not consider long-range dependence generated by fractionally integrated processes.

next state the 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 eighth-order 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, condition (i) assumes that u_t is a martingale difference sequence (MDS), and, in other words, the dynamic model (10) for the conditional expectation of y_t is correctly specified.[32] The mixing conditions in (ii) are standard regularity conditions, and they are used to establish uniform bounds for moments of sums (e.g. in establishing the law of large numbers for squared sequences to apply the MDS CLT). These conditions are stronger than strictly necessary.[33] The mixing conditions are widely used to control higher-order serial dependence in time series, e.g., Andrews and Guggenberger (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). Although unconditionally heteroskedastic errors are excluded by the stationarity condition on u_t, it should not affect the inference results presented later in the paper.[34] Condition (iii) rules out singular conditional variances. The moment condition (iv) on 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.

4.2 Local projection regression

In this subsection we study the asymptotic distribution of the point estimator of $\beta_1(h, \mu)$, and then construct the variance estimators and t-stat based inference procedures in the next subsection. Throughout the paper, the time series data available to the econometrician are

See Baillie, et al. (2017) for a recent study based on sieve VAR implication.

[32] Brüggemann, et al. (2016) used a weaker assumption, serially uncorrelated errors, and considered the inference of a conventional stationary finite-order VAR model.

[33] An alternative method is to directly impose assumptions on the summability of joint moments of the sequence $\{u_t\}$. We find that the approach based on mixing conditions is more intuitive and makes the proofs relatively transparent.

[34] The robustness to unconditional heteroskedasticity could be established (at a cost of more complicated proofs) since, as shown later, the effective regressors in the LP regression are only weakly dependent (although the data can be highly persistent); c.f. Phillips and Xu (2006).

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 $\beta_1(h, \mu)$, $\widehat{\beta}_1(h, \mu)$, a single-step OLS estimator, is obtained by OLS of (14) (for $t = p, \cdots, n - h$). We assume that the lag order p in the horizon-h regression is the same across h, so it is not a function of h.

By the partialling-out theorem, $\widehat{\beta}_1(h, \mu)$ can be rewritten as

$$\widehat{\beta}_1(h, \mu) = [\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)']^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h) y_{1t}(h, \mu), \tag{15}$$

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

The key insight of the uniform distributional theory for $\widehat{\beta}_1(h, \mu)$ lies in the 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 $p = p_{\text{true}} + 1$).[35]

Standard least squares algebra gives that

$$\widehat{\beta}_1(h, \mu) - \beta_1(h, \mu) = [\sum_{t=p}^{n-h} \widehat{u}_t(h)\widehat{u}_t(h)']^{-1} \sum_{t=p}^{n-h} \widehat{u}_t(h) \eta_{1t}(h, \mu). \tag{16}$$

Before stating the main theorem of the paper, we impose the following assumption.

Assumption 4. Let $\Delta y_t = y_t - y_{t-1}$. There exists a $K \times K$ nonsingular matrix of constants Q such that

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

where $\widetilde{X}_{t-1}^{Q}(1) = ((Qy_{t-1})', 1)'$ if $p = 1$, and $\widetilde{X}_{t-1}^{Q}(p) = ((Qy_{t-1})', \Delta y_{t-1}', \cdots, \Delta y_{t-p+1}', 1)'$ if

[35] In our setting, despite the analytically convenient (and seemingly two-step) form (15), the estimator $\widehat{\beta}_1(h, \mu)$ is in fact only one-step, OLS of (14). Breitung and Brüggemann (2023) study a more general projection-type estimator, formulated in two steps, the first of which estimates the shock that may need external information outside the VAR system. Also see their discussions, in a more structural setting, on how this first step (in our setting, the implicit partialling-out step) plays a key role in achieving the robustness to nonstationary data and also in simplifying the inference.

$p \geq 2$, $\Upsilon_n(p) = (n-p)^{1/2}\text{diag}(\Pi^Q(n)^{1/2}, \underbrace{\mathrm{I}_K, \cdots, \mathrm{I}_K}_{p-1 \text{ times}}, 1)$, with $\Pi^Q(n) = \text{diag}\{\pi_1^Q(n), \cdots, \pi_K^Q(n)\}$, $\pi_k^Q(n) = \sum_{i=0}^{n-1} |Q_k'\beta(i)|^2$, and Q_k' being the k-th row of Q, for $k = 1, \cdots, K$.

The eigenvalue condition in Assumption 4 is a rather technical but weak assumption on the DGP.[36] The intuition is analogous to that of the standard assumption which rules out asymptotically collinear "regressors". Note that the "regressors" in Assumption 4 are written in the augmented-Dickey-Fuller form so that the fast rate induced by potentially highly persistent data can be captured. If some component of y_t is integrated, cointegration of the system y_t of unknown rank is allowed by Assumption 4, thanks to the flexibility offered by the rotation matrix Q. If cointegration does not exist, Q is simply I_K. The presence of Q in Assumption 4 is crucial if there is non-trivial cointegration (i.e. the cointegration rank is strictly between 0 and K).[37] We now establish the asymptotic normality of $\nu_1'\widehat{\beta}_1(h, \mu)$, where ν_1 is a known nonzero $K \times 1$ vector.

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

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

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

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{\underline{p}\leq p\leq\overline{p}} \sup_{1\leq h\leq\overline{h}} \sup_{a\in\mathcal{A}} \left|\mathrm{P}\left(V^{-1/2}[\nu_1'\widehat{\beta}_1(h,\mu) - \nu_1'\beta_1(h,\mu)] \leq x\right) - \Phi(x)\right| = 0, \tag{17}$$

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

$$\underline{C}_V \leq \pi_1(h,\mu)^{-1}(n-h-p+1)V \leq \overline{C}_V, \tag{18}$$

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

The proof of Theorem 1 is given in on-line Appendix A.

Remark 1 (Convergence rate). Theorem 1 shows that the pointwise convergence rate

[36] MOPM also use a similar technical assumption (their assumption 3), which is argued as necessary for their uniform inference results. See also Montiel Olea and Plagborg-Møller (2022).

[37] The framwork of Montiel Olea and Plagborg-Møller (2021) allows some variables in y_t to be stationary and others to have unit roots. If all variables in y_t have unit roots, they model the first-order difference of y_t as the finite-order VAR, thereby ruling out non-trivial cointegration among y_t.

(determined by the asymptotic variance V) is $\pi_1(h,\mu)^{-1/2}n^{1/2}$, which depends on the degree of data persistence, the forecast horizon, and linear combination coefficients μ. The convergence rate for the cumulated response estimator is generally slower than that of the level response estimator. For example, 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. Under stationarity $|a_1| < 1$, both the level and cumulated response estimators achieve the standard $n^{1/2}$ rate, even when h is allowed to increase with the sample size. When the data are highly persistent ($a_1 \to 1$), the cumulative effects across horizons become dominant. Without assuming stationarity, the uniform convergence rates over the parameter space $|a_1| \leq 1$ are $h^{-1/2}n^{1/2}$ and $h^{-3/2}n^{1/2}$ for level and cumulated response estimators, respectively. Section 4.3 provides practical methods for estimating the asymptotic variance V.

Remark 2 (Finite-order DGP and finite p). Under a finite-order VAR data-generating process and a finite p, MOPM established a uniform asymptotic theory for the level response estimator. Specialized to the framework of MOPM, Theorem 1 extends their theory by incorporating a more general form of the asymptotic variance, whose validity does not rely on mean-independent shocks and thus accommodates general MDS shocks.

Remark 3 (Conditions on the lag order). For fixed horizons we only require $\overline{p}^2/n \to 0$ (Assumption 2(i)). Under stronger assumptions on data dependence (stationarity) and sometimes on shocks, the existing asymptotic theory for VAR-implied impulse response estimators often assumes a more stringent restriction on the lag order, $\overline{p}^3/n \to 0$. This restriction arises from the joint limit theory for slope matrices of all p lags of VAR model (Lewis and Reinsel, 1985, Gonçalves and Kilian, 2007), whereas our LP approach estimates only a single slope matrix—the coefficient matrix on y_t—and treats the remaining $p-1$ 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 proposed by Lusompa (2022) likewise requires $\overline{p}^4/n \to 0$ for his asymptotic theory under stationarity.

Remark 4 (Asymptotic variance). Theorem 1 holds under both (correctly specified) finite-order VAR and VAR(∞) models. Hence, in either asymptotic framework, the LP estimator exhibits the same form of asymptotic variance. This contrasts with the VAR-recursive impulse response estimator, whose asymptotic variance differs across the two asymptotic frameworks (Kilian and Lütkepohl, 2017, chapter 12.1.3). This discontinuity introduces a degree of subjectivity in choosing the asymptotic framework when the variance estimator relies on the explicit analytical form of the asymptotic variance (e.g. as in Kilian and Lütkepohl, 2017, p. 338).[38]

Remark 5 (Martingale representation of the score). The key tool used in the proof of Theorem 1 is a martingale representation. Examining (16), and neglecting terms of smaller order, the (effective) score $\sum_{t=p}^{n-h} u_t \xi_{1t}(h,\mu)$ which plays a central role in the asymptotic theory can be expressed as (see Lemma MART in Appendix A for a more general form)

$$\sum_{t=p}^{n-h} u_t \xi_{1t}(h,\mu) = \sum_{t=p+1}^{n} w_t, \tag{19}$$

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 (19) is that, although score contributions $u_t\xi_{1t}(h,\mu)$ are generally autocorrelated, the transformed summands w_t form (conditionally and unconditionally heteroskedastic) martingale differences with respect to their natural filtration, given that u_t is assumed to be an MDS. This representation enables the direct application of the martingale central limit theorem (CLT).

The approach of MOPM, in a more restricted setting, works directly with the raw (untransformed) score. To apply the CLT their method relies on a reverse-time argument, which consequentially requires shocks to be mean-independent of the future (in addition to the past), that is, $\mathrm{E}(u_t|u_s, s \ge t+1) = 0$, a.s. Eliminating such a full mean-independence assumption—though practically appealing—introduces additional technical challenges, requiring more involved uniform moment bound calculations in the proofs (see Appendix S1 in the Supplement).

Although the equality (19) is purely an algebraic rearrangement, it has deeper theoretical roots. The technique is closely related to the classical approach used to derive central limit theorems for stationary and serially correlated sequences, which one might, at least mechanically, apply to the score contributions $u_t\xi_{1t}(h,\mu)$ in our context (but we do not). A general version of this approach is due to Gordin (1969), in which a martingale approximation plays a central role.[39] Adapting this tool to the score contributions $u_t\xi_{1t}(h,\mu)$ yields the exact representation (19), rather than an approximation, owing to the moving-average structure of the LP(∞) error term $\xi_{1t}(h,\mu)$.[40] We next demonstrate that this representation is not merely

[38] We note that certain VAR-based inference methods avoid this discontinuity, as they are implemented using the conventional variance estimator or the bootstrap (Gonçalves and Kilian, 2004, 2007).

[39] See Beveridge and Nelson (1981), Phillips and Solo (1992), Wu and Woodroofe (2004), and Cuny and Merlevède (2014) for applications and further developments of the martingale approximation method.

[40] See Xu (2020b) for an earlier application of the martingale score representation in multiple-period return predictive regression, and the relation to the well-known Hodrick (1992) standard error in empirical asset

a technical trick but also provides the foundation for a new standard error construction.

4.3 A tale of two standard errors

To estimate the asymptotic variance V and construct test statistics for inference, heteroskedasticity and autocorrelation robust methods naturally arise (see Herbst and Johannsen, 2024, for related discussions in the LP context). In what follows, we consider two alternative approaches.

The first variance estimator for $\nu_1'\widehat{\beta}_1(h,\mu)$ is of the Eicker-Huber-White type, introduced by Montiel Olea and Plagborg-Møller (2021),

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

where $\widehat{\eta}_{1t}(h,\mu)$ is the OLS residual of (14), and $\widehat{\Sigma}(h) = (n-h-p+1)^{-1}\sum_{t=p}^{n-h}\widehat{u}_t(h)\widehat{u}_t(h)'$. We study $\widehat{V}_{HC}$ under the general framework introduced in Section 4.1. It turns out that $\widehat{V}_{HC}$ can recover V asymptotically even for regression (14) with serially dependent errors, but *only when* such serial dependence does not induce serial correlation in the score contributions—the latter potentially arising from conditional heteroskedasticity of unknown form. Therefore, although the variance estimator $\widehat{V}_{HC}$ is remarkably simple and robust, it is not truly "heteroskedasticity-consistent (HC)" for horizons $h \geq 2$ (unlike the case $h=1$), even though the notation retains the conventional HC subscript.

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.

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

The proof of Theorem 2 is given in on-line Appendix B. Theorem 2 shows that the consistency of the variance estimator $\widehat{V}_{HC}$ clearly requires the equality $\mathrm{Var}(\sum_{t=p}^{n-h}u_t\xi_{1t}(h,\mu)) = \sum_{t=p}^{n-h}\mathrm{E}\xi_{1t}(h,\mu)^2u_tu_t'$, which holds under the following Assumption 5.

pricing; see also West (1997).

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

Assumption 5 is implied by—and is therefore weaker than—commonly used assumptions on u_t such as conditional homoskedasticity or mean independence. Assumption 5 is relatively straightforward to verify empirically.

To provide concrete examples, consider a simple MDS $u_t = e_t e_{t-1}$, where e_t is I.I.D. (independent and identically distributed) 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 mean independence, which by definition requires u_t to be uncorrelated with any measurable function of random variables in the set $\{u_s : s \neq t\}$. Nevertheless, in this example the process $u_t\xi_t(h,\mu)$ remains serially uncorrelated, so Assumption 5 continues to hold.

Assumption 5 can, however, 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 previous example. In this case, it can be shown that the process $u_t\xi_t(h,\mu)$ exhibits serial correlation. This example serves as a simplified representation of an MDS with GARCH-type or stochastic volatility dynamics, which typically violate Assumption 5 when the innovation has a nonzero third moment. Nonzero—and even time-varying—third moments are frequently observed in empirical applications (Hansen, 1994, Conrad et al., 2013, Colacito et al., 2016).

To derive lower-level conditions for Assumption 5, denote score contributions as $s_t(h,\mu) = u_t\xi_{1t}(h,\mu)$. The corresponding autocovariance matrices of $s_t(h,\mu)$ are given by $\mathrm{E}s_t(h,\mu)s_{t-j}(h,\mu)' = \sum_{i=j+1}^{h} \mathrm{E}[u_{t+j}u_t'\varphi_{1i}'u_{t+i}u_{t+i}'\varphi_{1,i-j}]$, if $1 \leq j \leq h-1$, and equal to zero for $j \geq h$. Although the score contributions are serially uncorrelated beyond $h-1$ lags, the autocorrelations can play a role at smaller lags. Assumption 6 below imposes restrictions directly on the fourth joint cumulants of u_t, which are sufficient to ensure Assumption 5 holds for all horizons $h \geq 1$.

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

Like Assumption 5, Assumption 6 is weaker than imposing conditional homoskedasticity or future mean independence on u_t. The scalar version of this assumption has appeared in earlier studies to ensure valid inference in various time series contexts (e.g., Deo, 2000, condition A(vii)). Gonçalves and Kilian (2004, Assumption A'(iv')) show that a simliar condition is necessary for the validity of the recursive wild bootstrap applied to coefficient-based statistics in stationary AR models under conditional heteroskedasticity. Kuersteiner (2002, Assumption A2(iii)) also employs this condition in constructing optimal instruments for AR models.

We now propose a new standard error that does not rely on Assumption 5. By the martingale representation (19), we have $\mathrm{Var}\left(\sum_{t=p}^{n-h} u_t\xi_{1t}(h,\mu)\right) = \sum_{t=p+1}^{n} \mathrm{E}w_t w_t'$. This key identity motivates the following martingale (MG) variance estimator of V,

$$\widehat{V} = (n-h-p+1)^{-2}\nu_1'\widehat{\Sigma}(h)^{-1}\left(\sum_{t=p+1}^{n} \widehat{w}_t\widehat{w}_t'\right)\widehat{\Sigma}(h)^{-1}\nu_1, \tag{20}$$

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 (20), $\widetilde{\beta}_1(i)$ and $\widetilde{u}_t$ denote preliminary estimates of $\beta_1(i)$ and u_t, respectively, and both estimates are required to converge sufficiently fast so that Assumption 7 below is satisfied.

Assumption 7. The preliminary estimates $\widetilde{\beta}_1(i)$ and $\widetilde{u}_t$ in (20) satisfy

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

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

For an example of $\widetilde{\beta}_1(i)$, suppose that our primary interest lies in the level 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 as the residuals from the OLS VAR(p) regression (including an intercept). The following result establishes the uniform validity of inference based on the proposed 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 $\lim_{n\to\infty}\sup_{x\in\mathbb{R}}\sup_{\underline{p}\le p\le\overline{p}}\sup_{1\le h\le\overline{h}}\sup_{a\in\mathcal{A}}\left|\mathrm{P}(\widehat{S}\le x)-\Phi(x)\right| = 0.$

The proof of Theorem 3 is provided in on-line Appendix C.

Remark 6 (Negligibility). Theorem 3 shows that estimation of unknown parameters in the asymptotic variance of $\widehat{\beta}_1(h,\mu)$ has asymptotically negligible effects on the inference based on $\widehat{V}$ (provided Assumption 7 holds). This result contrasts with the effect of using these estimates in constructing the point estimator itself, as shown in online Appendix D, which generally has nontrivial—and typically inflationary—effects on the asymptotic variance.

Remark 7 (Cross equations). The MG variance estimator $\widehat{V}$ can be extended to conduct inference on cross-equation restrictions. Such restrictions are particularly useful when empirical researchers seek to assess the joint responses of multiple macroeconomic variables to a common shock. The parameter of interest is the response matrix $\beta(h,\mu)=\sum_{j=1}^{h}\mu_j\beta(j)$, which can be estimated by OLS on the regression (14) 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

$$\widehat{\Omega}=(n-h-p+1)^{-2}\nu(\widehat{\Sigma}(h)^{-1}\otimes \mathrm{I}_K)\left(\sum_{t=p+1}^{n}\widehat{W}_t\widehat{W}_t'\right)(\widehat{\Sigma}(h)^{-1}\otimes \mathrm{I}_K)\nu', \qquad (21)$$

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 $\widetilde{u}_t$ denoting preliminary estimates. The corresponding Wald statistic is $\{\nu\mathrm{vec}[\widehat{\beta}(h,\mu)-\beta(h,\mu)]\}'\widehat{\Omega}^{-1}\nu\mathrm{vec}[\widehat{\beta}(h,\mu)-\beta(h,\mu)]$. If $\nu\mathrm{vec}(\beta(h,\mu))$ reduces to the scalar $\nu_1'\beta_1(h,\mu)$ for some K-dimensional vector ν_1, the Wald statistic simplifies to the square of $\widehat{S}$ defined in Theorem 3.

Remark 8 (Finite-order models 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 relaxed to $\underline{p}\ge p_{\text{true}}$— meaning that lag augmentation is unnecessary — for the validity of the MG variance estimator $\widehat{V}$, provided that at least one lagged control is included in the LP regression (14) (i.e. $p\ge 2$).

To illustrate, let R_t be the (population) regression residual of y_t on $y_{t-1},\cdots,y_{t-p+1}$ and a constant. Requiring $p\ge 2$ guarantees that for inference of $\beta_1(h,\mu)$, the *effective* regressor R_t in the local projection regression (14) is stationary, provided y_t is not integrated or nearly integrated of order two or higher (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 (19) 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 justifying the MG variance estimator $\widehat{V}$ still applies, even though 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 depends only on current and past values of u_t. We thus continue to have the equality $\text{Var}\left(\sum_{t=p}^{n-h} R_t \xi_{1t}(h,\mu)\right) = \sum_{t=p+1}^{n} \mathrm{E} w_t^R w_t^{R\prime}$, which underlies the validity of the MG variance estimator $\widehat{V}$.

In contrast, as shown by MOPM, lag augmentation is crucial for the validity of $\widehat{V}_{HC}$. Without lag augmentation, R_t is not white noise, leading to non-zero serial correlation in $R_t \xi_{1t}(h,\mu)$ (*even if* u_t satisfies full mean independence). Consequently, $\text{Var}(\sum_{t=p}^{n-h} R_t \xi_{1t}(h,\mu)) \neq \sum_{t=p}^{n-h} \mathrm{E} R_t R_t' \xi_{1t}(h,\mu)^2$, thereby invalidating the variance estimator $\widehat{V}_{HC}$ and distorting inference based on $\widehat{S}_{HC}$.

Remark 9. Throughout the analysis, the system dimension K is assumed to be fixed. Allowing K to grow slowly with the sample size is possible, preserving the asymptotic validity of inference based on $\widehat{S}_{HC}$ and $\widehat{S}$, though it may require stronger assumptions (e.g., restricted horizons) and more complex proofs. Inference becomes challenging when K is large enough that Kp is of similar magnitude to n, even for fixed h. Cattaneo et al. (2018) show that in random-sampling settings, the standard HC variance estimator can be inconsistent when the number of covariates is of the same order as the sample size, and propose a dimension-robust alternative. Extending such ideas to large macro systems in our robust inference setting would be an interesting direction for future research.

Remark 10. We focus on impulse responses at a single horizon. Joint inference across multiple (fixed) horizons can be obtained by extending our single-horizon arguments—typically implemented in the visually intuitive form of a confidence band. This involves (i) establishing a joint Gaussian limit and estimating the variance matrix using one of our proposed methods, and (ii) constructing the band using the sup-t critical value of Montiel Olea and Plagborg-Møller (2019). Nevertheless, extending this to a growing set of horizons—subject only to the conditions in this paper—poses nontrivial challenges and merits further study.

Remark 11 (Structural responses). While our analysis focuses on reduced-form impulse responses, one can construct the MG standard error for the corresponding structural response identified through recursive zero restrictions.[41] Write the reduced-form error variance matrix Σ as $\Sigma = P\Sigma_{diag}P'$, where P is the $K \times K$ lower-triangular matrix of structural impact multipliers, and Σ_{diag} is a $K \times K$ diagonal matrix. All diagonal entries of P and Σ_{diag} are assumed to be positive. We distinguish between *the unit effect* (which sets all diagonal entries of P to one while leaving Σ_{diag} otherwise unrestricted; see Plagborg-Møller and Wolf, 2021, and Montiel Olea, et al., 2024, for LPs) and *the unit-standard-deviation effect* (which imposes $\Sigma_{diag} = \mathrm{I}_K$ while leaving the diagonal entries of P otherwise unrestricted; see Jordà,

[41] The recursive identification scheme is widely used; see Kilian, et al. (2025) for a recent discussion.

2005). This factorization can be obtained (directly or after transformation) via the Cholesky decomposition of Σ. Denote $P = P(\Sigma)$ to emphasize that P is a smooth function Σ.

Let P_{j_0} denote the j_0-th column of P. One may be interested in $\gamma_{1j_0}(h,\mu) = \beta_1(h,\mu)' P_{j_0}$, for $h \geq 1$. This object can be interpreted as the (level or cumulated, depending on the specification of μ) impulse response of the first outcome variable to the j_0-th of the structural shocks in $P^{-1}u_t$, which may or may not be normalized to have unit variance. As mentioned above, P_{j_0} is a smooth function of Σ, so write $P_{j_0} = P_{j_0}(\Sigma)$. Under the recursive design, the first $j_0 - 1$ entries of P_{j_0} are zero. Let $\widehat{\Sigma} = (n-p)^{-1} \sum_{t=p}^{n-1} \widehat{u}_t \widehat{u}_t'$, where $\widehat{u}_t$ is the LP regression residual for $h = 1$. The estimator of $\gamma_{1j_0}(h,\mu)$ is $\widehat{\gamma}_{1j_0}(h,\mu) = \widehat{\beta}_1(h,\mu)' \widehat{P}_{j_0}$, where $\widehat{\beta}_1(h,\mu)$ is the LP estimator defined in (15), and $\widehat{P}_{j_0} = P_{j_0}(\widehat{\Sigma})$. Under the recursive design, $\widehat{P}$ can be obtained from the Cholesky decomposition of $\widehat{\Sigma}$.

To obtain the MG standard error for $\widehat{\gamma}_{1j_0}(h,\mu)$, let vech be the half-vectorization operator of the variance matrix Σ and denote $\nabla P_{j_0} = \frac{\partial P_{j_0}(\Sigma)}{\partial \mathrm{vech}(\Sigma)'}$, which depends only on Σ and constants (like zeros and ones).[42] Assume $\nabla P_{j_0} \neq 0$. Define $\underset{K_0 \times 1}{\widehat{\psi}_t} = \begin{pmatrix} \widehat{w}_t \\ \mathrm{vech}(\widehat{u}_t \widehat{u}_t' - \widehat{\Sigma}) \end{pmatrix}$, for $t = p+1, \cdots, n$, where $\widehat{w}_t$ is defined in (20) and $K_0 = K + K(K+1)/2$. The MG variance estimator for $\widehat{\gamma}_{1j_0}(h,\mu)$ is obtained via the delta method as $\widehat{\mathcal{V}} = \widehat{\Lambda}' \widehat{\Xi} \widehat{\mathrm{LRV}}(\widehat{\psi}_t, J) \widehat{\Xi} \widehat{\Lambda}$ where,

$$
\begin{aligned}
\underset{K_0 \times 1}{\widehat{\Lambda}} &= \begin{pmatrix} \widehat{P}_{j_0} \\ (\widehat{\nabla P_{j_0}})' \widehat{\beta}_1(h,\mu) \end{pmatrix}, \quad \underset{K_0 \times K_0}{\widehat{\Xi}} = \begin{pmatrix} [\sum_{t=p}^{n-h} \widehat{u}_t(h) \widehat{u}_t(h)']^{-1} & 0 \\ 0 & \mathrm{I}_{K(K+1)/2} \end{pmatrix}, \\
\underset{K_0 \times K_0}{\widehat{\mathrm{LRV}}(\widehat{\psi}_t, J)} &= \widehat{\Gamma}_0 + \sum_{i=1}^{n-p-1} \mathcal{K}(i/J)(\widehat{\Gamma}_i + \widehat{\Gamma}_i'), \quad \widehat{\Gamma}_i = \sum_{t=p+1+i}^{n} \widehat{\psi}_t \widehat{\psi}_{t-i}',
\end{aligned}
$$

with $\widehat{\nabla P_{j_0}} = \frac{\partial P_{j_0}(\widehat{\Sigma})}{\partial \mathrm{vech}(\Sigma)'}$. In the long-run variance matrix estimator $\widehat{\mathrm{LRV}}(\widehat{\psi}_t, J)$, practitioners must specify the kernel function $\mathcal{K}(\cdot)$ and the truncation parameter J; their selection can follow the recommendations of Andrews (1991). The HAC-type (heteroskedasticity and autocorrelation consistent) long-run variance matrix estimator is generally required for valid inference on structural responses under general conditional heteroskedasticity.

Inference on structural responses can be greatly simplified if $\nabla P_{j_0} = 0$; in that case, the HAC long-run variance matrix estimator described above is unnecessary. When this condition holds, P_{j_0} is a vector of constants. Focusing on the first outcome variable in y_t, our inference results in Theorems 2 and 3 apply directly when $\nu_1 = P_{j_0}$. An especially

[42] The expression for ∇P_{j_0} can be found in standard textbooks. For example, if the unit-standard-deviation effect is of interest, ∇P_{j_0} is formed by the $[K(j_0-1)+1]$-th to the (Kj_0)-th rows of the matrix denoted as H in Lütkepohl (2005, page 111).

interesting case arises for the response of $y_{i,t+h}$, where $1 \leq i \leq K$, to the *last* structural shock (the K-th entry of $P^{-1}u_t$) under the *unit-effect* normalization, when the shocks are correctly ordered under the recursive identification scheme. In this case, having $j_0 = K$ and $P_{j_0} = (0, \cdots, 0, 1)'$, structural and reduced-form responses coincide, and all impact responses are zero except that of y_{Kt} (which equals one). For example, letting $i = 1$, for $h \geq 1$, the structural response of $y_{1,t+h}$ is given by the coefficient on y_{Kt} in the local projection regression (14).

5 Concluding remarks

Local projections are straightforward to implement. Recent research also highlights their advantages in conducting inference on impulse responses relative to alternative methods, particularly in terms of achieving uniform validity. This paper contributes to the literature by showing that, in realistic settings, local projections can be more efficient than previously recognized. In fact, under classical assumptions of stationarity and homoskedastic martingale-difference shocks, LPs can be among the most efficient estimators of impulse responses when the controlled lag order diverges.

Using a large number of lags is also practically justified, as finite-order VAR models are best viewed as approximations to the true data-generating processes implied by structural macroeconomic models. In this paper, we analyze the asymptotic properties of LP regressions and propose new inference methods that allow researchers to remain relatively agnostic about the degree of persistence and the nature of heteroskedasticity in the data.

We have discussed several potential extensions, and a few additional ones merit consideration. First, this paper establishes the asymptotic equivalence between the sieve LP and sieve VAR estimators under classical assumptions on the data-generating process. To examine whether this equivalence extends to settings with more realistic data features, a more general asymptotic theory for the sieve VAR estimator—particularly a closed-form expression for its asymptotic variance—is required. Some progress in this direction is reported in Xu (2025). Second, we focus on the simple OLS estimator due to its widespread practical use. However, a broader class of estimators—especially those based on shrinkage or machine learning methods—could be explored to accommodate higher-dimensional settings with many regressors. While related studies exist, as noted in the introduction, issues of uniform validity and robust inference remain open. Third, the framework developed here can be extended to a wider class of structural models and to panel data settings, building on the results presented in this paper.

Data Availability Statement

The MATLAB code for implementation and replication of the simulation results is available on Zenodo at https://dx.doi.org/10.5281/zenodo.18864155.

References

Andrews, D. W. K., 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 59, 817-858.

Adamek, R., S. Smeekes, I. Wilms, 2024. Local Projection Inference in High Dimensions, *Econometrics Journal* 27, 323-342.

Andrews, D. W. K., 2002. Higher-order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators, *Econometrica* 70, 119-162.

Andrews, D. W. K., P. Guggenberger, 2014. A Conditional-Heteroskedasticity-Robust Confidence Interval for the Autoregressive Parameter, *Review of Economics and Statistics* 96, 376-381.

Baillie, R., G. Kapetanios, F. Papailias, 2017. Inference for Impulse Response Coefficients from Multivariate Fractionally Integrated Processes, *Econometric Reviews* 36, 60–84.

Ballarin, G., 2021. On the Limiting Distribution of Sieve VAR(∞) Estimators in Small Samples. arXiv:2104.13693.

Beveridge, S., 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.

Breitung, J., R. Brüggemann, 2023. Projection Estimators for Structural Impulse Responses, *Oxford Bulletin of Economics and Statistics* 85, 1320-1340.

Brüggemann, R. C. Jentsch, C. Trenkler, 2016. Inference in VARs with Conditional Heteroskedasticity of Unknown Form, *Journal of Econometrics* 191, 69–85.

Brugnolini, L., 2018. About Local Projection Impulse Response Function Reliability, Manuscript, University of Rome Tor Vergata.

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

Cattaneo, M., M. Jansson, W. Newey, 2018. Inference in Linear Regression Models with Many Covariates and Heteroscedasticity, *Journal of the American Statistical Association* 113, 1350-1361.

Cha, J., 2024. Local Projections Inference with High-dimensional Controls without Sparsity, Manuscript, Vanderbilt University.

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

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

Cloyne, J., N. Dimsdale, P. Hürtgen, 2025. Are Tax Cuts Contractionary at the Zero Lower Bound? Evidence from a Century of Data, *Journal of Political Economy* 133(2).

Colacito, R., E. Ghysels, J. Meng, 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, E. Ghysels, 2013. Ex ante Skewness and Expected Stock Returns, *Journal of Finance* 68, 85-124.

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

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

Dinh, V. H., D. Nibbering, B. Wong, 2024. Random Subspace Local Projections, *Review of Economics and Statistics*, forthcoming.

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

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

Dufour, J.-M., E. Renault, V. Zinde-Walsh, 2025. Wald Tests when Restrictions are Locally Singular, *Annals of Statistics* 53, 457-476.

Dufour, J.-M., E. Wang, 2024. Simple Robust Two-stage Estimation and Inference for Generalized Impulse Responses and Multi-horizon Causality, Manuscript, McGill University.

Engle, R., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50, 987-1007.

Ferreira, L. N., S. Miranda-Agrippino, G. Ricco, 2025. Bayesian Local Projections, *Review of Economics and Statistics* 107, 1424-1438.

Gonçalves S., A. M. Herrera, L. Kilian, E. Pesavento, 2024. State-Dependent Local Projections. *Journal of Econometrics* 244, 105702.

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

Gonçalves, S., 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.

Gospodinov, N., 2004. Asymptotic Confidence Intervals for Impulse Responses of Near-Integrated Processes, *Econometrics Journal* 7, 505–527.

Gupta, A., M. H. Seo, 2023. Robust Inference on Infinite and Growing Dimensional Time-Series Regression, *Econometrica* 91, 1333–1361.

Hall, P., 1988. On Symmetric Bootstrap Confidence Intervals, *Journal of the Royal Statistical Society (Series B)* 50, 35–45.

Hall, P., J. Horowitz, 1996. Bootstrap Critical Values for Tests based on Generalized-method-of-moment Estimators, *Econometrica* 64, 891–916.

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

Herbst, E. P., B. K. Johannsen, 2024. Bias in Local Projections, *Journal of Econometrics* 240, 105655.

Hodrick, R. J., 1992. Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies* 5, 357-386.

Huber, F., C. Matthes, M. Pfarrhofer, 2025. General Seemingly Unrelated Local Projections, arXiv:2410.17105.

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

Inoue, A., B. Rossi, Y. Wang, 2024. Local Projections in Unstable Environments, *Journal of Econometrics* 244, 105726.

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

Jordà, Ò., 2023. Local Projections for Applied Economics, *Annual Review of Economics* 15, 607-631.

Jordà, Ò., 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.

Jordà, Ò., A. M. Taylor, 2025. Local Projections, *Journal of Economic Literature* 63, 59-110.

Känzig, D. R., 2021. The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements, *American Economic Review* 111, 1092-1125.

Känzig, D. R., 2023. The Unequal Economic Consequences of Carbon Pricing, NBER Working Paper 31221.

Kilian, L., 1998. Small-sample Confidence Intervals for Impulse Response Functions, *Review of Economics and Statistics* 80, 218-230.

Kilian, L., 2009. Not All Oil Price Shocks are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market, *American Economic Review* 99, 1053-1069.

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

Kilian, L., M. Plante, A. Richter, 2025. Macroeconomic Responses to Uncertainty Shocks: The Perils of Recursive Orderings, *Journal of Applied Econometrics* 40, 395-410.

Kolesár, M., and M. Plagborg-Møller, 2024. Dynamic Causal Effects in a Nonlinear World: the Good, the Bad, and the Ugly. arXiv:2411.10415v2.

Kuersteiner, G., 2002. Efficient IV Estimation for Autoregressive Models with Conditional Heteroskedasticity, *Econometric Theory* 18, 547-583.

Leeb, H., B.M. Pötscher, 2005. Model Selection and Inference: Facts and Fiction, *Econometric Theory* 21, 21–59.

Lewis, D. J., 2021. Identifying Shocks via Time-Varying Volatility, *Review of Economic Studies* 88, 3086-3124.

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

Li, D., M. Plagborg-Møller, C. K. Wolf, 2024. Local Projections vs. VARs: Lessons From Thousands of DGPs, *Journal of Econometrics* 244, 105722.

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, M. W. Watson, 2006. A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series, *Journal of Econometrics* 135, 499-526.

Mei, Z., L. Sheng, Z. Shi, 2023. Nickell Bias in Panel Local Projection, arXiv:2302.13455v3.

Metcalf, G. E., J. H. Stock, 2020. Measuring the Macroeconomic Impact of Carbon Taxes, *AEA Papers and Proceedings*, 110, 101–106.

Mikusheva, A., 2012. One-Dimensional Inference in Autoregressive Models With the Potential Presence of a Unit Root, *Econometrica* 80, 173-212.

Montiel Olea, J. L., M. Plagborg-Møller, 2019. Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs, *Journal of Applied Econometrics* 34, 1-17.

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

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

Montiel Olea, J. L., M. Plagborg-Møller, E. Qian, C. K. Wolf, 2024. Double Robustness of Local Projections and Some Unpleasant VARithmetic, August 4, 2024, arXiv:2405.09509v2.

Nakamura, E., J. Steinsson, 2018. Identification in Macroeconomics, *Journal of Economic Perspectives* 32, 59-86.

Obstfeld, M., H. Zhou, 2022. The Global Dollar Cycle, *Brookings Papers on Economic Activity*, Fall 2022.

Onatski, A., H. Uhlig, 2012. Unit Roots in White Noise. *Econometric Theory* 28, 485-508.

Pesavento, E., B. Rossi, 2007. Impulse Response Confidence Intervals for Persistent Data: What Have We Learned? *Journal of Economic Dynamics and Control* 31, 2398–2412.

Phillips, P. C. B., 1998. Impulse Response and Forecast Error Variance Asymptotics in Non-Stationary VARs, *Journal of Econometrics* 83, 21–56.

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

Phillips, P. C. B., K.-L. Xu, 2006. Inference in Autoregression under Heteroskedasticity, *Journal of Time Series Analysis* 27, 289-308.

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

Ringo, D., 2025. Monetary Policy and Home Buying Inequality, *Review of Economics and Statistics*, forthcoming.

Romano, J., M. Wolf, 2006. Improved Nonparametric Confidence Intervals in Time Series Regressions, *Journal of Nonparametric Statistics* 18, 199-214.

Romer, C. D., D. H. Romer, 2004. A New Measure of Monetary Shocks: Derivation and Implications, *American Economic Review* 94, 1055–84.

Schorfheide, F., 2005. VAR Forecasting under Misspecification, *Journal of Econometrics*, 128, 99–136.

Stock, J. H., M. W. Watson, 2018. Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments, *Economic Journal* 128, 917-948.

Velez, A., 2024. The Local Projection Residual Bootstrap for AR(1) Models, arXiv.

West, K. D., 1997. Another Heteroskedasticity-and Autocorrelation-Consistent Covariance Matrix Estimator, *Journal of Econometrics* 76, 171-191.

White, H., 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48, 817-838.

Wright, J. H., 2000. Confidence Intervals for Univariate Impulse Responses with a Near Unit Root, *Journal of Business & Economic Statistics* 18, 368–373.

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

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

Xu, K.-L., 2020b. Standard Errors for Predictive Regression with Overlapping Observations, SSRN 4470760.

Xu, K.-L., 2025. Impulse Responses, Efficiency Bounds, and Newey-West Inference, Manuscript, Indiana University Bloomington.