Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Micro Responses to Macro Shocks

Martín Almuzara and Víctor Sancibrián

Working Paper n. 732

This Version: July 2026

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Micro Responses to Macro Shocks*

Martín Almuzara† Víctor Sancibrián‡

July 2026

[Link to Appendix]

Abstract

We study panel data regression models when the shocks of interest are aggregate and there are omitted macro and micro-level shocks of any relative size. This speaks to a large empirical literature that targets impulse responses via panel local projections. We show how to interpret the estimated coefficients when responses are heterogeneous and that a simple recipe leads to uniformly valid inference over the macro–micro composition of the errors: including lags as controls and then clustering at the time level. Finally, we use our methods to reassess the role of firm financial frictions in shaping the transmission of monetary policy.

Keywords: Panel data, local projections, impulse responses, aggregate shocks, inference, macro–micro data, heterogeneity.

*We thank the Co-Editor, a second consulting Co-Editor and three anonymous referees for useful feedback and comments. We have also greatly benefited from discussions with Manuel Arellano, Dmitry Arkhangelsky, Stéphane Bonhomme, Richard Crump, Òscar Jordà, Daniel Lewis, Laura Liu, Geert Mesters, Mikkel Plagborg-Møller, Enrique Sentana, Andres Santos, Dmitriy Sergeyev, Thomas Winberry, Ke-Li Xu and seminar and conference participants at Banco Central de Reserva del Perú, Bocconi, Boston University, CEMFI, CUNY, Erasmus University Rotterdam, Georgetown, Philadelphia Fed, Princeton, Rutgers, Toulouse, Universidad Autónoma de Madrid and University of Pennsylvania, the 2023 EABCN-UPF Conference, 2023 Greater New York Metro Area Econometrics Colloquium, 2024 NBER Summer Institute, 2024 SED Annual Meeting, 2024 System Econometrics Conference at Richmond Fed, 2024 CEME Conference for Young Econometricians, 2025 Workshop on Applied Microeconometrics and Panel Data, 2026 ASSA Annual Meeting, and the 2026 Macro Week Conference at Universidad del Pacífico. Stefano Graziosi, Babur Kocaoglu and Jackson Owen provided outstanding research assistance. Sancibrián acknowledges financial support from Grant PRE2022-000906 funded by MCIN/AEI/10.13039/501100011033 and by "ESF +", the María de Maeztu Unit of Excellence CEMFI MDM-2016-0684, funded by MCIN/AEI/10.13039/501100011033, Fundación Ramón Areces and CEMFI. Part of this work was done during Sancibrián's PhD studies at CEMFI. The views expressed in this paper do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

†FRBNY: martin.almuzara@ny.frb.org

‡Bocconi University and IGIER: victor.sancibrian@unibocconi.it

1 Introduction

Applied macroeconomists are increasingly interested in empirical estimates of the transmission of aggregate shocks to individual outcomes, often in the form of impulse responses.

A popular approach is to formulate estimating equations of the form

$$Y_{i,t+h} = \beta(h) s_i X_t + \text{controls} + \xi_{it}(h), \tag{1}$$

where Y_{it} is a *micro outcome* for unit i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, T$) and X_t a *macro shock* of interest. Shocks are often interacted with unit-level covariates s_i to document heterogeneity in transmission along observables. For example, Ottonello and Winberry (2020), Cloyne, Ferreira, Froemel, and Surico (2023) and Jeenas and Lagos (2024) are interested in the heterogeneous effects of monetary policy shocks X_t on firm-level investment along different margins, such as firm age, leverage or liquidity. Estimates $\hat{\beta}(h)$ of the response at horizon h are obtained via least squares; a panel local projection version of Jordà (2005).

Despite its routine application, little is known about the statistical properties of $\hat{\beta}(h)$. The way standard errors are computed in the empirical literature illustrates this well: in our own survey of 61 recent papers, slightly less than half compute two-way clustered standard errors, more than one-third cluster within units only, and a sizable minority resort to Driscoll and Kraay (1998).[1] This reflects the vastly different ways in which researchers perceive the nature of shocks, the role of each dimension of the panel for precision, and the importance of aggregate variation in the data.

In this paper, we provide the first treatment of estimation and inference for this problem. We show how to interpret $\hat{\beta}(h)$ when impulse-response heterogeneity is unrestricted, and propose standard errors and confidence intervals that are easy to compute and robust to the relative importance of macro shocks in the microdata. This requires clustering standard errors at the time level and ex-ante including enough lags as controls. We refer to this as *time-clustered lag-augmented heteroskedasticity-robust (t-LAHR) inference.*

We establish our results in a comprehensive setup featuring heterogeneous responses, general forms of serial and cross-sectional dependence, and unobserved macro and micro shocks of any relative size. Our notion of a macro shock is that of an innovation uncorrelated

[1] The typical application uses administrative data on firms or households and tracks units for a limited number of periods (with a mode around $T \approx 30$ and several other applications around $T \approx 100$). The economic content of X_t is very diverse, including monetary and fiscal policy shocks, investment shocks, TFP and innovation shocks, temperature shocks, etc. We reproduce the full list in Appendix F.

to its own lags and leads and other shocks, as is standard in the time series literature (Leeper, Sims, and Zha, 1996; Ramey, 2016; Stock and Watson, 2018; Montiel Olea and Plagborg-Møller, 2021). This is an identifying assumption without which the estimand of $\hat{\beta}(h)$ may not be interpretable causally. Empirical researchers widely recognize this and devote great efforts to constructing and motivating X_t by leveraging macroeconometric methods, and our analysis extends these different strategies to a rich panel environment. Specifically, we cover settings where the shock of interest is directly observed (the most prevalent assumption in applications), those where it is recoverable from a macro system (via, say, recursive or long-run identification), and those where it is contaminated with measurement error but a proxy is available (as in local projection-instrumental variables; LP-IV for short).

In that setup, we first show that $\hat{\beta}(h)$ recovers the slope coefficient of a population linear projection of unit-specific impulse responses on the characteristics s_i, thereby formalizing what practitioners have in mind when including interactions in (1). This is also the case when the shock of interest is unobserved, such as in the IV setup where only a noisy measurement is available. If $s_i = 1$, the estimand boils down to the average response in the population. Crucially, since we place no restrictions on the underlying cross-sectional impulse-response heterogeneity or in s_i, our characterization of the estimand is effectively nonparametric.[2]

Inference with an arbitrary macro–micro structure. The relative importance of aggregate shocks in the microdata is a key dimension of the problem: Observable macro shocks drive identification; omitted macro shocks shape the extent of spatial dependence and dynamics in the errors. As such, how sizable these are determines both the strength of identifying variation and the nature of uncertainty.[3]

The notion of different macro-micro regimes has direct counterparts in applied work: the ubiquitous nature of regression models of the form in (1) contrasts with the breadth of empirical settings to which it is applied (workers, firms, counties), the richness of the data (cross-sectional vs. time-series), the range of specification choices (unit-level vs. aggregate controls) and the way standard errors are calculated, all of which point to an implicit stance on the relative importance of aggregate shocks in a given empirical setting. For example, clustering within units rules out spatial dependence induced by aggregate shocks, and two-

[2]We discuss extensions to (exogenous) time-varying characteristics s_{it} in Section 3 (Remark 3).

[3]It is immediate that if $s_i = 1$ in Equation (1), including time fixed effects causes collinearity. If s_i varies over units, for time indicators to remove all additional aggregate variation one would need the untenable assumption that *only* impulse responses to X_t at horizon h are heterogeneous (see also Remark 2). In our exposition, we always allow for time indicators as controls when s_i displays cross-sectional variation.

way clustering requires that any serial dependence is induced by micro shocks.

In this paper, we seek to understand the robustness of panel local projections to those features. Hence, one of our main contributions is to introduce a novel asymptotic framework where the macro–micro composition of the problem remains arbitrary in the limit. We achieve this by letting the joint distribution P_{TN} of observed and unobserved, macro and micro shocks drift with the sample size in an unrestricted manner, subject only to a minimum amount of identification and mild regularity conditions. For example, this device allows for data generating processes (DGPs) P_{TN} in a class $\mathcal{P}_{TN}$ that interpolates between settings where aggregate shocks explain a substantial part of the variation in the microdata and settings where they are almost imperceptible. It also accommodates arbitrary degrees of cross-sectional dependence and estimation errors that may be dominated by micro-level terms, a combination of micro and macro errors, or aggregate components only. On the contrary, standard asymptotic plans where DGPs are not indexed by the sample size fix the joint distribution of micro and macro shocks, missing these features and potentially leading to poor approximations in small samples.

Since the estimation error depends on P_{TN}, the natural question is which inference procedures (if any) are robust to different macro–micro regimes, thus dispensing practitioners from taking a particular stance on this structure. Our main result is that *t-LAHR inference is uniformly valid over $\mathcal{P}_{TN}$*, that is, t-LAHR confidence intervals have correct asymptotic coverage for the local projection estimand uniformly over DGPs in the class $\mathcal{P}_{TN}$.

To build intuition for this result, we show that panel local projections with macro shocks are equivalent to *synthetic* time series local projections with an appropriately aggregated dependent variable. Aggregating the microdata by collapsing the cross-sectional dimension of the panel and treating it as a time series yields valid inference for any P_{TN}. This is precisely what t-LAHR inference does, as time-level clustering in the panel problem is equivalent to heteroskedasticity-robust inference in the synthetic time series.

In fact, there is no need for more complicated procedures such as heteroskedasticity and autocorrelation robust (HAR) approaches or an additional clustering step within units: since X_t is a shock, including lags as controls renders the regression scores uncorrelated. Finally, we show that this also holds over long horizons when the DGP is well approximated by a low-order vector autoregression (VAR). That is, we prove that t-LAHR inference is also uniformly valid over $h \propto T$ within the VAR subclass of $\mathcal{P}_{TN}$, a result reminiscent of those in Montiel Olea and Plagborg-Møller (2021) for time series local projections.

Finite-sample performance. We complement our theoretical results with practical considerations, including guidance for lag length selection and finite-sample refinements (as in Imbens and Kolesár, 2016), and with simulation evidence showing an excellent finite-sample performance of t-LAHR inference compared to alternative methods across realistic designs, sample sizes and macro–micro regimes.

The finite-sample evidence matches the predictions of our asymptotic analysis. In particular, clustering only within units and two-way clustering — the two most common choices in empirical work — are not uniformly valid as they adjust for the wrong type of serial dependence (driven by idiosyncratic rather than aggregate shocks). In fact, clustering within units is superfluous, even in regimes where idiosyncratic shocks explain most of the variation in the microdata. Interestingly, t-HAR/HAC alternatives that substitute lag augmentation with HAR/HAC inference work in theory, since they are based on time-series inference on the synthetic scores (such as Driscoll-Kraay). In practice, they leave information on the table about the autocovariance function of the regression scores, struggle with realistic sample sizes and display severe coverage distortions. We also show that they inherit the well-documented finite-sample issues identified for HAR/HAC estimators in the time series literature (for a review, see Lazarus, Lewis, Stock, and Watson, 2018).

Revisiting the transmission of monetary policy to firm investment. We use our methods to provide a systematic reassessment of the role of financial frictions and firm heterogeneity in monetary policy transmission. Besides the traditional interest rate channel, recent contributions have emphasized a financial accelerator mechanism for more constrained firms using different measures of financial conditions (Ottonello and Winberry, 2020; Cloyne et al., 2023) and an asset price channel (Jeenas and Lagos, 2024). We revisit these three papers, which emphasize different mechanisms by showing that certain groups of firms are more responsive to exogenous changes in interest rates X_t. They use Compustat data to study comparable time periods from the 1990s to the 2010s but differ in the estimation methods (panel LP or LP-IV), choices of standard errors, nature and type of controls, and salience of aggregate shocks in the microdata. Our paper allows us to study all these specification choices and data environments within a unified framework.

We show that correctly accounting for the presence of an aggregate and possibly serially correlated component delivers several new empirical lessons. In particular, the robust t-LAHR method strengthens the case for a dynamic distance-to-default gradient, which suggests that monetary policy is less powerful for firms with higher default risk. Instead, and contrary to previous results, we do not find sufficient evidence for differential transmission for more constrained firms (in terms of age and dividend-paying status), for more levered firms and for those with higher stock turnover and lower liquidity.

Taken together, our findings reshape the picture of which firms are most responsive to monetary policy. Among the dimensions we revisit, distance to default emerges as the most empirically salient proxy for financial frictions. On the other hand, evidence for the leverage and credit-constraint margins of the financial accelerator and for a turnover-liquidity gradient remains elusive.

Related literature. Our paper contributes to various strands of the literature. First, it relates to the time series literature on local projection inference (Hansen and Hodrick, 1980; Jordà, 2005; Stock and Watson, 2018; Montiel Olea and Plagborg-Møller, 2021; Lusompa, 2023; Xu, forthcoming; Montiel Olea, Plagborg-Møller, Qian, and Wolf, forthcoming). Relative to this literature we are the first to deal with the panel data case with aggregate shocks.[4] We supplement the macroeconometric toolbox with results that are unique to the panel setting by accommodating heterogeneity in responses, spatial dependence and coexisting micro and macro shocks of any relative sizes. In a time series finite-order VAR setup, Montiel Olea and Plagborg-Møller (2021) show the uniform validity of heteroskedasticity-robust inference on lag-augmented local projections over the persistence of the data and horizon h. Our Proposition 2 can be interpreted as the panel version of their results. However, our focus is on uniformity with respect to macro–micro regimes, which has no obvious counterpart in time series, and we derive most of our results without assuming a VAR model.

Second, we contribute to the literature on estimation and inference with aggregate shocks and microdata. In stylized models, Hahn, Kuersteiner, and Mazzocco (2020) bring attention to the drastic consequences of drawing inferences from short panels with aggregate uncertainty. Chang, Chen, and Schorfheide (2024) propose a methodology based on functional vector autoregressions and repeated cross-sections to study the effects of aggregate shocks on cross-sectional distributions. We study the complementary problem of heterogeneity in transmission of aggregate shocks to unit-level outcomes via local projections.

[4] Our results on limited serial dependence in regression scores relate to the earlier multi-step forecast literature (Hansen and Hodrick, 1980), which relied on infinite lags to ensure forecast errors have an MA(h) representation. In the local projection context, Jordà (2005) arrived at a similar result under a finite-order VAR model while Lusompa (2023) provided a recent reformulation. Instead, we exploit the orthogonality properties of macro shocks to show that the *scores* have MA(h) dynamics. The distinction is reminiscent of the difference between design-based and model-based/conditional unconfoundedness assumptions.

An important empirical literature exploits regional-exposure designs with instruments of the form $s_i X_t$ used to identify the causal link between Y_{it} and a unit-level variable W_{it}, yielding a reduced-form equation identical to (1) for $h = 0$.[5] Recent work emphasizes the role of the exogeneity of X_t for credible identification (Adão, Kolesár, and Morales, 2019; Arkhangelsky and Korovkin, 2023; Majerovitz and Sastry, 2023). The broad empirical literature to which our paper speaks is related, but different in fundamental ways: panel local projections allow for dynamics, interpret aggregate shocks X_t in the macroeconometric sense, and aim at measuring their propagation via impulse responses. We add formal inference results in a rich panel environment with unrestricted heterogeneity, dynamics and different macro–micro regimes. In some cases, that literature has also relied on quasi-random variation in exposures s_i for identification. In our context, this would entail very strong requirements, analogous to finding a separate independent source of exogenous variation for every horizon of interest.[6] This is reminiscent of similar recent points in shift-share setups with heterogeneous treatment effects (Hahn, Kuersteiner, Santos, and Willigrod, 2024).

Last, our paper relates to the cross-sectional dependence literature that studies models where the scores feature spatial correlation (Driscoll and Kraay, 1998; Andrews, 2005; Pesaran, 2006; Gonçalves, 2011; Pakel, 2019). Our setting lies in the polar case where the shock of interest only varies over time, precluding solutions based on partialling out the common component from the regressors, as in Pesaran (2006). Our uniformity result (including robustness to any degree of spatial dependence) is new to this literature.

Outline. Section 2 motivates the problem and provides an overview of our results. This section is designed to be self-contained for readers interested in the practical aspects of our work. We summarize our recommended approach in Section 2.5. Section 3 presents our main inference result across general, heterogeneous dynamic models indexed by $P_{TN} \in \mathcal{P}_{TN}$. Section 4 discusses finite-sample performance and Section 5 the empirical applications. Proofs can be found in the (Supplemental) Appendix. MATLAB and R packages are available online at `https://github.com/TinchoAlmuzara/PanelLocalProjections`.

[5]Examples include regional fiscal multipliers (Nakamura and Steinsson, 2014) or the effects of foreign aid on conflict across countries (Nunn and Qian, 2014).

[6]Consider, for instance, the role of firm size (s_i) in mediating the transmission of monetary policy shocks to firm-level outcomes, as in Crouzet and Mehrotra (2020). For the cross-sectional dimension of the panel to help pin down the impulse response at, say, $h = 2$, firm size must be orthogonal to firm-level responses at any other horizon and to any other aggregate shock. This is, of course, hardly plausible. In fact, such exclusion restrictions would rule out dynamic effects over different horizons altogether — the object of empirical interest. We elaborate on this in Remark 2.

2 Overview of the results

Here we introduce the problem we study in this paper and outline our key results through the lens of a stylized model. We summarize our applied recommendations in Section 2.5, and leave the more technical econometric discussion for Section 3.

2.1 Setup

We observe an outcome Y_{it}, the macro shock of interest X_t, an attribute s_i and a vector of controls W_{it} for units $i = 1, \ldots, N$ and over periods $t = 1, \ldots, T$. Interest is in the dynamic propagation of aggregate shocks to individual outcomes, possibly allowing for heterogeneous responses across s_i. Specifically, for each horizon h, we fit the panel regression

$$Y_{i,t+h} = \hat{\beta}(h)s_i X_t + \hat{\psi}(h)' W_{it} + \hat{\mu}_i(h) + \hat{\nu}_t(h) + \hat{\xi}_{it}(h), \tag{2}$$

where $\hat{\mu}_i(h)$ and $\hat{\nu}_t(h)$ are unit and time fixed effects, and $\hat{\xi}_{it}(h)$ is the least squares residual. For clarity of exposition, we assume that Y_{it}, X_t and s_i are scalar. Although our results have natural multivariate versions, the scalar case covers most empirical applications. When $s_i = 1$, we omit time effects. This is a panel version of local projections (Jordà, 2005). Practitioners then report estimates $\hat{\beta}(h)$ together with a confidence interval of the form

$$\hat{C}_\alpha(h) = \left[\hat{\beta}(h) \pm z_{1-\alpha/2}\hat{\sigma}(h)\right], \tag{3}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution and $\hat{\sigma}(h)$ a standard error. Our goal is to assess this practice and provide guidance for estimation and inference.

We focus on two major questions. First, what is the interpretation of $\hat{\beta}(h)$ in a general panel model with cross-sectional heterogeneity? Second, which choices of $\hat{\sigma}(h)$ deliver valid confidence intervals? To address these questions, we study the properties of $\hat{\beta}(h)$ and of various choices of $\hat{\sigma}(h)$ in a broad class of data generating processes (DGPs) for Y_{it}, X_t, s_i and W_{it} that allows for both macro and micro shocks and rich forms of heterogeneity and dynamic transmission, including cases where the identification content of macro data is small. We develop our analysis in depth in Section 3; before, we discuss the main elements of the framework and illustrate our main results through the lens of a stylized model.

Shocks and identification. An important feature of the problem is that X_t is a shock, that is, it is unpredictable given its own history and given other macro and micro innovations.

This notion — made precise in Assumption 1 below — is a key identifying condition without which $\hat{\beta}(h)$ cannot be interpreted as estimating a dynamic causal effect (Stock and Watson, 2018, Section 1.2). Empirical researchers widely recognize this and, in practice, put great effort in constructing X_t by leveraging a variety of macroeconometric methods.

In that sense, our analysis covers three different settings that span the main macroeconometric shock identification procedures, and whose use in the panel context entails no conceptual difference. The first setting is one where X_t is directly observed, typically constructed from narrative or high-frequency methods, and which covers most empirical applications. The second setting is one where X_t can be recovered from a macro system via standard identification restrictions, e.g., recursive, short-run, long-run, or heteroskedasticity-based identification (see Ramey, 2016, for a review). The third setting is one where the researcher only observes the proxy $X_t^* = X_t + \nu_t$, that is, a measure of X_t contaminated with measurement error. In that case, it is more appropriate to think of X_t^* as an instrument in a panel version of local projection-instrumental variables (LP-IV) with outcome $Y_{i,t+h}$ (see Stock and Watson, 2018, for a review in the time series context). Our empirical applications in Section 5 are representative of these different settings.[7]

2.2 A stylized model

To sketch the key ideas of the paper, we now turn to a simple stylized model, which nonetheless preserves many of the main features of the framework. Suppose the micro outcome Y_{it} is related to the macro shock of interest X_t by

$$Y_{it} = \rho Y_{i,t-1} + \beta_i X_t + Z_t + u_{it}, \qquad |\rho| < 1, \tag{4}$$

where the terms Z_t and u_{it} capture, respectively, unobservable aggregate and unit-level shocks orthogonal to the shock of interest X_t. More precisely, X_t, Z_t and u_{it} are mutually and serially mean independent (Assumption 1). We allow for u_{it} to be spatially correlated. We also observe a covariate s_i, possibly correlated with β_i. Under suitable initial conditions,

[7] Specifically, we revisit Ottonello and Winberry (2020) and Jeenas and Lagos (2024), which treat X_t as observable and Cloyne et al. (2023), which reports panel LP-IV estimates. Similarly, an example where X_t is recovered from a macro system is Drechsel (2023), who imposes long-run restrictions on a structural VAR model to identify investment-specific technology shocks.

we can write (4) as

$$Y_{it} = \sum_{\ell=0}^{\infty} \rho^{\ell} \beta_i X_{t-\ell} + \sum_{\ell=0}^{\infty} \rho^{\ell} (Z_{t-\ell} + u_{i,t-\ell}), \tag{5}$$

so the model allows for persistence and dynamics in the transmission of both observed and unobserved, macro and micro shocks to unit-level outcomes Y_{it}. This macro–micro structure is empirically realistic and plays a key role for inference. For example, in applications that study the response of firm-level investment Y_{it} to monetary policy shocks X_t, the same data often contain other aggregate disturbances, such as TFP or oil supply shocks, captured here by Z_t. Similarly, a sizable share of the variation in Y_{it} reflects shocks u_{it} that operate at the unit level, possibly exhibiting cross-sectional dependence through sectoral or regional linkages. As we show below, the coexistence of aggregate and unit-level components creates a nonstandard macro–micro problem that is not adequately addressed by conventional inference methods.

2.3 Estimand

The fact that X_t is a shock allows us to interpret the coefficient on X_t and its lags in (5) as dynamic causal effects for unit i. This resembles a core insight from macroeconometrics that links causal interpretations to the availability of unpredictable variation (Leeper et al., 1996; Ramey, 2016; Stock and Watson, 2016, 2018; Plagborg-Møller and Wolf, 2021). Concretely, writing E_i for expectations conditional on β_i,

$$\beta_{ih} \equiv E_i[\, Y_{it} \mid X_{t-h} = 1\,] - E_i[\, Y_{it} \mid X_{t-h} = 0\,] = \rho^h \beta_i.$$

Summaries of the distribution of β_{ih} are objects of primary interest in applications. A central question is then how to interpret the estimator $\hat{\beta}(h)$ in a panel data context in terms of the underlying heterogeneous dynamic causal effects β_{ih}. We show that $\hat{\beta}(h)$ converges in probability to the following population object:

$$\beta(h) = \frac{\mathrm{Cov}(\, s_i, \beta_{ih}\,)}{\mathrm{Var}(\, s_i\,)}.$$

This has a simple interpretation: it is the population coefficient in an (infeasible) projection of the unobserved, unit-level response β_{ih} on s_i, including an intercept. In other words, panel local projections recover the best linear approximation to how the dynamic causal effects β_{ih}

vary, in expectation, with the observable attribute s_i. In particular,

$$\beta(h) \approx E[\,\beta_{ih} \mid s_i = \bar{s} + 1\,] - E[\,\beta_{ih} \mid s_i = \bar{s}\,]\,,$$

for a one-unit change in $s_i = \bar{s}$, and with equality when $E[\,\beta_{ih} \mid s_i\,]$ is linear in s_i. Thus, $\hat{\beta}(h)$ is informative about differential transmission of X_t for units with higher and lower values of s_i. This formalizes what practitioners have in mind when including interactions in (2), where s_i is chosen to capture an economically relevant dimension along which responses are heterogeneous. The same interpretation applies to both continuous and discrete attributes s_i — for instance, if $s_i = \mathbf{1}\{i \in G\}$, $\beta(h)$ is the average response of group G relative to the omitted group. Both cases appear frequently in applications.[8] Finally, when $s_i = 1$ for all i, the estimand becomes $\beta(h) = E[\,\beta_{ih}\,]$, the average impulse response.

Importantly, this characterization is not specific to the stylized model considered here and applies in the general framework of Section 3. In particular, the same interpretation carries over to setups where X_t is recovered from a macro system, to instrumental variables panel local projections estimators, and to alternative specifications of the regression such as including lagged outcomes $Y_{i,t-1}$ or controls orthogonal to the shock X_t. These controls affect the stochastic properties of the residual in (2), but not the population object to which $\hat{\beta}(h)$ converges. As we discuss next, while the estimand is invariant to the choice of controls, the behavior of standard errors is not.

2.4 Robust inference

A key result of our paper is that a simple strategy guarantees correct and robust inference in a wide range of situations where popular approaches fail. It consists of two ingredients: estimation including lags of the outcome Y_{it} and regressor of interest $s_i X_t$, and clustering the standard errors at the time level. Intuitively, clustering isolates the aggregate component of the regression score, while controlling for lags absorbs aggregate serial dependence.

In other words, in (2) we control for $W_{it} = (s_i X_{t-1}, Y_{i,t-1}, \ldots, s_i X_{t-p}, Y_{i,t-p}, \tilde{W}_{it}')'$, where lag length p is chosen as described in Section 2.5, and in (3) we construct $\hat{\sigma}(h)$ by time-level clustering. In our treatment, we also allow for other (macro and micro) controls through $\tilde{W}_{it}$. These are common in practice; for instance, in our empirical application to the sensitivity

[8] For example, Cloyne et al. (2023) study heterogeneity in the response of firm investment to monetary policy across groups defined by firm characteristics such as age or dividend-paying status, whereas Ottonello and Winberry (2020) use continuous measures of credit risk such as leverage or distance to default.

of firm investment to monetary policy, lagged micro indicators such as firm sales and lagged macro indicators such as GDP growth (interacted with s_i) are routine. By contrast, including lags of $s_i X_t$ and Y_{it} is much less common, but it turns out that this is the crucial step to obtain valid and informative confidence intervals.

We call this strategy time-clustered lag-augmented heteroskedasticity-robust (t-LAHR) inference. In contrast to other popular alternatives in applied practice, we argue below that from both a theoretical and practical standpoint, t-LAHR inference delivers valid inference regardless of the macro–micro composition of the error term; that is, regardless of whether micro or macro variation dominates the estimation error.

Robustness to the macro–micro structure of the errors. To illustrate the relevance of our macro–micro structure and the need for inference procedures that remain valid in any such regime, consider the cross-sectionally aggregated version of our stylized model (5):

$$\bar{Y}_t = \bar{\beta} X_t + \sum_{\ell>0}^{\infty} \rho^{\ell} \bar{\beta} X_{t-\ell} + \sum_{\ell=0}^{\infty} \rho^{\ell} (Z_{t-\ell} + \bar{U}_{t-\ell}), \tag{6}$$

where $\bar{Y}_t = \frac{1}{N}\sum_{i=1}^{N} Y_{it}$, $\bar{\beta} = \frac{1}{N}\sum_{i=1}^{N} \beta_i$ and $\bar{U}_t = \frac{1}{N}\sum_{i=1}^{N} u_{it}$. A key observation in our setting is that there is a tight connection between our underlying panel data problem and this (synthetic) time series version, which is made precise in Remark 1 of Section 3. For the purposes of this section, it suffices to note that we can equivalently use the (synthetic) model in (6) to understand the sources of estimation uncertainty. For example, the population errors from a regression on X_t (setting $s_i = 1$ and $h = 0$ for simplicity) would be given by $\bar{Y}_t - \bar{\beta} X_t$ in (6). It is then immediate that the relative sizes of its components regulate the nature of serial and cross-sectional dependence in the errors and the strength of identification, thereby directly dictating the validity of different choices of standard errors.

To illustrate the idea, let $\sigma_X^2 = \mathrm{Var}(X_t)$, $\sigma_Z^2 = \mathrm{Var}(Z_t)$ and $\sigma_{\bar{U}}^2 = \mathrm{Var}\left(\bar{U}_t\right)$. Now, consider a hypothetical R^2 capturing relative importance of macro shocks in the error term,

$$\bar{R}^2 \equiv 1 - \frac{\mathrm{Var}\left(\sum_{\ell=0}^{\infty} \rho^{\ell} \bar{U}_{t-\ell}\right)}{\mathrm{Var}\left(\bar{Y}_t - \bar{\beta} X_t\right)} = 1 - \frac{\sigma_{\bar{U}}^2}{\rho^2 \bar{\beta}^2 \sigma_X^2 + \sigma_Z^2 + \sigma_{\bar{U}}^2}, \tag{7}$$

which we also target in our simulations. The relative sizes of these parameters potentially imply very different inferential regimes. For example, a relatively large $\sigma_{\bar{U}}$ captures regimes where idiosyncratic shocks might dominate the estimation error ($\bar{R}^2 \approx 0$). These quantities are essential because the reliability of conventional standard errors depends on what the dominant source of variation is in a given problem.[9] Indeed, the ratio of unit-level to two-way clustered standard errors provides a (crude) approximation to $1-\bar{R}^2$. We find that this ratio varies greatly between 0 and 1 in our empirical applications in Section 5.

In our general formulation in Section 3, we consider a range of DGPs indexed by the joint distribution of $\{X_t, Z_t, \{u_{it}\}_{i=1}^N\}_{t=-\infty}^{\infty}$ (and the heterogeneity) $P \in \mathcal{P}$, where $\mathcal{P}$ is a rich class, and build this into our asymptotic approximations. For example, $\mathcal{P}$ allows for any $\bar{R}^2(P) \in [0,1]$ in the limit. Other properties of the data, such as the relative signal of X_t — that is, $\mathrm{Var}(X_t)/\mathrm{Var}(\bar{Y}_t)$ — are also allowed to vary, subject only to a minimum requirement ruling out weak identification. The natural question is whether any standard error is robust to these features, including an arbitrary macro–micro error composition. At a theoretical and practical level, we will see that only t-LAHR achieves such robustness.

Simulation evidence. We now return to the practical problem of doing inference about $\beta(h)$. We seek to construct a standard error $\hat{\sigma}(h)$ to obtain a confidence interval (3) with reliable coverage, reasonable length, and robustness to empirically relevant features of the DGP, such as the relative size of aggregate shocks or the form of cross-sectional dependence. Our main result is that the t-LAHR standard error in (11) achieves validity, informativeness and robustness, whereas other commonly used choices do not.

To illustrate the point, we simulate data from the stylized model (4). The experiment sets $T = 30$, in line with the modal time series dimension in our survey in Appendix F. We also fix $N = 1,500$ and $\rho = 0.7$. We take i.i.d. draws from $X_t \sim \mathcal{N}(0,1)$, $Z_t \sim \mathcal{N}(0,1)$, $\beta_i \sim \mathcal{N}(1,1)$ and $u_{it} \sim \mathcal{N}(0,\sigma_u^2)$, where σ_u^2 is calibrated to match $\bar{R}^2 \in \{0.1, 0.9\}$, the observed range of this statistic in our empirical applications in Section 5, as proxied by the ratio of one-way to two-way clustered standard errors.[10] For simplicity, we also set $s_i = \beta_i$, which yields the closed-form estimand $\beta(h) = \rho^h$.

Table 1 reports coverage rates for 90% confidence intervals obtained by different standard error choices based on $5,000$ simulated samples and over $h = \{0,1,2,5\}$. For illustration purposes, we present two t-LAHR versions: one which uses $p = 1$, and another (labeled as recommended) which applies the lag selection rule and small-sample refinement given in

[9]Moreover, in a previous version of this paper (Almuzara and Sancibrián, 2025, Section 2), we show explicitly how an unrestricted σ_U/σ_X (or σ_U/σ_Z) implies a discontinuity in the asymptotic distribution of the panel local projection estimator.

[10]In the present design, where u_{it} is i.i.d., we have $\sigma_u^2 = N(1-\bar{R}^2)/\bar{R}^2$, up to ρ and β_i. For cross-sectionally dependent u_{it}, $\bar{R}^2$ would be defined over the purely idiosyncratic component of u_{it} only. We keep it intentionally informal in this section for simplicity.

Section 2.5. We also report coverage-adjusted width averaged across h. This is a measure of statistical power that is computed rescaling critical values so as to eliminate potential coverage distortions, putting the procedures on a more comparable basis (see, for example, Lazarus, Lewis, and Stock, 2021).

TABLE 1. Performance of standard errors in the stylized model

	Coverage rate of 90% CI				Coverage-adjusted width
Standard error	$h=0$	$h=1$	$h=2$	$h=5$	average
$\bar{R}^2 = 0.9$					
Unit-level clustered	0.494	0.372	0.344	0.376	0.938
Two-way clustered	0.819	0.726	0.725	0.811	0.952
Driscoll–Kraay	0.716	0.725	0.695	0.782	1.007
t-LAHR (plain)	0.851	0.842	0.811	0.806	0.754
t-LAHR (recommended)	0.903	0.899	0.872	0.875	0.754
$\bar{R}^2 = 0.1$					
Unit-level clustered	0.888	0.878	0.866	0.880	2.583
Two-way clustered	0.841	0.829	0.814	0.820	2.823
Driscoll–Kraay	0.809	0.791	0.789	0.810	2.735
t-LAHR (plain)	0.848	0.849	0.836	0.835	2.595
t-LAHR (recommended)	0.903	0.902	0.891	0.895	2.615

Note: Results are based on 5,000 simulation runs. Unit-level, two-way and Driscoll-Kraay use $p = 0$, whereas plain t-LAHR uses $p = 1$. The row labeled `t-LAHR (recommended)` corresponds to the procedure discussed in Section 2.5. Coverage-adjusted width is the width of confidence intervals rescaled to attain correct coverage by modifying their critical value, averaged over $h = 0, 1, 2, 5$.

In terms of coverage, we highlight three findings. First, the recommended t-LAHR approach has coverage rates very close to 90% across h, with minimal distortions despite the small time series dimension. Lag selection and small-T refinements (which are straightforward to implement) deliver meaningful improvements, but even without them t-LAHR generally outperforms alternatives. Second, two-way clustering and Driscoll-Kraay display severe distortions, undercovering $\beta(h)$ by about 10 percentage points (pp), compared to 5 and 1 pp for plain and refined t-LAHR. In fact, Driscoll-Kraay shortfalls reach up to 20 pp, often doing worse than two-way clustering, which ignores (common) serial correlation altogether. Third, unit-level clustering is fragile. When $\bar{R}^2$ is low, its coverage is only moderately below nominal and close to that of unrefined t-LAHR, but when $\bar{R}^2$ is high, it shows severe undercoverage. This sensitivity is problematic since $\bar{R}^2$ is generally hard to know ex-ante in a given dataset. In contrast, t-LAHR performs well across low- and high-$\bar{R}^2$ designs.

These findings align with the large-sample theory of Section 3, which delivers approximations that are *uniformly valid* over classes of DGPs with different macro–micro error compositions, thus formalizing the idea of robustness (or lack of) to $\bar{R}^2$ illustrated in Table 1. Specifically, for confidence intervals $\hat{C}_\alpha(h)$ that use the t-LAHR standard error, we show in Proposition 1 that for a rich class $\mathcal{P}$ of DGPs and for $h \leq p$,

$$\sup_{P \in \mathcal{P}} \left| P\Big(\beta(h) \in \hat{C}_\alpha(h)\Big) - (1-\alpha) \right| \longrightarrow 0 \text{ as } T, N \to \infty, \tag{8}$$

that is, the worst-case discrepancy across DGPs between the realized and nominal coverage rate of $\hat{C}_\alpha(h)$ vanishes in large samples. This is a strong sense of robustness that extends to moderately long horizons $h \propto T$ (with additional structure) and to instrumental variable problems (Propositions 2 and 3). Instead, common alternatives in applied practice such as unit-level and two-way clustered inference do not enjoy the uniform guarantee (8) — in fact, the worst-case coverage of the former is close to zero for high $\bar{R}^2$. Driscoll–Kraay may be valid at short horizons, but performs poorly in the short time series typical of macro applications due to the difficulty of long-run variance estimation. Its performance also deteriorates with even moderate persistence (of macro or micro origin) in the data.

Turning to the informativeness of alternative approaches, the main takeaway is that the superior coverage properties of t-LAHR do not imply a sacrifice of statistical power. In Table 1, coverage-adjusted t-LAHR intervals are between 5% and 25% shorter than Driscoll–Kraay intervals. This is consistent with our theory: although the relative power of these methods can depend on the DGP, positive persistence — a realistic feature of macro and micro data — tends to favor t-LAHR. Unit-level clustered intervals may appear short when $\bar{R}^2$ is high, but only because they undercover severely; after size adjustment, they are not more informative than t-LAHR. Two-way clustering is also relatively inefficient by this metric.

Generality of the results. One might wonder how much the stylized nature of (4) limits these results. The rest of the paper shows that they extend to a substantially more general framework — allowing for rich forms of heterogeneity, dynamics, serial correlation, cross-sectional dependence, controls and identification strategy. Section 3 offers large-sample approximations and Section 4 presents finite-sample evidence across a wide range of designs.

2.5 Step-by-step implementation

We conclude by summarizing our recommended implementation of t-LAHR inference:

1. Select a candidate number of lags $\hat{p}^{AIC}$ using the Akaike Information Criterion on the

synthetic time series[11] and apply the following horizon-dependent rule:

$$\hat{p}(h) = \max\{1, \min(h, \hat{p}^{AIC})\}.$$

If the shock X_t is retrieved with error from a macro system, consider the inclusion of additional lags as discussed in Section 3.3.

2. Estimate panel local projections on $s_i X_t$ at horizon h including $\hat{p}(h)$ lags of $s_i X_t$ and Y_{it}. Optional: compute the synthetic time-series (see Remark 1 and Figure 3) as a visual, low-dimensional representation of the estimation problem.

3. Cluster residuals at the time level. Alternatively, aggregate the regression scores to the time level and compute heteroskedasticity-robust standard errors.

4. If T is small, report confidence intervals as in (3) using the refinements advocated by Imbens and Kolesár (2016) for heteroskedasticity-robust inference.

All these steps are implemented in our own Matlab and R packages available online at `https://github.com/TinchoAlmuzara/PanelLocalProjections`. Additional information and discussion on these implementation details is given in Section 4 and in Appendix B.

3 Econometric framework

This section contains the econometric results of the paper in a general setting. Section 3.1 introduces the class of data generating processes (DGPs) we consider. Section 3.2 presents our main estimand and inference results when the shock of interest X_t is observed. We then extend the analysis to settings in which X_t is identified from a structural VAR (Section 3.3) or used as a proxy in panel LP-IV (Section 3.4). All proofs are in Appendix A.

The material in this section is more technical than the rest of the paper. Readers interested primarily in the interpretation and implementation of the results may refer to Sections 2.3–2.5, which provide an overview of the main ideas and practical recommendations.

[11] Specifically, let $\hat{\xi}_{\text{VAR},it}(p)$ denote the residuals on a panel VAR(p) and define $\hat{s}_i = s_i - \frac{1}{N}\sum_{i=1}^{N} s_i$. Then, for some upper bound $\bar{p}$, let

$$\hat{p}^{AIC} = \underset{0 \le p \le \bar{p}}{\operatorname{argmin}} \log \hat{\sigma}^2_{\text{VAR}}(p) + \frac{2p}{T},$$

where $\hat{\sigma}^2_{\text{VAR}}(p) = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{\xi}_{\text{VAR},t}(p)^2$ and $\hat{\xi}_{\text{VAR},t}(p) = (\sum_{i=1}^{N} \hat{s}_i^2)^{-1} \sum_{i=1}^{N} \hat{s}_i \hat{\xi}_{\text{VAR},it}(p)$. Note that, contrary to a standard implementation in a panel data problem, we use the synthetic residuals $\hat{\xi}_{\text{VAR},t}(p)$ and a penalty proportional to T^{-1} (rather than $(NT)^{-1}$).

3.1 Model and assumptions

We observe an outcome Y_{it}, the shock of interest X_t, an attribute s_i, and controls W_{it} for units $i = 1, \ldots, N$ and periods $t = 1, \ldots, T$. In this subsection and the next, we treat X_t as directly observed—we study cases where X_t is not observed in Sections 3.3 and 3.4. The controls take the form $W_{it} = (s_i X_{t-1}, Y_{i,t-1}, \ldots, s_i X_{t-p}, Y_{i,t-p}, \tilde{W}'_{it})'$ where $\tilde{W}_{it}$ contains a fixed number d of additional macro and micro controls. We model Y_{it} and $\tilde{W}_{it}$ as

$$\begin{aligned} Y_{it} &= \mu_i + \sum_{\ell=0}^{\infty} \beta_{i\ell} X_{t-\ell} + \sum_{\ell=0}^{\infty} (\gamma'_{i\ell} Z_{t-\ell} + \delta'_{i\ell} u_{i,t-\ell}), \\ \tilde{W}_{it} &= \mu_{w,i} + \sum_{\ell=1}^{\infty} \beta_{w,i\ell} X_{t-\ell} + \sum_{\ell=0}^{\infty} (\gamma'_{w,i\ell} Z_{t-\ell} + \delta'_{w,i\ell} u_{i,t-\ell}). \end{aligned} \tag{9}$$

Here Z_t and u_{it} are vectors of dimension d_Z and d_u that collect unobserved aggregate and micro-level shocks orthogonal to X_t. The first equation allows Y_{it} to respond dynamically to the shock of interest X_t and to other aggregate and micro disturbances. The second imposes that the additional controls $\tilde{W}_{it}$ are predetermined with respect to X_t, while allowing them to have arbitrary exposures to lagged X_t and to the histories of Z_t and u_{it}.

The coefficients in (9) allow the transmission of all shocks to vary across units. We collect all unit-level objects into $\theta_i = (\mu_i, \mu_{w,i}, \{\beta_{i\ell}, \beta_{w,i\ell}, \gamma_{i\ell}, \gamma_{w,i\ell}, \delta_{i\ell}, \delta_{w,i\ell}\}_{\ell \geq 0})$ with the convention that $\beta_{w,i0} = 0$. Importantly, we leave the joint distribution of the observed attribute s_i and the unobserved heterogeneity θ_i largely unrestricted. Thus, s_i need not summarize all heterogeneity in responses. Rather, the purpose of panel local projections is to estimate how the heterogeneous impulse responses β_{ih} vary, on average, with s_i.

The representation in (9) accommodates rich dynamics. The infinite-order moving averages may arise, for example, as Wold representations of more primitive stationary macro and micro state variables, including heterogeneous versions of familiar time-series models such as VARs. Consequently, Y_{it} may display substantial persistence even though the primitive shocks X_t, Z_t, and u_{it} are serially unpredictable. The specification of $\tilde{W}_{it}$ also covers both pure micro controls and aggregate controls interacted with unit-level attributes.

DGPs. Let P denote the joint distribution of the shock processes $\{X_t, Z_t, \{u_{it}\}_{i=1}^N\}_{t=-\infty}^{\infty}$ and cross-sectional heterogeneity $\{s_i, \theta_i\}_{i=1}^N$. Through (9), P determines the distribution of the observed sample. We let $\mathcal{P}$ denote the class of distributions satisfying the shock and regularity conditions stated below.[12] The class $\mathcal{P}$ formalizes the dimensions of robustness we wish to study: it allows the relative size of aggregate and micro-level shocks, the form of time series and cross-sectional dependence, and the heterogeneity to vary across P. We therefore study the large-sample behavior of $\hat{\beta}(h)$, $\hat{\sigma}(h)$, and $\hat{C}_\alpha(h)$ *uniformly* over $P \in \mathcal{P}$.

Our main identifying assumption is that, conditional on the heterogeneity, X_t is a shock relative to all other primitive innovations. Throughout this section, expectations (and other operators) are computed under a fixed $P \in \mathcal{P}$, which we indicate by writing E_P.

Assumption 1 (shocks and mean independence). *Under each distribution $P \in \mathcal{P}$, the aggregate and micro-level shocks $\{X_t, Z_t, \{u_{it}\}_{i=1}^N\}_{t=-\infty}^\infty$ satisfy*

$$E_P\Big[X_t \;\Big|\; \{X_\tau\}_{\tau \neq t}, \{Z_\tau, \{u_{i\tau}\}_{i=1}^N\}_{\tau=-\infty}^\infty, \{s_i, \theta_i\}_{i=1}^N \Big] = 0, \tag{10}$$

and

$$E_P\Big[Z_t \;\Big|\; \{Z_\tau\}_{\tau \neq t}, \{X_\tau, \{u_{i\tau}\}_{i=1}^N\}_{\tau=-\infty}^\infty, \{s_i, \theta_i\}_{i=1}^N \Big] = 0_{d_Z \times 1},$$
$$E_P\Big[u_{it} \;\Big|\; \{\{u_{j\tau}\}_{j=1}^N\}_{\tau \neq t}, \{X_\tau, Z_\tau\}_{\tau=-\infty}^\infty, \{s_j, \theta_j\}_{j=1}^N \Big] = 0_{d_u \times 1}, \text{ for all } i.$$

Condition (10) is the key identifying restriction. It formalizes the notion, standard in macroeconometrics, that impulse responses have a structural interpretation only when the variable of interest is an unpredictable innovation orthogonal to other shocks (e.g., Leeper et al., 1996; Ramey, 2016; Stock and Watson, 2016; Plagborg-Møller and Wolf, 2021). The formulation imposes mean independence with respect to leads and lags. This is stronger than a martingale-difference condition but weaker than iidness, since it still permits dependence in higher moments.[13] For symmetry, Assumption 1 imposes analogous restrictions on Z_t and u_{it}. These restrictions apply to the primitive innovations, and not to the observed regression objects — controls and regression residuals — which may still exhibit rich autocorrelation.

Each $P \in \mathcal{P}$ also satisfies the regularity conditions stated in Appendix A.1 (Assumptions A.1, A.2 and A.3). These conditions impose limited time series persistence (ruling out unit roots), bounded moments and cumulants, summable dynamic responses, and nondegenerate

[12] Formally, as in standard double-array asymptotics, both the class $\mathcal{P} = \mathcal{P}_{TN}$ and its elements $P = P_{TN}$ may depend on (N, T), while the constants appearing in the moment, summability, and nondegeneracy conditions in Assumption A.3 are fixed across sample sizes.

[13] This type of lead-lag mean independence is convenient in local-projection settings, where the regression residual may contain both leads and lags of primitive innovations; see, for example, Montiel Olea and Plagborg-Møller (2021). It is also analogous to strict exogeneity conditions in panel data models.

regressors and regression scores, ensuring a well-defined inference problem.[14]

We also leave other features of the DGP largely unrestricted, including the relative variances of all elements in (9), the extent of cross-sectional dependence in u_{it} (ranging from purely idiosyncratic to strongly dependent shocks), the size and nature of individual-level responses and the macro–micro composition of the error term. As a result, several interesting summaries of the distribution of the data are allowed to vary freely across DGPs, such as the relative importance of aggregate shocks in the microdata, which we denoted with $\bar{R}^2(P)$ in Section 2. Our estimation and inference results are uniform over this class.

3.2 Estimand and uniform inference

We now state the main econometric result of the paper. We aim to answer two questions. First, how should the coefficient of a panel local projection on an aggregate shock be interpreted as a population object? Second, how can one construct confidence intervals that remain valid under empirically relevant forms of aggregate and unit-level variation? Recall that the t-LAHR standard error is based on time clustering of the regression scores $\hat{x}_{it}(h)\hat{\xi}_{it}(h)$, which corresponds to

$$\hat{\sigma}(h) = \sqrt{\frac{\hat{V}(h)}{(T-h)\hat{J}(h)^2}}, \quad \hat{V}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{x}_{it}(h)\hat{\xi}_{it}(h)\right)^2, \tag{11}$$

where $\hat{\xi}_{it}(h)$ are the panel local projection residuals in (2) and $\hat{J}(h) = \sum_{t=1}^{T-h}\sum_{i=1}^{N}\hat{x}_{it}(h)^2/[N(T-h)]$ with $\hat{x}_{it}(h)$ the residual from regressing $s_i X_t$ on W_{it} with unit and time fixed effects. Proposition 1 answers both questions by characterizing the large-sample behavior of $\hat{\beta}(h)$, $\hat{\sigma}(h)$, and $\hat{C}_\alpha(h)$ uniformly over $P \in \mathcal{P}$. Throughout, we let $N, T \to \infty$ and take the number of controls d and the lag length p as fixed.[15]

Proposition 1. *Let Assumptions 1, A.1, A.2 and A.3 hold and fix $h \le p$. For all $\epsilon > 0$,*

$$\lim_{T,N\to\infty}\sup_{P\in\mathcal{P}} P\left(\left|\hat{\beta}(h) - \beta(h)\right| > \epsilon\right) = 0.$$

[14] That is, for every $P \in \mathcal{P}$ we require that $\mathrm{Var}_P(X_t)$ and the variance of the estimation error are bounded away from zero and infinity, thus ruling out weak identification in the limit. Nonetheless, t-LAHR is still uniformly valid in the sense of Equation (12) even without this requirement; see Almuzara and Sancibrián (2025). Naturally, this comes at the expense of uniform convergence in probability.

[15] A natural question is whether in our general linear heterogeneous model (9) one can accommodate $h \to \infty$ by allowing p to grow with T. The analysis in Xu (forthcoming) and our simulation study in Section 4 suggest that this may be possible, but we leave this important issue for future research.

Moreover, for the t-LAHR confidence interval $\hat{C}_\alpha(h)$ *in* (3)*, for all* $0 < \alpha < 1/2$*,*

$$\lim_{T,N\to\infty} \sup_{P\in\mathcal{P}} \left| P\Big(\beta(h) \in \hat{C}_\alpha(h) \Big) - (1-\alpha) \right| = 0. \tag{12}$$

The first part of Proposition 1 shows that $\hat{\beta}(h)$ converges uniformly in probability to $\beta(h) = \mathrm{Cov}_P(\, s_i, \beta_{ih}\,) / \mathrm{Var}_P(\, s_i\,)$. Thus, even with unrestricted heterogeneity in unit-level responses, the panel local projection coefficient has a simple nonparametric interpretation: it is the slope of the best linear approximation to the heterogeneous dynamic effects β_{ih} as a function of s_i. As discussed in Section 2.3, this object measures differential transmission of the shock X_t between units with high and low values of the observed attribute s_i.

The second part of Proposition 1 establishes the uniform validity of the t-LAHR confidence interval for the estimand $\beta(h)$. Uniformity is important because it implies robustness to features that are difficult to know ex ante, such as the relative size of aggregate and unit-level shocks, summarized by $\bar{R}^2(P)$, and the strength of cross-sectional dependence in u_{it}. This is the formal counterpart to the robustness documented in Table 1 and explored further in the simulations of Section 4.

Where is the validity of t-LAHR coming from? The relevant object for inference is the aggregated regression score $X_t \sum_{i=1}^N \hat{s}_i \xi_{it}(h,P)$, where $\hat{s}_i$ is the demeaned s_i and $\xi_{it}(h,P)$ denotes the population regression residual. When $p \geq h$, the scores are serially uncorrelated, even though the residual itself may be correlated. Lag augmentation therefore transforms a problem with potentially persistent residuals into one where Eicker–Huber–White inference applied to time-clustered scores suffices, which links t-LAHR inference in panel local projections to heteroskedasticity-robust inference in a synthetic time-series local projection (see also Remark 1). Moreover, the aggregated score remains uncorrelated even with $\bar{R}^2(P) \approx 0$ and u_{it} independent over i. That is, t-LAHR still delivers correct inference even when time-clustering is not needed; in this sense, our uniformity result follows.

By contrast, unit-level and (non-lag-augmented) two-way clustered standard errors are not uniformly valid. Unit-level clustering can have worst-case coverage close to zero when the aggregate component dominates the regression score, while two-way clustering does not address the serial correlation of the score. Time-clustered HAC or HAR inference (t-HAR) provides a valid alternative in principle: analogues of Proposition 1 can be obtained (with slightly stronger conditions) by replacing $\hat{V}(h)$ in (11) with Hansen–Hodrick or Newey–West/Driscoll–Kraay long-run variance estimators. In finite samples, however, Section 4 shows that t-LAHR typically delivers better coverage and shorter coverage-adjusted intervals.

Remark 1 (Synthetic time series). A useful intuition for panel local projections is the following representation. The residual $\hat{x}_{it}(h)$ can be written as $\hat{x}_{it}(h) = \hat{s}_i(X_t - \bar{X}_0(h)) - \hat{\pi}(h)'W_{it} \approx \hat{s}_i X_t$ where $\hat{s}_i$ is the demeaned s_i, $\bar{X}_0(h)$ is the sample mean of X_t and $\hat{\pi}(h)$ is the coefficient on W_{it}, which then gives

$$\hat{\beta}(h) = \frac{\sum_{t=1}^{T-h} \sum_{i=1}^{N} \hat{x}_{it}(h) Y_{i,t+h}}{\sum_{t=1}^{T-h} \sum_{i=1}^{N} \hat{x}_{it}(h)^2} \approx \frac{\sum_{t=1}^{T-h} X_t \hat{Y}_{t+h}}{\sum_{t=1}^{T-h} X_t^2},$$

that is, the time series local projection estimator that regresses cross-sectional coefficients $\hat{Y}_{t+h} = (\sum_{i=1}^{N} \hat{s}_i^2)^{-1} (\sum_{i=1}^{N} \hat{s}_i Y_{i,t+h})$ on X_t controlling for $1, X_{t-1}, \ldots, X_{t-p}$.[16] The standard error $\hat{\sigma}(h)$ in (11) is also the Eicker–Huber–White standard error calculated on the time series LP residuals $\hat{\xi}_t(h) = (\sum_{i=1}^{N} \hat{s}_i^2)^{-1} \sum_{i=1}^{N} \hat{s}_i \hat{\xi}_{it}(h)$. Hence, panel t-LAHR inference and time-series lag-augmented heteroskedasticity-robust inference are intimately related.

Remark 2 (s_i and precision). The synthetic time-series representation also clarifies why the effective sample size is often governed by T, not by NT. Faster rates would require s_i to isolate variation in β_{ih} while being orthogonal to all other exposures to aggregate shocks. In that sense, s_i would need to act like a cross-sectional instrument that is relevant for the horizon-h response but excluded from the responses to all other aggregate innovations and horizons. These conditions are strong and are ruled out by Assumption A.3(iii). This suggests caution regarding the conventional wisdom in many empirical applications that a larger cross-sectional dimension somehow compensates for a shorter time series.

Remark 3 (Time-varying responses). Some applications study whether responses differ across units with different values of a time-varying covariate s_{it} that is exogenous to X_t. The same logic applies after extending (9) to allow for time-varying impulse responses,

$$Y_{it} = \mu_i + \sum_{\ell=0}^{\infty} \beta_{it\ell} X_{t-\ell} + \sum_{\ell=0}^{\infty} \gamma_{it\ell} Z_{t-\ell} + \sum_{\ell=0}^{\infty} \delta_{it\ell} u_{i,t-\ell}.$$

Collecting all unit-level responses into θ_i, the analysis is otherwise unchanged. In particular, letting $\hat{s}_{it} = s_{it} - N^{-1} \sum_{j=1}^{N} s_{jt}$, the coefficient on $\hat{s}_{it} X_t$ is the slope coefficient in the linear projection of β_{ith} on time fixed effects and s_{it}, and standard errors in Equation (11) are calculated using $\hat{s}_{it} X_t$ in place of $\hat{s}_i X_t$. Some specifications in our empirical analysis in Section 5 take this form, and we revisit it in simulations in Section 4.[17]

[16] Arkhangelsky and Korovkin (2023) derive a similar connection in a regional-exposure design context.

[17] In this paper, we focus on linear models. When interest is in a possible feedback loop from s_{it} to Y_{it}, a

VAR DGPs. Proposition 1 gives uniform guarantees for short horizons $h \leq p$. There is an important class of DGPs for which the result extends to long horizons: finite-order VARs. This case is useful for two reasons. First, VARs are a natural benchmark for macroeconomic dynamics and for the question of lag-length selection. Second, they provide a bridge to the structural-shock identification approaches considered below. Specifically, we model

$$Y_{it} = m_i + \sum_{\ell=1}^{p} A_\ell Y_{i,t-\ell} + A_{w0}\tilde{W}_{it} + \sum_{\ell=0}^{p} B_{i\ell}X_{t-\ell} + C_{i0}Z_t + D_{i0}u_{it}. \tag{13}$$

If the autoregressive polynomial $A(L) = 1 - \sum_{\ell=1}^{p} A_\ell L^\ell$ is stable, this model can be inverted into the distributed-lag representation (9). Let $\mathcal{P}_{\text{VAR}} \subset \mathcal{P}$ be the subclass of such VAR processes satisfying Assumption 1. For this subclass, the uniformity result extends to $h > p$.

Proposition 2. *Let Assumptions 1, A.1, A.2 and A.3 hold and fix $\phi < 1$. Then, for all $\epsilon > 0$ and $0 < \alpha < 1/2$,*

$$\lim_{T,N\to\infty} \sup_{P\in\mathcal{P}_{\text{VAR}}} \sup_{0\leq h\leq\phi T} P\Big(\Big|\hat{\beta}(h) - \beta(h)\Big| > \epsilon\Big) = 0,$$
$$\lim_{T,N\to\infty} \sup_{P\in\mathcal{P}_{\text{VAR}}} \sup_{0\leq h\leq\phi T} \Big|P\Big(\beta(h) \in \hat{C}_\alpha(h)\Big) - (1-\alpha)\Big| = 0.$$

In the VAR model (13), the lag-augmented regression score is serially uncorrelated at all horizons, not only for $h \leq p$. Thus, when a low-order VAR provides a good approximation, controlling for a small number of lags of the outcome and the shock delivers inference that is robust over long horizons, uniformly over the macro-micro error composition.

In practice, the main implementation question is the choice of lag length. Section 4 studies this issue and proposes a simple-to-implement information-criterion rule.[18]

3.3 Shocks identified from structural VAR models

We next consider the case in which the shock of interest is identified outside the panel regression from a structural VAR. This covers standard macroeconometric identification strategies,

nonlinear DGP is the starting point. Rambachan and Shephard (2021, Section 3.4) provide a nonparametric characterization of local-projection estimands with endogenous states in a time-series potential-outcomes framework; see also Gonçalves, Herrera, Kilian, and Pesavento (2024) and Winkler (2026).

[18]The results in Montiel Olea and Plagborg-Møller (2021) suggest uniform validity of $\hat{C}_\alpha(h)$ over a VAR parameter space that includes (near) unit roots. In turn, the results in Montiel Olea et al. (forthcoming) suggest that, for fixed h, t-LAHR inference would also remain valid if the VAR model (13) were locally misspecified. The simulation evidence in Section 4 is consistent with these observations.

including recursive, short-run, long-run, and heteroskedasticity-based identification. Our results above continue to apply when the panel local projection uses an estimated structural shock, provided the panel regression includes the variables needed to absorb the generated-regressor error (as in Breitung and Brüggemann, 2023, in a time series context).

Let $\boldsymbol{R}_t$ be an n-dimensional vector of macroeconomic variables observed by the researcher and suppose it admits the structural VAR representation

$$\alpha_0 \boldsymbol{R}_t = \sum_{\ell=1}^{p_R} \alpha_\ell \boldsymbol{R}_{t-\ell} + \boldsymbol{\varepsilon}_t,$$

with $\boldsymbol{\varepsilon}_t$ a vector of structural shocks satisfying the restrictions in Assumption 1. Let the shock of interest be the j-th entry, $X_t = \varepsilon_{jt} = \alpha_{0,j\bullet} \boldsymbol{R}_t - \sum_{\ell=1}^{p_R} \alpha_{\ell,j\bullet} \boldsymbol{R}_{t-\ell}$ where $\alpha_{\ell,j\bullet}$ is the j-th row of α_ℓ. The remaining entries of ε_t may overlap or correlate with the shocks in Z_t. An identification strategy pins down $\{\alpha_{\ell,j\bullet}\}_{\ell=0}^{p_R}$ from the autocovariances of $\boldsymbol{R}_t$ and allows the researcher to construct X_t. In practice, the researcher replaces these coefficients by estimates $\{\hat{\alpha}_{\ell,j\bullet}\}_{\ell=0}^{p_R}$ from a macro sample $\{\boldsymbol{R}_t\}_{t=1}^{T_R}$, typically with $T_R \gg T$, and constructs

$$\hat{X}_t = \hat{\alpha}_{0,j\bullet} \boldsymbol{R}_t - \sum_{\ell=1}^{p_R} \hat{\alpha}_{\ell,j\bullet} \boldsymbol{R}_{t-\ell}.$$

Propositions 1 and 2 extend to panel local projections that replace X_t by $\hat{X}_t$ in (2) as long as the regression includes the corresponding contemporaneous and lagged macro controls.

The recursive case makes this connection especially transparent. If the structural shock is ordered j-th, then

$$R_{jt} = -\sum_{k=1}^{j-1} \alpha_{0,jk} R_{kt} + \sum_{\ell=1}^{p_R} \alpha_{\ell,j\bullet} \boldsymbol{R}_{t-\ell} + X_t.$$

Thus, instead of explicitly constructing $\hat{X}_t$, one may replace X_t in (2) by R_{jt} and control for the interactions of s_i with the contemporaneous $R_{1t}, \ldots, R_{j-1,t}$ and the lags $\boldsymbol{R}_{t-1}, \ldots, \boldsymbol{R}_{t-p_R}$. The coefficient on $s_i R_{jt}$ then has the same interpretation as the coefficient on $s_i X_t$, and the t-LAHR confidence interval enjoys the same uniform guarantees of Propositions 1 and 2.[19]

[19] Controlling for $s_i R_{1t}, \ldots, s_i R_{j-1,t}$ and lags of $s_i \boldsymbol{R}_t$ ensures that the generated-regressor error $X_t - \hat{X}_t$ is asymptotically orthogonal to the residualized regression score. If $T/T_R \to 0$, this additional adjustment is not needed because first-stage estimation errors are asymptotically negligible.

3.4 Proxy shocks and panel LP-IV

Most empirical applications of panel local projections treat the shock of interest as observed. In some settings, however, the available shock measure is better viewed as a noisy proxy for an underlying structural shock. This creates a measurement-error problem that can be handled by using the proxy as an instrument for the endogenous aggregate variable, as in time-series LP-IV designs (Ramey, 2016; Stock and Watson, 2018).

Suppose the researcher observes Y_{it} and s_i, but not the structural shock X_t itself. Instead, she observes an endogenous aggregate state variable $\tilde{X}_t$ and a proxy shock X_t^*. In our empirical illustration based on Cloyne et al. (2023), for example, $\tilde{X}_t$ is the federal funds rate and X_t^* is an imperfect measure of monetary policy surprises. We augment (9) with

$$\tilde{X}_t = \sum_{\ell=0}^{\infty} b_\ell X_{t-\ell} + \sum_{\ell=0}^{\infty} c_\ell Z_{t-\ell}, \qquad X_t^* = a_0 X_t + \nu_t, \tag{14}$$

where ν_t is measurement or shock-elicitation error. Let $\mathcal{P}_{\text{IV}}$ denote the corresponding class of DGPs, including the joint distribution of ν_t.

Assumption 2 (Panel LP-IV). *For all $P \in \mathcal{P}_{\text{IV}}$, $a_0 \neq 0$, $b_0 = 1$, the mean independence restrictions in Assumption 1 hold with the conditioning sets augmented by $\{\nu_\tau\}_{\tau \neq t}$, and*

$$E_P\Big[\nu_t \;\Big|\; \{\nu_\tau\}_{\tau\neq t}, \{X_\tau, Z_\tau, \{u_{i\tau}\}_{i=1}^N\}_{\tau=-\infty}^{\infty}, \{s_i, \theta_i\}_{i=1}^N\Big] = 0.$$

The normalization $b_0 = 1$ fixes the scale of the impulse response, since LP-IV only identifies relative impulse responses. The condition $a_0 \neq 0$ ensures instrument relevance, and we focus on the strong-instrument case where a_0 is uniformly bounded away from zero as $N, T \to \infty$ (see Appendix A.1). The last condition is the exclusion restriction. It implies that, after accounting for controls, X_t^* is contemporaneously correlated with the panel outcome only through X_t. This is the standard lead-lag exogeneity requirement in time-series LP-IV designs; see, for example, Stock and Watson (2018, p. 924) and Plagborg-Møller and Wolf (2021, p. 970).[20] Additional regularity conditions are stated in Appendix A.1.

The panel LP-IV estimator $\hat{\beta}^{\text{IV}}(h)$ is the coefficient on $s_i\tilde{X}_t$ in

$$Y_{i,t+h} = \hat{\beta}^{\text{IV}}(h) s_i \tilde{X}_t + \hat{\psi}^{\text{IV}}(h)' W_{it} + \hat{\mu}_i^{\text{IV}}(h) + \hat{\nu}_t^{\text{IV}}(h) + \hat{\xi}_{it}^{\text{IV}}(h),$$

[20] The setup can be extended to serially correlated ν_t and to proxies that are valid only after conditioning on additional controls. Intercepts can also be added to (14) without changing the arguments.

where $W_{it} = (s_i\tilde{X}_{t-1}, Y_{i,t-1}, \ldots, s_i\tilde{X}_{t-p}, Y_{i,t-p}, \tilde{W}'_{it})'$. The endogenous regressors and controls are instrumented by $s_iX^*_t$ and $W^*_{it} = (s_iX^*_{t-1}, Y_{i,t-1}, \ldots, s_iX^*_{t-p}, Y_{i,t-p}, \tilde{W}'_{it})'$. Inference is obtained by time-clustering the corresponding IV score. Let $\hat{\sigma}^{\mathrm{IV}}(h)$ and $\hat{C}^{\mathrm{IV}}_{\alpha}(h)$ be the t-LAHR standard error and confidence interval—explicit formulas are given in Appendix A.

Proposition 3. *Let Assumptions 2, A.1, A.2 and A.3 (adapted to the LP-IV setting as indicated in Appendix A.1) hold and fix $h \leq p$. Then, for all $\epsilon > 0$ and $0 < \alpha < 1/2$,*

$$\lim_{T,N\to\infty} \sup_{P\in\mathcal{P}_{\mathrm{IV}}} P\Big(\Big|\hat{\beta}^{\mathrm{IV}}(h) - \beta(h)\Big| > \epsilon\Big) = 0,$$
$$\lim_{T,N\to\infty} \sup_{P\in\mathcal{P}_{\mathrm{IV}}} \Big|P\Big(\beta(h) \in \hat{C}^{\mathrm{IV}}_{\alpha}(h)\Big) - (1-\alpha)\Big| = 0.$$

Proposition 3 shows that the estimand and inference results for panel local projections extend to the IV case. Under the normalization $b_0 = 1$, the LP-IV estimand is the same object as before. This equivalence is not automatic under heterogeneity. In cross-sectional IV settings, treatment-effect heterogeneity typically leads to weighted local average treatment effects (Angrist and Imbens, 1995; Angrist, Imbens, and Graddy, 2000). Here the first stage is aggregate. As a result, the instrument corrects measurement error in the aggregate shock without reweighting units by heterogeneous first-stage responses. This is another consequence of the macro–micro structure studied in this paper. Finally, note that uniformity over horizons can also be obtained by imposing a finite-order VAR structure on (9) and (14), as in Equation (13).

4 Simulation study

In this section, we show through simulations that the proposed t-LAHR procedure has excellent finite-sample properties across realistic designs, sample sizes and macro–micro regimes, and compare it with those of alternative methods. Here we provide a summary and defer additional detail and results to the Appendix.

4.1 Implementation details

Designs. Our study relies on two different DGPs. The first is the general setup (9) supplemented with (14) to cover the endogenous case. We begin by simulating shocks $\{X_t, Z_t, \nu_t, \{u_{it}\}_{i=1}^N\}$ as mutually and serially independent $N(0,1)$ random variables, and by drawing $\{\theta_i, s_i\}_{i=1}^N$ independently across units. To ensure correlation between observed and unobserved heterogeneity we use a technique described in Appendix D. We calibrate the distribution of $\{\beta_{i\ell}, \gamma_{i\ell}, \delta_{i\ell}\}$ and $\{b_\ell, c_\ell\}$ to produce realistic degrees of shock persistence.

Given these elements, we generate the inputs for panel LP and LP-IV procedures, namely Y_{it}, X_t, s_i, $\tilde{X}_t$, X_t^*. We also simulate the time-varying covariate $s_{it} = s_i + \zeta_{it}$ (where ζ_{it} is such that s_{it} remains strictly exogenous) to compare panel LPs on $s_i X_t$ and $s_{it} X_t$ — this illustrates the point we made in Remark 3.

The second DGP is the VAR model (13). Again we generate shocks as i.i.d. $N(0,1)$ and we simulate the heterogeneity as detailed in Appendix D. When calibrating the VAR parameters $\{A_\ell\}$ we allow the largest AR root to be $1 - c/T$ (we use $c = 5$) to capture the essence of a near non-stationary environment.

The results below are based on $n_{\mathrm{MC}} = 5,000$ Monte Carlo samples. Motivated by our survey of the empirical literature, we look at designs with $T = 30$ and $T = 100$. We set $N/T = 50$, so that $N = 1,500$ when $T = 30$ and $N = 5,000$ when $T = 100$, and we calibrate $\mathrm{Var}_P(\, u_{it}\,)$ to take values consistent with $\bar{R}^2 \in \{0.99, 0.66, 0.33\}$. (As a reference, $\bar{R}^2 = 0.66$ corresponds to the one-third of aggregate fluctuations explained by micro shocks suggested by Gabaix, 2011 for GDP growth.)

We provide additional details in Appendix D and consider alternative designs with time-varying s_{it}, conditionally heteroskedastic shocks, non-Gaussian shocks featuring skewness and heavy tails, spatially correlated micro shocks u_{it}, more persistent macro shocks in the general DGP in (9), proxy shocks recovered from (approximate) SVARs, and impulse responses over longer horizons. All results reinforce our conclusions in this section.

***t*-LAHR inference.** Our preferred implementation uses a lag-selection rule given by

$$\hat{p}(h) = \max\{1, \min(h, \hat{p}^{AIC})\}, \tag{15}$$

where $\hat{p}^{AIC}$ is the number of lags selected by the AIC. The reasoning for this rule combines data, theory and comprehensive simulation evidence. First, we include at least one lag of the outcome and the shock at all horizons. This provides insurance against near unit-root scenarios via lag augmentation[21] and reflects efficiency considerations in realistic empirical designs (see the discussion below). Second, we use a data-driven selection rule via AIC to

[21] This is motivated by the results in Montiel Olea and Plagborg-Møller (2021) and Montiel Olea et al. (forthcoming), which we expect to hold in our setup (by means of the time-series representation of the problem). We also show very good performance of t-LAHR inference in simulation designs with high degrees of persistence; see Appendix D.6.

pick the maximum lag length. In the Appendix, we show that this delivers consistent lag selection (Section B.1) and that it provides excellent performance in simulations (Section B.3). Third, we leverage our result that the scores have limited leftover serial correlation of order h to ensure that we limit lag length appropriately.

On top of this, we adjust t-LAHR confidence intervals with the finite-sample adjustments advocated by Imbens and Kolesár (2016), which combine an HC2 leverage adjustment with a Student-t critical value based on effective degrees of freedom. These are very simple to implement and leverage the connection with the synthetic time-series representation. Additional details are reported in Appendix B.2, whereas simulation evidence comparing t-LAHR with and without finite-sample refinements is offered in B.3. Whereas these adjustments are minor for $T = 100$, they prove important for $T = 30$ and ensure that t-LAHR inference provides excellent coverage guarantees even for settings with as few as $T - H \approx 25$ effective (synthetic) observations.

Other methods. We compare t-LAHR inference with one-way (1W(i)), two-way (2W), and Driscoll-Kraay (DK98) confidence intervals. These are not lag-augmented, following standard practice. We set the bandwidth in Driscoll-Kraay to $\lceil 4((T-h)/100)^{2/9} \rceil$ following Newey and West (1994) and the default Stata implementation.

For illustrative purposes, we also include t-HR (the non-lag-augmented counterpart to t-LAHR) and other autocorrelation-robust methods following the discussion in Section 3.2. In particular, we include a variant using the Hansen and Hodrick (1980) variance estimator in (t-HAC) and another using the equally-weighted cosine approach (Müller, 2004) with the choice of tuning parameter recommended in Lazarus et al. (2018) (t-HAR).

4.2 Results

In Figures 1 and 2, we report pointwise coverage rates for horizons $0 \leq h \leq 0.15T$ for $T = 100$ and $T = 30$, respectively. These correspond to 90% confidence intervals for panel LP and LP-IV using s_i to interact the aggregate shock. Panels (a)-to-(c) display LP while (d)-to-(f) display LP-IV in the general DGP; panels (g)-to-(i) display LP in the VAR DGP.

The main takeaway from these results is that t-LAHR inference exhibits excellent finite-sample performance, both in absolute terms across scenarios and also relative to all other methods. Coverage is close to the nominal rate even in low-signal cases and for horizons h well beyond p. Its mean absolute coverage distortion never exceeds 3 pp for $T = 100$ and 5 pp for $T = 30$ (and is often much smaller), consistently ahead of the second-best option, which nonetheless exhibits substantial size distortions. Importantly, the second-best alternative differs across specifications, estimation methods, sample sizes and macro–micro regimes. We now provide additional discussion on the performance of these other methods.

Comparison with within-units and two-way clustering. The first and second most popular inference methods in empirical work guard against the wrong type of correlation in the score, are not uniformly valid and display substantial coverage distortions across designs.

One-way clustering is very sensitive to $\bar{R}^2$, suffering major distortions in intermediate- and high-$\bar{R}^2$ cases. In fact, as $\bar{R}^2 \to 1$, its worst-case coverage probability tends to zero. What is more, it is outperformed by t-LAHR even if micro shocks explain the majority of aggregate variation.

Two-way clustering provides some insurance against different macro–micro regimes, but does not consistently estimate the autocovariance function of the scores (see Lemma A.1 in Appendix A.2) and displays substantial distortions in our simulations (up to 13 pp with $T = 100$ and 23 pp with $T = 30$). Note that it is usually close to t-HR, its non-i-clustered version; another indication that there is no clear advantage in clustering by units. In fact, in certain occasions (mainly low-signal and near non-stationary designs), 2W gives worse inferences than t-HR or 1W(i) alone. In environments with (near) micro unit roots, variance estimators tend to perform poorly, and we document that 2W is not positive definite in a non-negligible fraction of these scenarios (see also Appendix D.9).

Comparison with *t*-HAR/HAC/DK98 methods. Methods that estimate the long-run variance of the (synthetic) scores (such as Driscoll-Kraay standard errors) could, in principle, be a valid alternative to lag augmentation. In practice, however, they struggle with realistic sample sizes, display severe coverage distortions, and are clearly dominated by t-LAHR.

Indeed, Driscoll-Kraay confidence intervals display pronounced size distortions in our simulations — up to 11 pp with $T = 100$ and 25 pp with $T = 30$. DK98 has comparable performance (oftentimes noticeably worse) to 2W across designs and DGPs — a method that ignores serial dependence and does not deliver a consistent estimator of the autocovariance of the scores. In fact, in low-signal scenarios and for $T = 30$ it performs even worse than 1W, which underscores the subpar behavior of these methods in finite samples. Let us elaborate on the reasons behind this:

1. t-HAR/HAC estimators inherit the well-documented finite-sample issues identified for HAR/HAC estimators in the time series literature (Lazarus et al., 2018), which stem

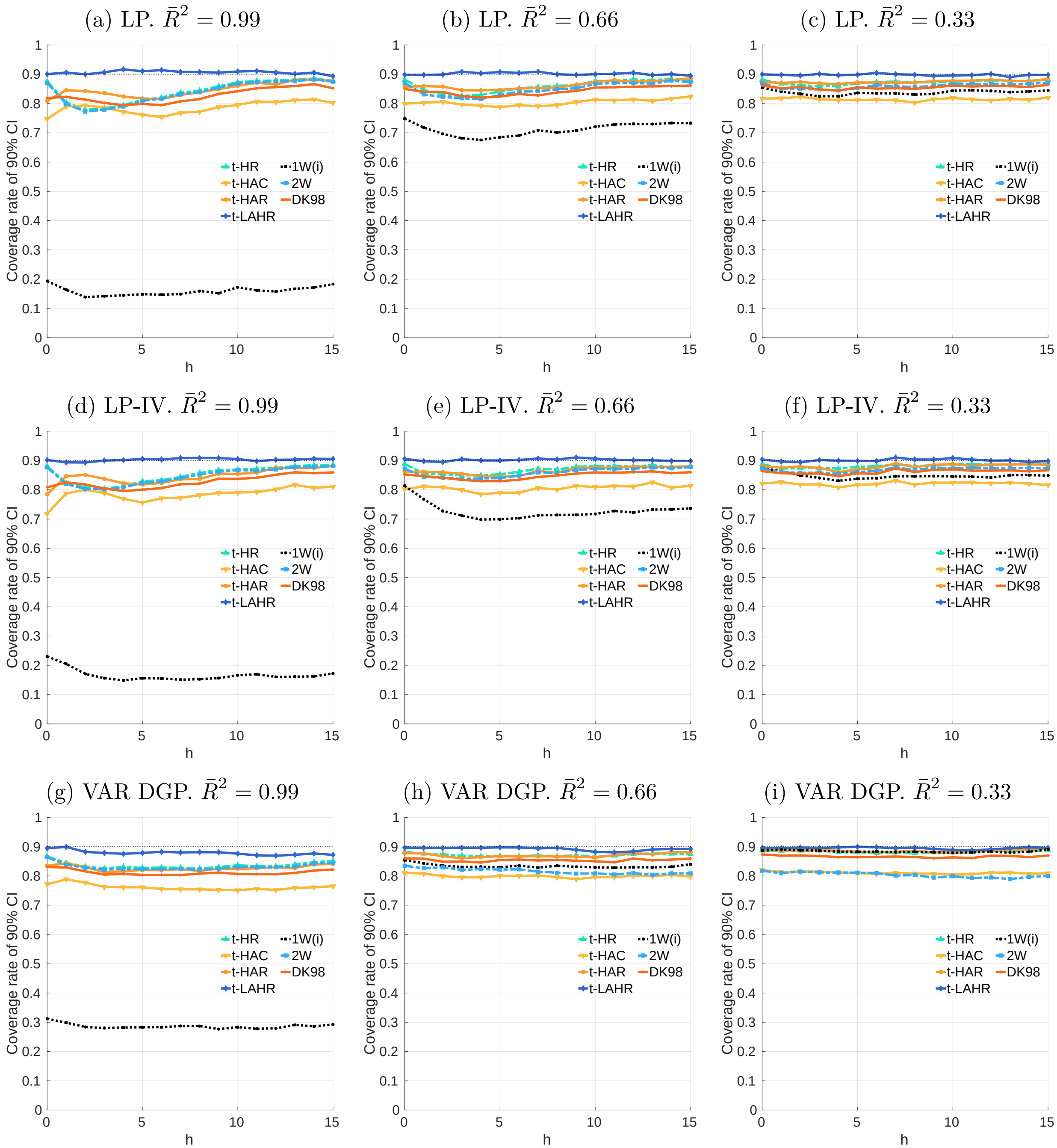


FIGURE 1. Coverage rates of 90% confidence intervals for $T = 100$.

Note: 1W(i) refers to one-way (unit-level) clustering, 2W to two-way clustering, DK98 to Driscoll–Kraay, and t-HR/t-LAHR/t-HAR/t-HAC to the time-level clustering approaches discussed in the text.

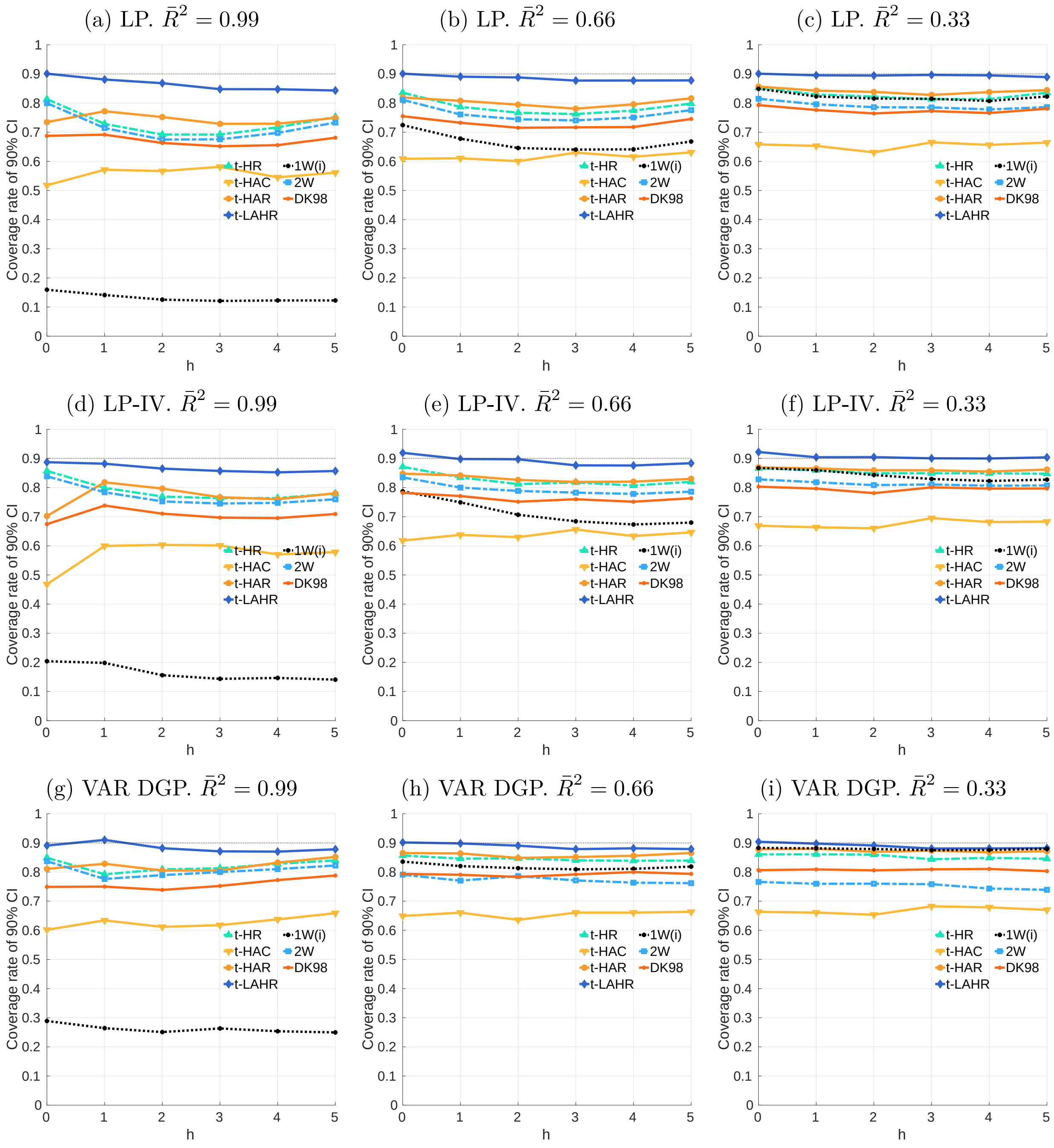


FIGURE 2. Coverage rates of 90% confidence intervals for $T = 30$.

Note: 1W(i) refers to one-way (unit-level) clustering, 2W to two-way clustering, DK98 to Driscoll–Kraay, and t-HR/t-LAHR/t-HAR/t-HAC to the time-level clustering approaches discussed in the text.

from the difficulty of estimating autocovariances in small samples (see also Herbst and Johannsen, 2023) and of choosing kernel and truncation parameters. We show this in Appendix C by mapping well-known results in the time series literature (Sun, Phillips, and Jin, 2008; Sun, 2014) to the synthetic time series representation of our problem, and illustrate it in simulations.

Our results also highlight the sensitivity of t-HAR/HAC to the choice of tuning parameters. For example, t-HAC estimators use a uniform kernel instead of a triangular kernel and display even larger coverage distortions, in part due to the fact that these variance estimators are not necessarily positive definite in finite samples.

2. t-HAR/HAC estimators are more difficult to refine in finite samples. Whereas t-HAR (which builds on state-of-the-art results from the HAR literature) performs better than Driscoll-Kraay, it nonetheless displays significant distortions even with $T = 100$, often comparable to 2W clustering — which, again, completely ignores autocorrelation in the scores. In contrast, t-LAHR only requires simple, standard adjustments insofar as it is in the class of heteroskedasticity-robust methods (on the synthetic time series).

3. t-HAR/HAC and t-LAHR echo recent results from the time series literature favoring lag augmentation over autocorrelation-robust variance estimation (Montiel Olea and Plagborg-Møller, 2021; Montiel Olea et al., forthcoming), particularly with highly persistent data. Driscoll-Kraay standard errors fail in (near) non-stationary environments, which we illustrate in Figure D.7 in Appendix D.6, where even with $T = 300$ DK98 undercovers by up to 10 pp. In contrast, t-LAHR hits nominal coverage across all horizons and macro–micro regimes.

4. t-HAR/HAC variance estimators are less efficient than t-LAHR and non-lag-augmented estimators are more biased in empirically realistic designs. It is easy to show that lag augmentation delivers efficiency gains and bias-reduction unless impulse responses display alternating patterns (sign oscillations from one horizon to the next); see Lemma A.1 in Appendix A.2. We report length of confidence intervals in Appendix D.9 and bias and RMSE in Appendix D.10. We show that t-LAHR dominates all other methods. In particular, coverage-adjusted t-LAHR confidence intervals are up to 70% shorter than Driscoll-Kraay confidence intervals, and lag augmentation is associated with notable gains in bias and RMSE, particularly over short horizons.

In sum, the small-sample evidence provided in this section shows that the large-sample approximations of Section 3 provide reliable guidance for understanding estimation and inference in panel regressions with aggregate shocks, illustrates the importance of achieving uniformity with respect to the macro-micro composition of the errors, reinforces our practical recommendations and delivers a clear methodological prescription: t-LAHR inference.

5 Empirical illustrations

We now replicate three relevant papers that explore the importance of firm-level heterogeneity in shaping the investment channel of monetary policy and provide a systematic reassessment of the available empirical evidence. Our methods qualify the significance of several established results and deliver new lessons on the role of financial frictions in the transmission of monetary policy.

5.1 Data and background

From a policy perspective, investment is key to the effectiveness of monetary policy. However, empirical evidence about the specific channels through which it operates and the type of firms that are most responsive to policy interventions has remained inconclusive.

As a result, a growing empirical literature has offered evidence on the importance of financial frictions in accounting for the differential responsiveness of firms to monetary policy. In an influential paper, Ottonello and Winberry (2020) focus on the role of default risk and find that firms with low leverage and those with high "distance to default" are more responsive to monetary shocks. They build on these results to embed default risk into a heterogeneous firm New Keynesian model. Instead, Cloyne et al. (2023) focus on credit constraints and document that younger firms and those not paying dividends adjust their investment significantly more than older firms paying dividends. On a different line, Jeenas and Lagos (2024) highlight the importance of the asset price channel. They find that stock market turnover has persistent effects on investment following a monetary policy shock for firms with a relatively low liquidity ratio (cash and short-term investments to total assets), but this is not the case for those with better access to internal financing.

We now revisit these findings through the lens of our framework. All three papers use Compustat quarterly data, a panel of publicly listed U.S. firms, and retrieve surprises to the Fed Funds rate as measures of monetary policy shocks,[22] which cover comparable timeframes

[22] In particular, Ottonello and Winberry (2020) use the shock series from Gürkaynak, Sack, and Swanson

from the 1990s to the 2010s. Additional details on sample sizes are reported in Table 2 below.

More specifically, these studies consider the differential firm-level response of a measure of the investment rate at horizon h, which we denote as $\Delta k_{i,t+h}$,[23] to monetary policy shocks X_t along different margins of heterogeneity captured by a possibly time-varying vector s_{it}, which can be cast into the now familiar form

$$\Delta k_{i,t+h} = a_{h,i} + d_{h,g(i)t} + \beta(h)' s_{i,t-1} X_t + \eta(h)' W_{it} + v_{h,it}, \tag{16}$$

where $a_{h,i}$ are firm fixed effects and $d_{h,g(i)t}$ are time indicators, which might be group g-specific (say, industry-by-time or sector-by-time dummies), and W_{it} a set of (lagged) micro and macro controls, possibly including lags of the outcome or shock of interest. Table 2 reports additional details on these choices. Micro-level controls are ubiquitous, unlike interacting s_{it} with aggregate variables. Similarly, lags of the outcome or the shock, which play an important role in our inference procedures, are rare. Equation (16) is the main estimating equation in both Ottonello and Winberry (2020) and Jeenas and Lagos (2024), whereas it corresponds to the reduced-form in Cloyne et al. (2023), who report panel LP-IV estimates along the lines of Section 3.4 treating the one-year changes in the interest rate as the endogenous variable.

In what follows, we focus on the following choices for s_{it}. In Ottonello and Winberry (2020), s_{it} is either a (demeaned) measure of the firm's leverage or distance to default; in Cloyne et al. (2023) it refers to group indicators depending on whether a firm is young (less than 15 years since incorporation) and whether it pays dividends; in Jeenas and Lagos (2024) it interacts a measure of stock turnover with an indicator of whether firm i either belongs to the bottom or top half of the liquidity ratio distribution. While in some cases the authors consider other outcomes and explore other margins of heterogeneity, these are in our view the main mechanisms. These are accordingly emphasized throughout the papers, for instance, by assessing model fit relative to these estimated IRFs.

(2005) and Gorodnichenko and Weber (2016), Cloyne et al. (2023) follow Gertler and Karadi (2015) and Jeenas and Lagos (2024) retrieve the time series of shocks from Jarociński and Karadi (2020).

[23] In particular, Ottonello and Winberry (2020) use $\Delta k_{i,t+h} = \log K_{i,t+h} - \log K_{i,t-1}$, where K_{it} is the value of the tangible capital stock at the end of quarter t; Cloyne et al. (2023) use $\Delta k_{i,t+h} = I_{i,t+h}/K_{i,t+h} - I_{i,t-1}/K_{i,t-1}$, where I_{it} are capital expenditures and K_{it} is physical capital at the beginning of t; Jeenas and Lagos (2024) let $\Delta k_{i,t+h} = \log I_{i,t+h} - \log K_{i,t+h}$, where again I_{it} is a measure of capital expenditures and K_{it} a measure of the stock of capital at the beginning of t.

TABLE 2. Sample details and specification choices across empirical studies.

	OW20	CFFS23	JL24
Micro controls	Yes	Yes	Yes
Macro controls	Yes	No	No
Lags of outcome	No	No	Yes (one)
Lags of $s_{i,t-1}X_t$	No	No	No
Standard errors	Two-way clustering	Driscoll-Kraay	Within-units clustering
Sample period	1990Q1–2007Q4 ($T = 72$)	1986Q1–2016Q4 ($T = 124$)	1990Q1–2016Q4 ($T = 108$)
Number of firms	$N = 6{,}522$	$N = 15{,}460$	$N = 5{,}236$

Note: OW20 stands for Ottonello and Winberry (2020), CFFS23 refers to Cloyne et al. (2023), and JL24 refers to Jeenas and Lagos (2024). Controls refer to variables beyond fixed effects, as in Equation (16). Time series sizes T refer to the entire sample period; the panels are unbalanced.

Synthetic time series representation. A fundamental insight of our paper is that the synthetic time series form of the microdata is a sufficient statistic for the panel LP, yielding a low-dimensional representation of a highly complex, unbalanced dataset.[24]

For illustration, Figure 3 displays the representation for leverage in Ottonello and Winberry (2020), while those for the other two applications are reported in Figure E.1 in the Appendix. Recall that impulse responses can also be recovered from local projections on these synthetic time series. It is clear that movements in synthetic outcomes concurrent with surprise cuts in policy rates, mostly around recessions, are the main source of identification. Still, it is difficult to visually discern a clear correlation between the synthetic outcome and monetary policy shocks, anticipating an overall noisy relationship. The sub-

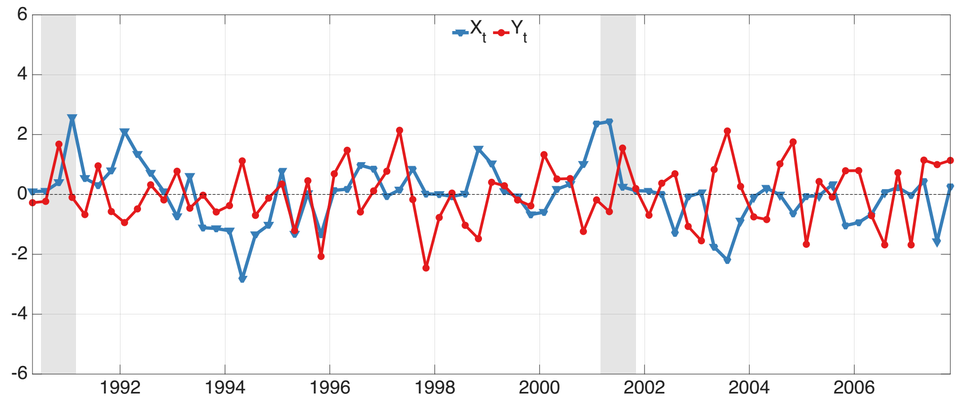
[24]Remark 1 generalizes as follows with unbalanced panels and time-varying s_{it}. Let $d_{it} = 0$ indicate a missing observation; $d_{it} = 1$ otherwise. Abstracting from controls, the panel local projection estimator is

$$\hat{\beta}(h) = \frac{\sum_{t=1}^{T-h} \sum_{i=1}^{N} d_{it} s_{it} X_t Y_{i,t+h}}{\sum_{t=1}^{T-h} \sum_{i=1}^{N} d_{it} s_{it}^2 X_t^2} = \frac{\sum_{t=1}^{T-h} \omega_t X_t \hat{Y}_{t+h}}{\sum_{t=1}^{T-h} \omega_t X_t^2},$$

where $\omega_t = \sum_{i=1}^{N} d_{it} s_{it}^2$ and $\hat{Y}_{t+h} = (\sum_{i=1}^{N} d_{it} s_{it}^2)^{-1} \sum_{i=1}^{N} d_{it} s_{it} Y_{i,t+h}$. This is a weighted least squares regression of slope coefficients $\hat{Y}_{t+h}$ on X_t. Note that if $s_{it} = 1$ the weights boil down to the number of non-missing observations $\omega_t = \sum_{i=1}^{N} d_{it}$, as intuition suggests. Our theory applies with data missing at random. Note that Figure 3 directly plots rescaled synthetic outcomes $\sqrt{\omega_t}\hat{Y}_{t+h}$ and shocks $\sqrt{\omega_t}X_t$.

stantial variation in synthetic outcomes unrelated to X_t indicates the presence of omitted aggregate or non-negligible idiosyncratic shocks, which are central elements of our framework and are appropriately accounted for in our estimation and inference theory.

FIGURE 3. Synthetic time series representation; leverage, Ottonello and Winberry (2020).



Note: The figure corresponds to Figure 1(a) in Ottonello and Winberry (2020). Grey areas are NBER-dated recessions. X_t and Y_t are standardized to zero mean and unit variance, and $X_t > 0$ indicates a surprise cut in the Fed Funds rate.

5.2 Revisiting the investment channel of monetary policy

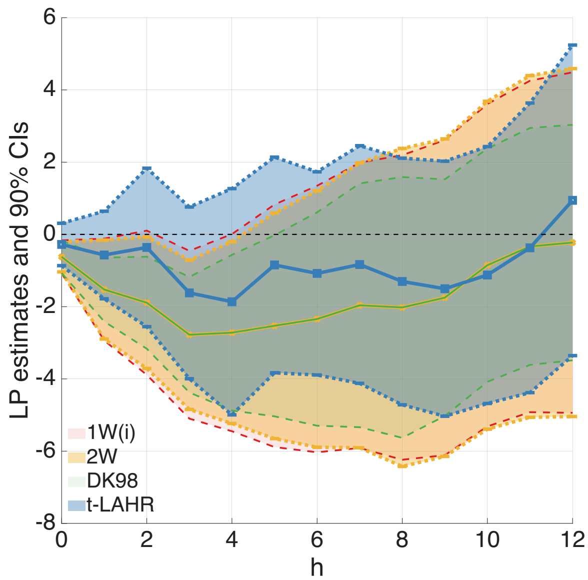
We report impulse responses for the different s_{it} in Figure 4 (for Ottonello and Winberry, 2020), Figure 5 (for Cloyne et al., 2023), and Figure 6 (for Jeenas and Lagos, 2024).

Solid lines indicate point estimates. From Section 3, we know that under unrestricted heterogeneity these IRFs recover meaningful cross-sectional moments: They are informative about differential responses for firms with larger versus smaller values of these covariates at each horizon. For example, an impact effect of 0.4 in Figure 4(b) implies that the semi-elasticity of investment to monetary policy is 0.4 pp larger for firms that are one standard deviation further from default than usual (the authors standardize s_{it} over the entire sample).

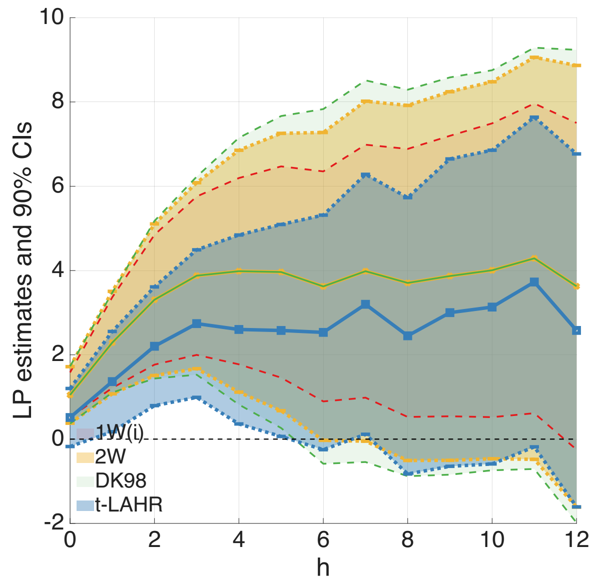
Comparing inference methods. In the figures, we also report t-LAHR bands together with the three most popular alternatives in empirical practice: clustering within-units (1W(i)), two-way clustering (2W), and Driscoll and Kraay (1998) (DK98).

These three papers cover the three different alternatives: Ottonello and Winberry (2020) cluster standard errors two-way within firms and quarters, Cloyne et al. (2023) use Driscoll-Kraay and set the number of autocovariances to 12, and Jeenas and Lagos (2024) cluster standard errors within firms, with an additional clustering step at the industry-by-quarter level (which will nonetheless neglect economy-wide aggregate shocks). For each different paper, we highlight the author's choice and report it following their exact specification; all other alternatives are reported in a lighter shade and following standard implementations (see Section 4). For t-LAHR inference we use the lag selection rule in (15).[25]

FIGURE 4. Impulse responses, Ottonello and Winberry (2020)



(a) Leverage



(b) Distance to default

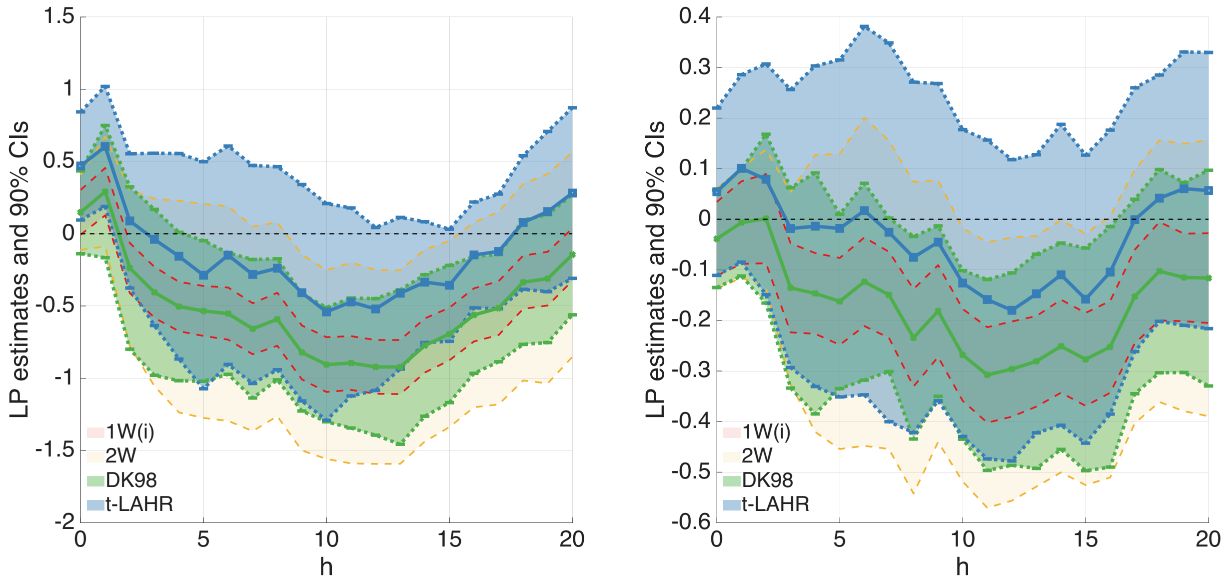
Note: This figure corresponds to Figure 1 in Ottonello and Winberry (2020). Solid lines correspond to point estimates and dashed lines to 90% confidence bands.

The main message from this section is that popular methods can deviate significantly from the (robust) t-LAHR method and lead to different conclusions about the role of various margins of heterogeneity — like default risk, credit constraints or firm liquidity — in shaping the transmission of monetary policy for firm-level investment. In particular, our results reinforce the existence of a dynamic, distance-to-default gradient for monetary policy, whereas we do not find sufficient evidence in these data of differential responses along firm leverage, age, dividend-paying status and turnover-liquidity, contrary to previous results.

As a simple, illustrative calculation, there are 44 different statistically significant horizons using the original authors' choice of confidence bands across Figures 4, 5 and Figure 6. Using

[25]Specifically, $\hat{p}^{AIC} = 1$ in Figures 4(a) and 6(a), $\hat{p}^{AIC} = 2$ in Figures 4(b) and 5(b), and $\hat{p}^{AIC} = 4$ in Figure 5(a) and 6(b). BIC and AIC coincide in all cases but Figure 5(a), where BIC chooses two lags (and delivers essentially identical results).

FIGURE 5. Impulse responses, Cloyne et al. (2023)



(a) Young, non-dividend-paying firms

(b) Old, dividend-paying firms

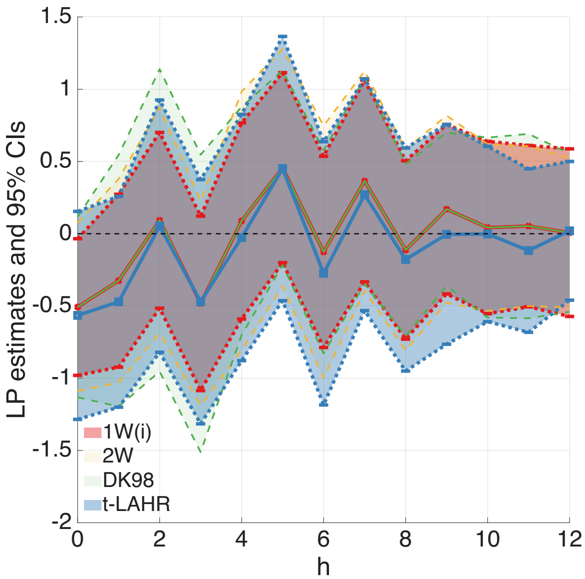
Note: This figure corresponds to Figure 3 in Cloyne et al. (2023). Solid lines correspond to point estimates and dashed lines to 90% confidence bands.

t-LAHR bands, only 6 of those remain. Importantly, this is not necessarily a consequence of t-LAHR bands being unnecessarily wider across horizons and designs, nor is it possible to rank any of the alternatives in terms of width of confidence intervals. This is documented in Table 3, which reports width relative to t-LAHR bands averaged over all horizons in each figure. For every alternative, it is possible to find instances where on average it delivers both wider and tighter confidence intervals.

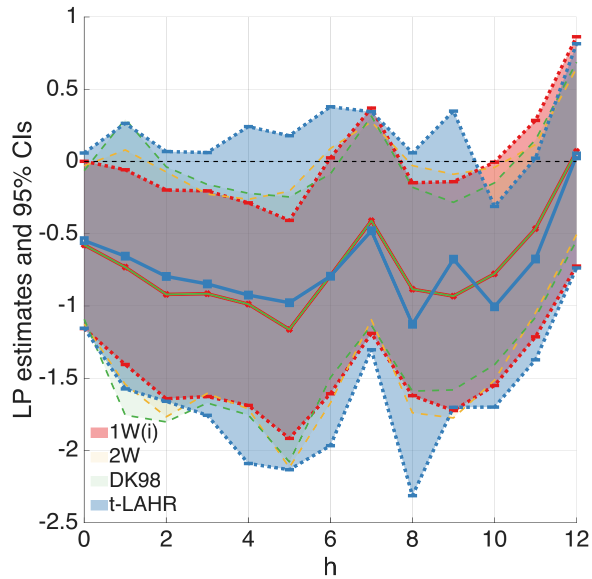
As an illustration, note that Driscoll and Kraay standard errors deliver noticeably smaller confidence intervals than one-way, two-way and t-LAHR in Figure 4(a), and would suggest a significant differential effect for more levered firms on impact and up to five quarters later; an even stronger pattern than originally reported by Ottonello and Winberry (2020). Driscoll and Kraay bands are also 35% wider on average than t-LAHR bands in Figure 4(b), and instead only about two thirds the width of t-LAHR's in Figure 5(b). Overall, we document that Driscoll and Kraay standard errors can be misleading and lead to opposite conclusions even in settings with $T > 100$, in line with our simulation evidence.

In a similar way, both one-way within-units and two-way clustering can produce intervals that are too short or too wide. Lag augmentation, which is mostly absent from current practice (see again Table 2), ensures the validity of t-LAHR inference relative to the latter.

FIGURE 6. Impulse responses, Jeenas and Lagos (2024)



(a) High liquidity



(b) Low liquidity

Note: This figure corresponds to Figure 2C in Jeenas and Lagos (2024). Solid lines correspond to point estimates and dashed lines to 95% confidence bands.

Similarly, t-LAHR intervals deliver robust inference across studies featuring different relative sizes of micro and macro variation. In the last row of Table 3 we report the ratio of within-units to two-way clustering, which is a heuristic estimate of the relative importance of micro shocks (and its complement of that of aggregate variation).[26] There is substantial variation across studies, from low macro signal environments where within-units and two-way confidence intervals are comparable (as in Ottonello and Winberry, 2020; Jeenas and Lagos, 2024) to those where aggregate variation plays an important role in the microdata (as in Cloyne et al., 2023). As such, the relative signal of macro shocks proves to be a key ingredient in our framework, and our results suggest a simple recipe that ensures robustness to any such regime: t-LAHR inference.

Taking stock of the results. In this section, we revisited the existing empirical evidence for the role of heterogeneity in the transmission of monetary policy to firm-level investment. This evidence is used to calibrate and assess the predictions of quantitative models, and is

[26]Consider the expression for $\bar{R}^2$ in (7) in Section 2, where the numerator is given by $\sigma^2_{\bar{U}}$, which regulates the size of idiosyncratic shocks. Essentially, the within-units clustered variance (1W(i)) accounts for this, whereas two-way clustering takes care of the denominator. Rearranging, we obtain $\bar{R}^2 = 1 - \frac{1\mathrm{W(i)}}{2\mathrm{W}}$.

TABLE 3. Relative width of confidence intervals across empirical studies

	Figure 4		Figure 5		Figure 6	
	(a)	(b)	(a)	(b)	(a)	(b)
$\frac{\text{1W(i)}}{t\text{-LAHR}}$	1.13	0.96	0.32	0.31	0.81	0.80
$\frac{\text{2W}}{t\text{-LAHR}}$	1.11	1.25	1.13	0.86	0.93	0.81
$\frac{\text{DK98}}{t\text{-LAHR}}$	0.85	1.35	0.79	0.65	0.95	0.79
$\frac{\text{1W(i)}}{\text{2W}}$	1.01	0.77	0.28	0.36	0.87	0.98

Note: Entries report relative CI widths, averaged over all horizons, for each panel in Figures 4, 5 and 6.

part of a large, growing body of empirical evidence which aims at answering a key policy question: Which firms are the most responsive to changes in monetary policy?

While interested in a similar question, we documented very different setups and design choices across applications. These papers differ in the estimation method (panel LP or LP-IV), inference method (clustering within-units, two-way clustering or Driscoll-Kraay), choice of micro and macro controls, and relative importance of aggregate shocks. Our framework explicitly incorporates all these ingredients and allows for a unified assessment of the existing empirical evidence.

In particular, after accounting for micro and macro shocks of any relative size and for the possible presence of a serially correlated aggregate component, our results reinforce the case for a dynamic distance-to-default channel. Moreover, this emerges as the most empirically salient proxy for financial frictions in these data, as the evidence for differential responses by leverage, liquidity, or credit constraints is substantially weaker than previously documented.

6 Conclusion

The use of microdata to answer macro questions offers an exciting avenue to study how agents respond to economy-wide policies. Possibilities include a better understanding of the dynamic transmission of shocks and the nature of heterogeneity.

Challenges are ubiquitous too. We propose a disciplined approach to estimation and inference when both aggregate and idiosyncratic shocks coexist and interest lies in parameters identified principally by macro shocks. The size of aggregate shocks relative to micro-level noise determines both the strength of identifying variation and the nature of estimation error. Since this macro–micro composition is unknown to the researcher and context-specific, it is important to design econometric procedures that are uniformly valid across all such regimes.

One such scenario is the estimation of impulse responses to macro shocks via panel local projections when rich microdata and measurements of the shock of interest are available. This includes cases where the shock of interest is directly observed, retrieved from a macro system, or contaminated with measurement error. Despite the complex environment, inference is simple and robust: it involves including lags as controls and clustering at the time level, and is valid regardless of the relative importance of aggregate shocks in the microdata.

Our framework generalizes beyond the empirical applications we have focused on. Other, related literatures where identification comes from randomness in group-level shocks include regional-exposure and shift-share designs. In fact, impulse responses are sometimes of interest too, as in the literature on cross-sectional fiscal multipliers (Chodorow-Reich, 2019).

We also leave some interesting dimensions for future research. Quantifying the relative importance of macro shocks (perhaps a lower bound) seems relevant in settings where uniform inference is not possible; these issues are becoming more salient as macroeconomists embrace the use of microdata to sharpen identification (Nakamura and Steinsson, 2018). On a different note, strong persistence of micro-level shocks is likely a feature of many datasets, and this is only captured in an indirect sense by our macro–micro device. Formalizing the idea of (possibly heterogeneous) non-stationarities along these lines seems promising and full of empirical content. Finally, extensions to simultaneous inference over horizons are possible building on the techniques in Jordà (2009) and Montiel Olea and Plagborg-Møller (2019).

References

Adão, R., M. Kolesár, and E. Morales (2019): "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics*, 134, 1949–2010.

Almuzara, M. and V. Sancibrián (2025): "Micro Responses to Macro Shocks." Federal Reserve Bank of New York Staff Report 1090.

Andrews, D. K. (2005): "Cross-section regression with common shocks." *Econometrica*, 73, 1551–1585.

Angrist, J. D. and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association*, 90, 431–442.

ANGRIST, J. D., G. W. IMBENS, AND K. GRADDY (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish." *Review of Economic Studies*, 67, 499–527.

ARKHANGELSKY, D. AND V. KOROVKIN (2023): "On Policy Evaluation with Aggregate Time-Series Shocks." Working Paper arXiv:1905.13660.

BREITUNG, J. AND R. BRÜGGEMANN (2023): "Projection estimators for structural impulse responses." *Oxford Bulletin of Economics and Statistics*, 85, 1320–1340.

CHANG, M., X. CHEN, AND F. SCHORFHEIDE (2024): "Heterogeneity and Aggregate Fluctuations." *Journal of Political Economy*, 132.

CHODOROW-REICH, G. (2019): "Geographic Cross-Sectional Fiscal Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy*, 11, 1–34.

CLOYNE, J., C. FERREIRA, M. FROEMEL, AND P. SURICO (2023): "Monetary Policy, Corporate Finance, and Investment." *Journal of the European Economic Association*, 21, 2586–2634.

CROUZET, N. AND N. R. MEHROTRA (2020): "Small and Large Firms over the Business Cycle." *American Economic Review*, 110, 3549–3601.

DRECHSEL, T. (2023): "Earnings-Based Borrowing Constraints and Macroeconomic Fluctuations." *American Economic Journal: Macroeconomics*, 15, 1–34.

DRISCOLL, J. AND A. KRAAY (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics*, 80, 549–560.

GABAIX, X. (2011): "The Granular Origins of Aggregate Fluctuations." *Econometrica*, 79, 733–772.

GERTLER, M. AND P. KARADI (2015): "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

GONÇALVES, S. (2011): "The moving blocks bootstrap for panel linear regression models with individual fixed effects." *Econometric Theory*, 27, 1048–1082.

GONÇALVES, S., A. M. HERRERA, L. KILIAN, AND E. PESAVENTO (2024): "State-dependent local projections," *Journal of Econometrics*, 244, 105702.

GORODNICHENKO, Y. AND M. WEBER (2016): "Are Sticky Prices Costly? Evidence from the Stock Market." *American Economic Journal: Macroeconomics*, 8, 160–199.

Gürkaynak, R. S., B. Sack, and E. T. Swanson (2005): "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking*, 1, 55–93.

Hahn, J., G. Kuersteiner, and M. Mazzocco (2020): "Estimation with Aggregate Shocks." *Review of Economic Studies*, 87, 1365–1398.

Hahn, J., G. Kuersteiner, A. Santos, and W. Willigrod (2024): "Overidentification in Shift-Share Designs," Working Paper arXiv:2404.17049.

Hansen, L. P. and R. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." *Journal of Political Economy*, 88, 829–853.

Herbst, E. P. and B. K. Johannsen (2023): "Bias in Local Projections." Working paper.

Imbens, G. W. and M. Kolesár (2016): "Robust Standard Errors in Small Samples: Some Practical Advice." *Review of Economics and Statistics*, 98, 701–712.

Jarociński, M. and P. Karadi (2020): "Deconstructing Monetary Policy Surprises—The Role of Information Shocks." *American Economic Journal: Macroeconomics*, 12, 1–43.

Jeenas, P. and R. Lagos (2024): "Q-Monetary Transmission." *Journal of Political Economy*, 132, 971–1012.

Jordà, O. (2009): "Simultaneous Confidence Regions for Impulse Responses." *The Review of Economics and Statistics*, 91, 629–647.

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

Lazarus, E., D. J. Lewis, and J. H. Stock (2021): "The Size-Power Tradeoff in HAR Inference." *Econometrica*, 89, 2497–2516.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018): "HAR Inference: Recommendations for Practice." *Journal of Business and Economic Statistics*, 36, 541–559.

Leeper, E. M., C. A. Sims, and T. Zha (1996): "What Does Monetary Policy Do?." *Brookings Papers on Economic Activity*, 27, 1–78.

Lusompa, A. (2023): "Local Projections, Autocorrelation, and Efficiency." *Quantitative Economics*, 14, 1199–1220.

Majerovitz, J. and K. A. Sastry (2023): "How Much Should We Trust Regional-Exposure Designs?." Working paper.

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

Montiel Olea, J. L. and M. Plagborg-Møller (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, E. Qian, and C. K. Wolf (forthcoming): "Double robustness of local projections and some unpleasant VARithmetic." *Econometrica*.

Müller, U. K. (2004): "A theory of robust long-run variance estimation." Working paper.

Nakamura, E. and J. Steinsson (2014): "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review*, 104, 753–792.

——— (2018): "Identification in Macroeconomics." *Journal of Economic Perspectives*, 32, 59–86.

Newey, W. K. and K. D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation." *The Review of Economic Studies*, 61, 631–653.

Nunn, N. and N. Qian (2014): "US Food Aid and Civil Conflict." *American Economic Review*, 104, 1630–1666.

Ottonello, P. and T. Winberry (2020): "Financial Heterogeneity and the Investment Channel of Monetary Policy." *Econometrica*, 88, 2473–2502.

Pakel, C. (2019): "Bias reduction in nonlinear and dynamic panels in the presence of cross-section dependence." *Journal of Econometrics*, 213, 459–492.

Pesaran, M. H. (2006): "Estimation and inference in large heterogeneous panels with a multifactor structure." *Econometrica*, 74, 967–1012.

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

Rambachan, A. and N. Shephard (2021): "When do common time series estimands have nonparametric causal meaning?" Working paper.

Ramey, V. (2016): "Macroeconomic shocks and their propagation." in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, chap. 2.

Stock, J. H. and M. W. Watson (2016): "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics." in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, chap. 8.

——— (2018): "Identification and estimation of dynamic causal effects in macroeconomics using external instruments." *Economic Journal*, 128, 917–948.

Sun, Y. (2014): "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference." *Journal of Econometrics*, 178, 659–677.

Sun, Y., P. C. B. Phillips, and S. Jin (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing." *Econometrica*, 76, 175–194.

Winkler, V. (2026): "When and Why State-Dependent Local Projections Work." Working Paper arxiv:2601.01622.

Xu, K.-L. (forthcoming): "Local Projection-Based Inference under General Conditions." *Review of Economic Studies*.