Programa de Doctorado en Economía y Gobierno

Tesis Doctoral

Essays on the econometrics of heterogeneous agents: finite populations, macro shocks, and subjective expectations

Víctor Sancibrián Lana

Centro de Estudios Monetarios y Financieros

2025

Programa de Doctorado en Economía y Gobierno

Tesis Doctoral

Essays on the econometrics of heterogeneous agents: finite populations, macro shocks, and subjective expectations

Víctor Sancibrián Lana

Director:

Manuel Arellano González

Centro de Estudios Monetarios y Financieros

2025

A mi madre y a mi padre

AGRADECIMIENTOS

Sobra decir que ninguno de los méritos de esta tesis me pertenece por completo. Irónicamente, esta es a su vez el resumen de una etapa larga que culmina en puerto de montaña: el doctorado. Como tal, deja entrever, pero no resalta, el papel indispensable de todos aquellos que han sido parte esencial del recorrido. Estas líneas pretenden dejar constancia de ello.

Con Manuel Arellano mi deuda es incalculable. Su influencia en mi formación como económetra y en mi modo de entender la disciplina es difícil de exagerar, y cualquier posible virtud de esta tesis lleva su impronta. Me cuesta encontrar palabras que dignifiquen la excepcional combinación de sabiduría y generosidad que entalla Manolo. De un lado, encarna el equilibrio entre precisión y vocación por la sustancia; de otro, la empatía y la paciencia, que he agradecido enormemente en tantas fases de una aventura como el doctorado. Contarme entre sus estudiantes es uno de los grandes privilegios de mi vida.

Dmitry Arkhangelsky también ha tenido un papel fundamental en mi paso por el doctorado. En él he encontrado a una suerte de codirector en la práctica, a un apasionado de la investigación y a un colega siempre entusiasta a la hora de compartir ideas. Entre mis mentores cuento también a Enrique Sentana, incisivo y siempre dispuesto a ayudar, y a Mikkel Plagborg-Møller, por cuya hospitalidad en la Universidad de Princeton y apoyo durante la fase final del doctorado le estoy profundamente agradecido. Y, por supuesto, a Tincho Almuzara, que ha terminado siendo también un amigo y un coautor del que aprendo cada día.

El CEMFI se ha convertido en un lugar muy especial para mí. Tras las puertas de Casado del Alisal 5 se genera un ambiente enormemente estimulante que va de la mano de la calidad humana de todos sus miembros. Me gustaría mencionar a todo el equipo no académico, que no son sino los cimientos del CEMFI: Miguel, Nathalia, Eva, María, Javier L., Javier P., Daniel, Eloy, Sonia, Isa, Araceli —mención especial por su inestimable ayuda con todos los asuntos relacionados con el doctorado—, Gema, Irene, Ulises y Teresa.

La convivencia con mis compañeros de promoción y con los demás doctorandos ha sido una de las dimensiones más gratificantes de esta etapa. Resalto a Pello Aspuru y a Sevin Kaytan, y, especialmente, a Vedant Agarwal, con quien he compartido muchas noches, sábados y domingos en la oficina a lo largo de estos años.

Finalmente, mi más sentido agradecimiento va para aquellos que forman mi círculo más cercano. A mis amigos: David y Andrea, Marc, Josep, Juanje, Kike, y todos aquellos que han sido, de una u otra manera, mi anclaje a la vida fuera del doctorado. A Elena, por acompañarme en este viaje. A mi hermano, Oasis. A mi abuela Carmen y a mi tía Chose. Y, especialmente, a mi madre, Irantzu, por su lucha y sacrificio incansables para salir adelante y por su apoyo incondicional; y a mi padre, Juan, por inculcarme la curiosidad por el saber y el interés por la lectura. Esta tesis va por vosotros.

I acknowledge financial support from PRE2022-000906 grant, funded by MICIU/AEI/ 10.13039/501100011033 and by "ESF+"; the Maria de Maeztu Unit of Excellence grants CEMFI MDM-2016-0684 and CEMFI CEX2020-001104-M, funded by MICIU/AEI/10.13039/501100011033; the PI2022-143184NA-100 grant, funded by MICIU/AEI/ 10.13039/501100011033 and by "ERDF/EU"; and the graduate scholarship grants funded by Fundación Ramón Areces and CEMFI.

RESUMEN

El uso de microdatos tiene una larga tradición en el trabajo empírico en economía, y entre sus hallazgos principales se encuentra la amplia heterogeneidad que existe en el comportamiento observado de los agentes. En paralelo, la literatura econométrica ha evolucionado para abordar algunos de los problemas metodológicos que surgen al analizar microdatos y al tratar o documentar la heterogeneidad. A pesar de estos esfuerzos, muchas preguntas continúan abiertas, alimentadas por el constante desarrollo de la investigación empírica.

En esta tesis, abordo algunas cuestiones metodológicas que surgen en contextos de microdatos ricos, los cuales aparecen de manera natural al trabajar con formas modernas de microdatos, como los censos o los datos administrativos o encuestas extensas que incluyen nuevas mediciones económicas. Como se verá, estos entornos suponen retos importantes para la teoría econométrica, con la heterogeneidad como un factor central de los mismos. Dichos retos abarcan desde la inferencia estadística y la cuantificación de incertidumbre hasta la identificación y estimación de modelos económicos con nuevos tipos de datos.

En particular, estudio aquí tres de estos contextos: entornos de *población finita*, donde todas (o la mayoría de) las unidades de interés se muestrean; la econometría de *shocks agregados* al usar microdatos para responder preguntas macro; y el uso de nuevas medidas económicas como datos de *expectativas subjetivas*. La heterogeneidad microeconómica está íntimamente ligada a estos temas: se conecta con la forma en que concebimos la aleatoriedad en los datos, define los objetos poblacionales que podemos recuperar, y orienta el desarrollo de nuevas formas de documentar y resumir regularidades empíricas.

En el Capítulo 1, presento brevemente estos temas en el marco de un modelo simple de coeficientes aleatorios para datos de panel. De hecho, un tema recurrente a lo largo de la tesis es que la disponibilidad de observaciones repetidas permite desarrollar métodos econométricos prácticos para abordar estas cuestiones.

En el Capítulo 2 (*Estimation uncertainty in repeated finite populations*), estudio las situaciones de población finita —aquellas donde la muestra es una fracción significativa de la población de interés—, que se dan frecuentemente con datos censales o administrativos, como aquellos que contienen registros casi universales de empresas o trabajadores. Considero el caso empíricamente relevante de una población finita que coexiste con un problema de medición, de modo que las características de interés no son necesariamente observables aún cuando se muestree la población completa. La práctica convencional en estos entornos consiste en directamente no comunicar mediciones de incertidumbre estadística o en proceder como si la muestra fuese una fracción infinitesimal de una superpoblación hipotética. En este capítulo, muestro que esto o bien ignora el problema de medición, o bien conduce a inferencia conservadora, en un grado que depende de cuánta heterogeneidad poblacional exista en dichas características latentes de interés.

A continuación, propongo Correcciones de Población Finita (*Finite Population Corrections*, FPCs) que garantizan inferencia no conservadora cuando se dispone de mediciones repetidas. Las FPCs requieren dependencia débil entre mediciones, como la que implican procesos de medias móviles, y son muy fáciles de implementar mediante restricciones de covarianza. Finalmente, aplico estos métodos a dos contextos empíricos donde la incertidumbre se ha venido entendiendo de distintas maneras: la predicción de encuentros letales con la policía usando datos de todos los departamentos policiales de EE.UU., y el estudio de la asignación ineficiente de *inputs* de empresas con un censo de empresas grandes en Indonesia. La inferencia de población finita produce intervalos de confianza hasta un 50% más cortos en el primer caso y muestra la necesidad de incorporar la incertidumbre de medición en el segundo.

El Capítulo 3, titulado *Micro responses to macro shocks*, es un trabajo conjunto con Martín Almuzara. Surge motivado por la creciente literatura empírica sobre la transmisión heterogénea de la incertidumbre agregada, como los cambios en la política monetaria, a variables microeconómicas. A pesar de su popularidad, se sabía poco sobre las propiedades estadísticas de los estimadores de *local projections* para funciones impulso-respuesta con datos de panel. De hecho, en nuestra propia revisión de unos cincuenta trabajos empíricos recientes, encontramos gran dispersión en cómo se calculan los errores estándar y posturas muy diferentes en cuanto al papel de cada dimensión del panel para cuestiones de precisión estadística y la importancia de la variación agregada en los datos. Motivados por esta observación, consideramos procesos generadores de datos que permiten una relación señal-ruido de los *shocks* macro en los microdatos sin restricciones.

Así, caracterizamos los objetos poblacionales cuando las respuestas a impulsos son heterogéneas y proponemos un método de inferencia con validez uniforme sobre dicha relación señal-ruido. Basta con incluir rezagos como variables de control y luego agrupar los datos (*cluster*) a nivel temporal. Complementamos estos resultados con una aplicación empírica sobre el papel de la heterogeneidad de las empresas y las fricciones financieras en la propagación de la política monetaria, ilustrando cómo los métodos de inferencia más populares pueden diferir sustancialmente del nuestro, que es (asintóticamente) robusto.

El Capítulo 4, titulado *Estimating flexible income processes from subjective expectations data: evidence from India and Colombia*, se basa en un trabajo conjunto con Manuel Arellano, Orazio Attanasio y Sam Crossman. Comienza con la observación de que muchas encuestas recaban información sobre escenarios hipotéticos o creencias subjetivas que resultan directamente informativas sobre el riesgo y la incertidumbre tal como los perciben los hogares. En particular, nos motiva la literatura sobre procesos de ingresos, que tradicionalmente se basa en las realizaciones efectivas de ingresos para caracterizar el riesgo y la persistencia.

Desarrollamos una metodología para modelar procesos de ingresos cuando se dispone de evaluaciones probabilísticas subjetivas sobre ingresos futuros, lo que permite estimar de manera flexible las *cdf* condicionales usando dichas probabilidades y obtener mediciones empíricas de riesgo y persistencia subjetivos. Luego, aplicamos esta metodología a dos encuestas longitudinales llevadas a cabo en zonas rurales de India y Colombia. Nuestros resultados indican que los procesos de ingresos lineales se rechazan en favor de versiones más flexibles en ambos casos; y que las distribuciones subjetivas de ingresos presentan heterocedasticidad, asimetría condicional y persistencia no lineal.

Finalmente, el Capítulo 5 contiene las *Conclusiones* de la tesis, donde recapitulo sus principales aportes, reflexiono sobre los desafíos y oportunidades que plantean los entornos modernos de microdatos y subrayo la importancia de mantener la heterogeneidad como elemento central en la investigación econométrica.

SUMMARY

Economists have long embraced microdata in empirical research. A central insight from this work is the widespread heterogeneity in the observed behavior of economic agents. In parallel, the econometrics literature has evolved to address some of the methodological issues that arise in analyzing microeconomic data and dealing with or documenting heterogeneity. Despite these efforts, many questions remain open, fueled by an ever-evolving body of empirical research.

In this thesis, I address some methodological questions posed by rich microdata environments that arise naturally when dealing with modern forms of microdata, such as census or administrative data or large surveys that include new economic measures. As will become clear, these are challenging environments from the standpoint of econometric theory, and heterogeneity lies at the core of many of these issues. These range from questions of statistical inference and valid uncertainty quantification to identification and estimation of economic models with new types of data.

In particular, here I study three such setups: *finite population* environments, where all (or most) units of interest are sampled, the econometrics of *aggregate shocks* when microdata is used to answer macro questions, and the use of new economic measures such as *subjective expectations* data. Microeconomic heterogeneity is deeply connected to these setups: it is intrinsically linked to the way we think of randomness in our data, shapes the population-level objects we can recover, and guides the development of new ways to document and summarize empirical regularities.

I briefly introduce these topics in Chapter 1 through the lens of a simple panel data random coefficients model. Indeed, a recurrent theme throughout the thesis is that the availability of repeated observations for the same units helps us develop new, practical econometric methods to address these questions.

In Chapter 2 (*Estimation uncertainty in repeated finite populations*), I study finite population setups — those where the sample is a large fraction of the population of interest. This is a prevalent feature of many census or administrative datasets, such as those containing nearly universal records of firms or workers. I consider the empirically relevant case where a finite population coexists with a measurement problem, in that the features of interest are not necessarily observable even if the entire population is sampled. Conventional practice in these setups is to either ignore uncertainty quantification altogether or proceed as if the sample were an infinitesimal fraction of a hypothetical superpopulation. I show that this either disregards the presence of a measurement problem or results in conservative inference, the extent to which depends on how heterogeneous these latent features of interest are in the population.

Then, I propose Finite Population Corrections (FPCs) that guarantee non-conservative inference when

repeated measurements are available. FPCs rely on weak dependence across measurements, such as that implied by moving-average processes, and are very easy to implement through covariance restrictions. Finally, I apply these methods to two empirical settings where uncertainty has been previously understood in different ways: predicting lethal encounters with police using data on all U.S. police departments, and studying firm misallocation with a census of large Indonesian firms. Finite-population inference leads to confidence intervals that are up to 50% shorter in the former and illustrates the need to account for measurement uncertainty in the latter.

Chapter 3 is titled *Micro responses to macro shocks* and is based on joint work with Martín Almuzara. This is motivated by the increasing empirical literature on the heterogeneous transmission of aggregate uncertainty, such as changes in monetary policy, to microeconomic outcomes. Despite its popularity, little was known about the statistical properties of panel local projections estimators of impulse responses in this context. Indeed, in our own survey of around fifty recent empirical papers, we document large dispersion in the way practitioners compute standard errors and vastly different stances on the role of each dimension of the panel for precision and the importance of aggregate variation in the data. Motivated by this observation, we consider data generating processes that allow for an unrestricted signal-to-noise ratio of macro shocks in the microdata.

We characterize the population objects when impulse responses are heterogeneous and provide a recipe for uniformly valid inference over signal-to-noise. This simply entails including lags as controls and then clustering at the time level. We complement our results with an empirical application to the role of firm heterogeneity and financial frictions in the propagation of monetary policy and illustrate how popular inference alternatives can deviate substantially from our (asymptotically) robust procedure.

Chapter 4 is titled *Estimating flexible income processes from subjective expectations data: evidence from India and Colombia* and is based on joint work with Manuel Arellano, Orazio Attanasio and Sam Crossman. It starts with the observation that many questions in household surveys elicit information on hypothetical scenarios or subjective beliefs that are directly informative about risk and uncertainty as perceived by households. In particular, we are motivated by the literature on income processes, which has long relied on income realizations to characterize income risk and persistence.

We develop a methodology for modeling household income processes when subjective probabilistic assessments of future income are available, which allows us to flexibly estimate conditional *cdf* s directly using elicited individual subjective probabilities, and to obtain empirical measurements of subjective risk and persistence. We then use two longitudinal surveys collected in rural India and rural Colombia to explore the nature of income dynamics in those contexts. Our results suggest linear income processes are rejected in favor of more flexible versions in both cases; subjective income distributions feature heteroskedasticity, conditional skewness, and nonlinear persistence.

Finally, Chapter 5 contains the *Conclusions* of the thesis. Here I revisit its main contributions, reflect on the challenges and opportunities afforded by modern microdata environments, and emphasize the importance of keeping heterogeneity at the forefront of econometric research.

CONTENTS

CHAPTER 1

INTRODUCTION

Economists have long welcomed the use of microdata, which contains disaggregated information on economic agents such as households or firms. The wealth of information provided by large surveys, census or administrative datasets enables the study of complex relationships among agents, documenting patterns and regularities among otherwise different units. There is now consensus that heterogeneity in observed behavior is widespread, and explicitly taking it into account has become deeply ingrained for anyone working with microdata.[1]

Such data often have a longitudinal structure that tracks the same units over time, and it becomes possible to learn about dynamic effects or to tell apart permanent unobserved heterogeneity from state dependence. In parallel, a well-established econometrics literature has laid the foundations for sound statistical analysis, with an emphasis on modeling and dealing with heterogeneity (Arellano, 2003).

This thesis deals with econometric challenges in *rich* microdata environments. This does not necessarily refer to the dimensionality of the data but to its characteristics, to new applications and to new forms of data, and to how these interact with the potentially heterogeneous nature of the underlying units. Here I study three such setups: *finite population* environments, where all (or most) units of interest are sampled, the econometrics of *aggregate shocks* when microdata is used to answer macro questions, and the use of new economic measures such as *subjective expectations* data. Chapters 2, 3, and 4 deal with each of these topics, respectively.

In this chapter, I briefly introduce the setup under the umbrella of repeated measurement models. Indeed, it is the repeated observation of certain features for the same units that helps us develop new, practical econometric methods to address these methodological challenges. I then summarize some of the particularities of each setup and outline the methodological and empirical contributions of the thesis.

[1] In his Nobel lecture, Heckman (2001) states "the most important discovery [from the use of microeconomic data is] the evidence on the pervasiveness of heterogeneity and diversity in economic life"; see also Browning and Carro (2007).

Repeated measurement models

Consider a setup where the researcher observes outcomes Y_{it} and a vector of characteristics X_{it} ($\dim X_{it} = q$) for units $i = 1, \ldots, N$ and over measurements $t = 1, \ldots, T$, often corresponding to a time dimension. Outcomes and characteristics are sampled from an unspecified probability distribution and are related according to

$$Y_{it} = X_{it}'\theta_i + v_{it}, \tag{1.1}$$

with zero conditional mean errors $E\left[v_{it}|X_i, \theta_i\right] = 0$, where $X_i = (X_{i1}, \ldots, X_{iT})'$ are strictly exogenous.

The presence of unit-specific coefficients θ_i together with the zero conditional mean assumption defines a random coefficients model in a panel data context. For instance, if $q = 2$ and X_{it} contains a constant, equation (1.1) allows for a heterogeneous level and slope in the relationship between Y_{it} and a given characteristic of interest. Put differently, these can be seen as repeated — but noisy — measurements of θ_i for each unit.

In the context of equation (1.1), interest is often on the effect of an exogenous change in X_{it} on Y_{it} summarized by the average partial effect (APE), which can be defined as:

$$\beta \equiv E\left[\frac{\partial E\left[Y_{it}|X_i = x, \theta_i\right]}{\partial x_t'}\right] = E\left[\theta_i\right], \tag{1.2}$$

where the outer expectation is taken with respect to the distribution of θ_i over the population of interest. Depending on the setup, these partial effects have different economic content and interpretation. Below we will see three different such cases.

Under a correlated fixed effects approach, where the dependence of θ_i on X_i is left unrestricted, the simple projection of Y_{it} on X_{it} will generally not recover the APE; this is often known as fixed effects endogeneity. If interest is in keeping unit-specific responses unrestricted, one possibility is to average over unit-by-unit least squares estimators as follows. Assuming that $T \geq q$ and $\det X_i'X_i \neq 0$, let

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i'X_i\right)^{-1}X_i'Y_i, \tag{1.3}$$

where $Y_i = (Y_{i1}, \ldots, Y_{iT})'$. This simple yet powerful tool is referred to as the between-groups estimator (see, for instance, Arellano and Bonhomme, 2012). To gain additional insight, it is worth decomposing the estimation error as

$$\hat{\beta} - \beta = \left(\frac{1}{N}\sum_{i=1}^{N}\theta_i - \beta\right) + \frac{1}{N}\sum_{i=1}^{N}\left(X_i'X_i\right)^{-1}X_i'v_i, \tag{1.4}$$

where $v_i = (v_{i1}, \ldots, v_{iT})'$. Given the strict exogeneity assumption, it is easy to see that the last term is zero mean, and so is the first one under mild regularity conditions on sampling. Many of the methodological contributions in this thesis relate to the concentration properties of these two terms in different environments; namely, how fast their variances shrink as more and more data accumulates. In Chapter 2, I consider the situation where the analyst has access to a sample that is a non-negligible fraction of the population of interest, whereas in Chapter 3 I explore estimation and inference in a context APEs to macro-level characteristics X_t are of interest. In Chapter 4, instead, units are allowed to have different responses depending on the level of the characteristics themselves, and we leverage new measurements of the outcomes of interest to recast dynamic panel data models into models similar to those introduced here.

Sampling entire populations

In Chapter 2, titled *Estimation uncertainty in repeated finite populations*, I consider the empirically relevant case where the sample at hand is a non-negligible fraction of the population of interest. A relevant particular case is when the entire population is sampled; consider having data on all 50 U.S. states or drawing nearly universal records on firms or workers.

This is clearly not the textbook case where the data are treated as an infinitesimal sample from a much larger, hypothetical superpopulation. In fact, it is intuitive to expect sampling uncertainty to disappear as a larger fraction of the population is sampled, which raises the question of how (and whether) to perform statistical inference in this context. Indeed, routine practice is to either ignore this additional information on the sampling process and proceed as if sampling from a superpopulation, or to not report measures of uncertainty quantification altogether.

I consider the case where a finite population coexists with a measurement problem, in that the features of interest are not necessarily observable even if the entire population is sampled. A way to operationalize this is via the repeated measurement model in equation (1.1), where only noisy counterparts of the unit-specific responses θ_i can be recovered. Consider a population of size n described by features $\{\theta_i\}_{i=1}^{n}$; the APE in equation (1.2) then becomes $\beta = n^{-1} \sum_{i=1}^{n} \theta_i$, and the estimation error is

$$\hat{\beta} - \beta = \left(\frac{1}{N} \sum_{i=1}^{N} \theta_i - \frac{1}{n} \sum_{i=1}^{n} \theta_i \right) + \frac{1}{N} \sum_{i=1}^{N} \left(X_i' X_i \right)^{-1} X_i' v_i.$$

It is easy to see that when $N = n$, the first term disappears. More generally, the “size” of the first term depends on the sample-to-population fraction $f = N/n$ and on how heterogeneous is the population.

I show that conventional standard errors remain generally conservative in this context and propose Finite Population Corrections (FPCs) that guarantee non-conservative inference for any sample-to-population fraction f when repeated measurements are available. Intuitively, conventional practice ignores an important piece of information about the sampling framework; internalizing it allows us to sharpen statistical inference. FPCs rely on weak dependence across measurements and are very simple to implement via covariance restrictions on v_{it}.

Finally, I apply these methods to two empirical settings where uncertainty has been previously understood in different ways: predicting lethal encounters with police using data on all U.S. police departments, and studying firm misallocation with a census of large Indonesian firms. Whereas uncertainty quantification is first-order in the prediction exercise, inference has often been overlooked in the latter, where it is common to sample all firms in a given sector or country. Finite-population inference allows for a systematic approach to uncertainty quantification in these cases: it accounts for the presence of a measurement problem, while at the same time leads to substantially more precise and meaningful assessments compared to conventional, superpopulation-based confidence intervals.

Dealing with aggregate shocks

Chapter 3 is titled *Micro responses to macro shocks* and is based on coauthored work with Martín Almuzara. The starting point is the growing empirical literature that is interested in the transmission of aggregate uncertainty to individual outcomes. Examples include the effectiveness of monetary policy of the consequences of business cycle, TFP or oil price shocks, and entail the use of microdata in order to study how these shocks trickle down through the economy or how heterogeneity shapes these responses.

Through the lens of the random coefficients model in equation (1.1), this is a setup where the regressor of interest is common to all units and only varies over time, that is, $X_{it} = X_t$. (Set $q = 1$ for simplicity and $X = (X_1, \ldots, X_T)'$.) Furthermore, since it is hard to ex-ante rule out the presence of other, unobserved aggregate shocks Z_t with possibly heterogeneous exposures γ_i, a simple representation of the model becomes

$$\begin{aligned} Y_{it} &= X_t \theta_i + v_{it}, \\ v_{it} &= Z_t \gamma_i + u_{it}, \end{aligned} \tag{1.5}$$

where u_{it} are unobserved idiosyncratic shocks. In most application, X_t represents the (perhaps noisy) measurement of a shock, and so exogeneity assumptions along the lines of those discussed above are reasonable. In this context, the simple panel regression of Y_{it} on X_t is equivalent to the between-groups estimator in (1.3), so that the estimation error representation in (1.4) is still of use:

$$\hat{\beta} - \beta = \left(\frac{1}{N} \sum_{i=1}^{N} \theta_i - \beta \right) + \frac{1}{N} \sum_{i=1}^{N} \left(X'X\right)^{-1} X' \left(Z\gamma_i + u_i\right),$$

with obvious definitions. Visually inspecting the second term then suggests that the nature of the estimation error is intrinsically linked to the presence of aggregate shocks and to how sizable they are relative to micro shocks.

In Chapter 3, we generalize these ideas to a more general context where interest is in the dynamic response of Y_{it} to an exogenous change in X_t over different horizons $h = 1, \ldots, H$; $\hat{\beta}$ can be seen as the contemporaneous impulse response ($h = 0$). In this context, APEs are average impulse responses over heterogeneous units and least-squares estimators are known as (panel) local projection estimators (Jordà, 2005). Despite its popularity, little was known about the statistical properties of these estimators. In our own empirical recollection of about 50 recent papers, we document large dispersion in the way practitioners compute standard errors, which is symptomatic of vastly different implicit stances on the importance of aggregate variation in the data.

We then study estimation and inference in a general context where the signal-to-noise of macro shocks in the microdata is left unrestricted. In particular, we scale the variance of idiosyncratic noise u_{it} in (1.5) by a parameter that can drift with the sample size, which allows for regimes where the signal value of aggregates may be arbitrarily low (or high) in the limit. In this framework, we show how to obtain standard errors that are uniformly valid over this parameter. Finally, we illustrate our methods in an empirical application to the role of financial frictions and firm heterogeneity in the transmission of monetary policy in the U.S.

Leveraging subjective expectations data

Another example of rich microdata environments are survey datasets that contain both standard variables — say, information on household income, consumption and demographics — with subjective expectations questions, which directly ask respondents for their own assessment of a future or uncertain event. In Chapter 4, titled *Estimating flexible income processes from subjective expectations data: evidence from India and Colombia*, we explore the identifying content of such new economic measures. This is based on joint work with Manuel Arellano, Orazio Attanasio and Sam Crossman.

This is motivated by the literature on income processes, which has traditionally relied on income realizations to shed light on the heterogeneous dynamics, persistence, and income risks faced by households when making consumption and saving decisions. Such an indirect approach requires both assumptions about the nature of expectation formation processes and modeling the dynamics of realized variables by collecting income realizations over several panel waves. In terms of the model in equation 1.1, we let y_{it} denote log household income, and set $X_{it} = (1, y_{i,t-1})$ and $Y_{it} = y_{it}$, which leads to the canonical panel autoregressive income process

$$y_{it} = \theta_{i1} + \theta_2 y_{i,t-1} + v_{it},$$

where we are setting $\theta_{i2} = \theta$ for simplicity. Note that in this context strict exogeneity is unrealistic, and, generally speaking, at least $T = 3$ is needed to identify the APE $\beta = \theta_2$, which in this context is a measure of persistence of the income process.

Instead, suppose that households are asked to report their own subjective assessment p_{it} that their income next period will not exceed a given threshold r_{it}. This is directly informative on the uncertainty faced by households on their future income as they perceive it. Furthermore, the statistical problem itself is different, since the stochastic nature of the relationship between p_{it} and $y_{i,t-1}$ is now tied to errors in the elicitation process. It turns out that letting $Y_{it} = h(p_{it})$ (for some given transformation $h(\cdot)$) and $X_{it} = (1, r_{it}, y_{i,t-1})$ it is possible to recast the problem into a static panel data model with strictly exogenous regressors.

In Chapter 4, we generalize these ideas to setups where several probabilistic assessments of future income are available and where we incorporate heterogeneity in responses by allowing for nonlinear income processes where persistence itself might depend on $y_{i,t-1}$ or the size of hypothetical income shocks. We develop a methodology that combines data on income expectations and realizations to flexibly estimate conditional *cdf* s, which allows us to obtain empirical measurements of subjective risk and persistence that might differ across households depending on both observed and unobserved heterogeneity.

In our empirical analysis, we exploit subjective expectations data from two surveys conducted in rural India and Colombia to document the characteristics of (perceived) income processes in a developing context. Our results suggest sizeable non-linearities and dynamics that are not compatible with canonical income processes, such as the presence of substantial heteroskedasticity, conditional skewness and nonlinear persistence.

CHAPTER 2

ESTIMATION UNCERTAINTY IN REPEATED FINITE POPULATIONS

2.1 Research context

Empirical researchers are often interested in features of finite populations — those for which all or a non-negligible number of units are sampled: all schools in a district, most households in a village, nearly universal records on firms or workers...

When these features are directly observable upon sampling, the usual standard error formulas need to be adjusted down to reflect this abundance of information.[1] This is, however, of limited applicability in many relevant problems: school quality might not be directly observable even if all schools of reference were to be sampled; instead, we might only have access to imperfect measurements such as average test scores for different student cohorts. Similar ideas apply to learning about household-level preferences or about the frictions firms face in a particular sector.

In this chapter, I propose new methods to assess estimation uncertainty in a framework where a finite population coexists with a measurement problem — where even if we observe the entire population, we may only have access to a few noisy measurements of the underlying attributes of interest. I show that conventional inference methods remain generally conservative in this context and propose Finite Population Corrections (FPCs) that lead to asymptotically correct inference for any sample-to-population fraction. FPCs rely on weak dependence across measurements and are very simple to implement.

I apply these methods to two empirical settings where uncertainty has been previously understood in different ways: predicting lethal encounters with police using data on all U.S. police departments, and studying firm misallocation with a census of large Indonesian firms. Inference is of primary interest in the former, and I show that FPCs lead to up to 50% shorter confidence intervals. Inference is usually second-order in the latter, where full-population datasets are common. When a measurement problem is nonetheless present,

[1]Such results belong to a long-standing statistical literature; Cochran (1977) is a classical reference. The earlier work was done in the context of survey sampling, see for instance Neyman (1934), Hansen and Hurwitz (1943) and Horvitz and Thompson (1952).

finite-population confidence intervals correctly reflect the dominant source of estimation uncertainty.

Setup and scope for empirical work. The methods in this chapter are relevant for applications that share two key ingredients. First, there is a well-defined population of units indexed by a set of characteristics — or attributes — and the object of interest is defined over this population: say, an average response coefficient or a regression parameter. [2] The analyst has access to a (random) sample of units from this population. Using a sample to learn about a population introduces *sampling uncertainty*; the extent of this is determined by the sample-to-population fraction $f \in [0, 1]$. Here $f = 1$ captures full population setups, such as with data on all U.S. counties, whereas $f = 0$ is appropriate for CPS data, where it is reasonable to view it as a random sample from a much larger, "infinite" population.

Second, some of these attributes might remain unobserved even if a given unit is sampled; instead, a few error-ridden measurements of the underlying attributes are available. These are then used to estimate the parameter of interest, introducing *measurement uncertainty*. We will require that these are "good measurements" in the sense that it is possible to construct unbiased estimators of the underlying attributes. For this purpose, I consider a general class of measurement models that are affine in the underlying attributes of interest, analogous to random coefficient models in the panel data literature (Chamberlain, 1992; Arellano and Bonhomme, 2012). This is a different setup from the one considered in experimental analyses in finite populations, where uncertainty is induced by treatment randomization (Neyman, 1923/1990; Abadie, Athey, Imbens, and Wooldridge, 2020). My setup is in the model-based tradition where policy variation is not exploited for inference.

These two ingredients are prominent in the two empirical applications I consider. The first one is based on Montiel-Olea, O'Flaherty, and Sethi (2021), who draw from records on all local police departments in the U.S. to study the determinants of police use of deadly force and conduct prediction exercises involving these agencies. Some of these predictors are directly observable (such as regional laws), while others are not (such as departmental culture). The authors use a panel of lethal encounters over 2013–2018 and a measurement system analogous to a heterogeneous Poisson model to disentangle their separate effects.

The second is in the spirit of a large literature following Hsieh and Klenow (2009). In a nutshell, firms face frictions that prevent them from choosing their inputs optimally, and these translate into firm-specific "wedges" in marginal products relative to the optimal allocation. Interest is here on investigating how these frictions relate to firm characteristics or on quantifying their cross-sectional dispersion, which is directly informative on aggregate TFP losses from misallocation. Measuring these underlying frictions is challenging: I use census panel data for manufacturing Indonesian firms from Peters (2020) and consider a persistent–transitory (fixed-effects) decomposition, following recent approaches in the literature (David and Venkateswaran, 2019; Chen, Restuccia, and Santaeulàlia-Llopis, 2022; Adamopoulos, Brandt, Leight, and Restuccia, 2022; Nigmatulina, 2023).

More generally, the analysis here is relevant for a large class of problems involving latent variables, fixed effects, factor models and random coefficient models.[3] Note that repeated measurements need not have a

[2] Sometimes it is not obvious whether one should adopt a finite-population perspective or treat the sample as drawn from a larger population that includes new, hypothetical units. This might be a useful conceptual exercise, see the discussion in Section 3.2 (Remark 2.1). The methods in this chapter allow to quantify and decompose estimation uncertainty under both approaches.

[3] Additional examples include heterogeneous earnings profiles (Guvenen, 2009), school or teacher value-added models

clear time ordering as in panel data; measurements over space or parallel measurements are also common. For instance, Kline et al. (2022) are interested in studying firm-level discrimination for a finite population of 108 Fortune 500 U.S. firms and have job-level repeated measurements for each company.

Finite-population inference. I propose consistent variance estimators for a fixed number of measurements in a method-of-moments framework that incorporates these two ingredients. The parameters of interest include finite-population estimands, defined by linear instrumental-variable moment conditions for the attributes of interest, and common parameters of the measurement system.[4]

The proposed finite-population variance estimator is constructed such that it accounts for sampling-based and measurement-based uncertainty in the right proportions, that is, it is indexed by $f \in [0, 1]$. In essence, it exploits a parallel between the *measurement–sampling* decomposition in the asymptotic variance of the estimator and a (generalized) *within–between* variance decomposition. The "within" part embeds the notion of measurement uncertainty: residual variation around the underlying latent attributes of interest. The "between" part captures the idea of sampling uncertainty: differences between sample and population attributes. The finite-population variance estimator weights the latter by $(1-f)$; this generalizes the standard Finite Population Correction (FPC) to problems where the features of interest are not directly observable upon sampling.

The within-between decomposition requires limited dependence across measurements. This is the main assumption in the chapter and the point of departure relative to conventional variance estimators. I specify weak dependence as linear restrictions on the covariance matrix of the measurement errors, such as those implied by m-dependent processes.[5] In other words, measurement errors should be not too dependent relative to the number of measurements, and fewer restrictions are needed as more measurements become available — a common notion of weak dependence in the time-series literature. This is also natural to many problems with repeated measurements; for instance, it is the key assumption in deconvolution problems, in which one is interested in the distributional characteristics of latent variables. Importantly, all other elements in the covariance matrix remain completely unrestricted and free to vary with observable and unobservable attributes. For instance, in the context of our school quality example, it is reasonable to assume that within variation in average test scores for different cohorts is uncorrelated, while we allow for dispersion to vary over cohorts and to select on school quality. The generalization of within-between decompositions to weak dependence of this form is established by Arellano and Bonhomme (2012) in the context of estimation of distributional characteristics of random coefficient models.

The resulting variance estimator takes the form of a FPC applied to the conventional "sandwhich" esti-

---

(Gilraine, Gu, and McMillan, 2022; Hahn, Singleton, and Yildiz, 2023), modelling skill and scalability in mutual funds (Barras, Gagliardini, and Scaillet, 2022), microforecasting (??Giacomini, Lee, and Sarpietro, 2023), state or country-level regressions (Villacorta, 2021), total factor productivity estimation (Klette, 1999; Combes, Duranton, Gobillon, Puga, and Roux, 2012), risk-sharing in village economies (Townsend, 1994; Schulhofer-Wohl, 2011; Chiappori, Samphantharak, Schulhofer-Wohl, and Townsend, 2014), firm-level discrimination audits (Kline, Rose, and Walters, 2022), meta-analyses (Meager, 2022), heterogeneity in returns to technology adoption in developing countries (Suri, 2011), schooling models as in Magnac, Pistolesi, and Roux (2018) or the difference-in-differences model in Bonhomme and Sauder (2011), elicitation of preferences and risk attitudes (Barsky, Juster, Kimball, and Shapiro, 1997; Andreoni and Samuelson, 2006; Ahn, Choi, Gale, and Kariv, 2014), low-rank models for time-varying treatment effects (Bonhomme and Denis, 2024), and many others.

[4] I extend the framework to nonlinear transformations of the latent attributes such as variances in Remark 2.8. Such objects are relevant in the empirical application in Section 2.5.2.

[5] This is a popular approach in minimum-distance estimation of covariance structures, see Arellano (2003, Chapter 5).

mator and is very simple to implement: FPCs only require to inversion of a selection matrix specifying the restrictions on the measurement part of the model. A drawback of the resulting estimator is that is that it is not guaranteed to be positive semi-definite in finite samples, although this is not a problem in the simulations and the empirical illustrations in this chapter. The presence of other observable attributes can be exploited to construct conservative estimators following similar proposals in the design-based literature (Fogarty, 2018; Abadie et al., 2020).

The asymptotic approximations in this chapter are established for a sequence of growing finite populations such that the limiting sample-to-population fraction remains representative of the sampling framework, an embedding often referred to as in the literature as finite-population asymptotics (Lehmann, 1975; Li and Ding, 2017; Abadie et al., 2020; Xu, 2021), and for a fixed number of measurements. Finite-population inference via FPCs is correct in large samples for any $f \in [0, 1]$, unlike conventional inference methods which implicitly set $f = 0$ and remain generally conservative. An exception are common parameters: intuitively, uncertainty about the measurement system itself should not depend on f, since the measurement problem is present regardless of the sampling framework. The same is true for finite-population estimands in the absence of heterogeneity in the underlying attributes of interest. Conversely, FPCs tend to be larger the more dispersed population attributes are, the less noise there is in the measurement system and the more measurements are available.

I complement these theoretical results with simulations for realistic designs to study the finite-sample properties of the proposed FPCs. The results show that finite-population inference maintains correct (nominal) coverage even for relatively small sample sizes ($N = 200$) and for different sample-to-population fractions, while the coverage probability for conventional confidence intervals is often one. The designs are calibrated to match reasonable signal-to-noise ratios (in the sense of relative weights of sampling and measurement components), which map to the relative width of conventional and finite-population confidence intervals in line with the discussion above.

Empirical illustrations. The framework developed in this chapter has practical implications for a wide range of applications. To illustrate this, I revisit two very different empirical problems: a prediction exercise where uncertainty quantification is first-order and an investigation of firm-level frictions and misallocation, where empirical moments are often reported without measures of estimation uncertainty.

In the first exercise, I apply Finite Population Corrections to the results in Montiel-Olea et al. (2021). Here the population of interest are local U.S. police departments, and the data comes from the Law Enforcement Agency Identifiers Crosswalk dataset (LEAIC), which compiles information on all state and local law enforcement agencies; here we set $f = 1$. The final dataset contains 7,585 agencies and information on the number of yearly lethal encounters with police and a number of covariates including local demographics, the number of officers per thousand inhabitants and state-level dummies on the severity of laws regarding officer misconduct. The authors posit an exponential model and obtain coefficient estimates via nonlinear least squares. Next, they propose a method to obtain predictions for counterfactual-like questions of the form "what would happen to number of lethal encounters if all 10 largest agencies had the department-specific attributes of the Chicago Police Department?"

Uncertainty is here of primary interest — more so than point prediction or statistical significance — and thus the authors directly report sampling-based confidence intervals (CIs). Finite-population inference instead identifies the right source of estimation uncertainty for this problem: the fact we only observed error-ridden measures of agency-specific baseline police violence. Applying Finite Population Corrections, I find that the conventional variance estimators were overly pessimistic about prediction uncertainty: standard errors for the projection coefficients are between 20% and 60% smaller, and this in turn leads to shorter prediction intervals. For instance, the conventional 90% CI for the question above is $(545, 700)$; the finite-population 90% CI is $(565, 667)$. From a policy perspective, the finite-population CI now excludes the realized number of lethal encounters during this period, whereas the effect remains ambiguous when treating the data as a negligible sample from a hypothetical population of U.S. local police departments.

In the second exercise, I apply these methods to a measurement exercise concerned with labor "wedges" for manufacturing Indonesian (formal) firms. I follow Peters (2020), who uses census data from Statistik Industri and focuses on young firms. The final dataset is an unbalanced yearly panel covering $17,000$ firms that enter the market over 1991–1999. Following recent contributions (David and Venkateswaran, 2019; Chen et al., 2022; Adamopoulos et al., 2022; Nigmatulina, 2023), I allow for measurement error in firm-level marginal revenue products of labor and focus on fixed-effect measures of wedges. I then explore the relationship between firm-level labor wedges and firm size upon entry, which might be suggestive of size-dependent policies and regulations (Guner, Ventura, and Yi, 2008). Similar exercises are commonplace in the literature (Yeh, Macaluso, and Hershbein, 2022; Gorodnichenko, Revoltella, Svejnar, and Weiss, 2021). I also extend the framework to cover wedge dispersion statistics, an often reported measure of "allocative efficiency" that can be mapped to the TFP loss from misallocation (Hsieh and Klenow, 2009).

Finite-population inference provides again a clear recipe for uncertainty quantification: despite having data on all firms we are interested in, measurement-based uncertainty needs to be accounted for. The results point at a very imprecise relationship between firm size and labor wedges for smaller firms — even more so if one were to calculate confidence intervals as if the sample was drawn from a superpopulation of firms. The emphasis on measurement problems also has implications for allocative efficiency calculations: fixed-effects measures revise down the TFP losses from misallocation of labor to about 15% from a (biased) baseline of around 20% on average across different size groups. Finite-population confidence intervals also suggest that this difference is statistically meaningful.

In essence, finite-population inference provides a unified approach to uncertainty quantification in problems where estimation uncertainty has been previously understood in very different ways.

Related literature. This chapter contributes to various strands of the literature.

First, it relates to the longstanding statistics literature on finite population analysis (Neyman, 1934; Hansen and Hurwitz, 1943; Horvitz and Thompson, 1952; Hájek, 1960; Erdős and Rényi, 1959; Li and Ding, 2017), which laid out the foundations of survey sampling and developed limit theorems under simple random sampling for growing sequences of finite populations; see Lehmann (1975) for a review. The focus of this literature has been on inference under sampling-based uncertainty, whereas I consider problems where sampling and measurement uncertainty coexist.

The discussion of measurement issues in this literature has revolved around the biases introduced by different forms of survey errors on estimation and prediction, see for instance Hansen, Hurwitz, and Bershad (1961) for an early contribution on (across units) interviewer bias. The literature has also noted the validity of standard variance formulas that ignore the presence of (classical) measurement error altogether as long as the

Finite Population Correction is negligible, see for instance Fuller (1995). Detailed treatments of measurement issues can be found in Cochran (1977, Chapter 13) and Mukhopadhyay (2001, Chapter 7). My focus is on inference with unbiased repeated measurements, which I exploit to propose consistent standard errors for a large class of empirically relevant models.[6]

Second, this chapter relates to the literature on design-based inference, which starting with Neyman (1923/1990) has been traditionally studied in a potential outcomes finite population context. These are (quasi)experimental setups where a source of uncertainty arises from randomized treatment assignment. In a context where both sampling and design uncertainty coexist, Neyman (1923/1990) noted the conservativeness of conventional variance estimators in a binary treatment setting. Later contributions generalized this setup to regression models with additional covariates, general sampling frameworks and assignment mechanisms, panel experiments and nonlinear models (Freedman, 2008; Rosenbaum, 2002; Abadie, Imbens, and Zheng, 2014; Fogarty, 2018; Abadie et al., 2020; Abadie, Athey, Imbens, and Wooldridge, 2023; Bojinov, Rambachan, and Shephard, 2021; Xu, 2021). An interesting set of extensions (Deeb and de Chaisemartin, 2022; Startz and Steigerwald, 2023, 2024) allows for stochastic potential outcomes (say, due to post-randomization aggregate shocks in an RCT); this is in the spirit of measurement-based uncertainty in a cross-sectional setting. Fogarty (2018), Abadie et al. (2020) and Xu (2021) also propose conservative finite-population variance estimators exploiting the predictive power of (observable) fixed attributes in different contexts.

I regard my framework as complementary to results in this tradition, both conceptually and in practice. The conceptual difference between measurement and experimental variation is clear, and which is more appropriate is application-specific. In practice, the two setups afford different tools for estimation and inference. For instance, exact inference for sharp nulls is possible if the analyst has access to the randomization distribution in the design world (Rosenbaum, 2002; Bojinov et al., 2021). In a model-based framework, I show that the availability of repeated measurements leads to asymptotically non-conservative finite-population inference.

Third, this chapter connects with the literature on fixed-T panel data random coefficient models. In particular, within-group and between-group transformations to deal with permanent unobserved heterogeneity are at the heart of this literature (Chamberlain, 1992; Arellano, 2003; Arellano and Bonhomme, 2012; Graham and Powell, 2012), and weak dependence over measurements has proved useful when estimation of distributional characteristics is of interest (Kotlarski, 1967; Arellano and Bonhomme, 2012). It turns out that Finite Population Corrections can be written as a variance over heterogeneous unit-level moment conditions; I leverage these insights to study inference in a finite population context.

Finally, the framework developed in this chapter allows us to reinterpret some of the existing results on inference in fixed-effects models as finite-population inference in the limit case with no sampling uncertainty. This is the case in the "many covariates" literature, which is concerned with linear regression models with a growing number of parameters (Cattaneo, Jansson, and Newey, 2018a,b; Kline, Saggio, and Sølvsten, 2020). Intuitively, removing the incidental parameters problem in this setup is analogous to removing sampling-based uncertainty. In practice, this only makes a difference for inference when the objects of interest involve

[6]The use of the term "measurement error" in my setup should be understood in a broad sense; it refers to any source of random variation that contaminates the underlying features of interest.

functions of the large-dimensional part of the model, as in Kline et al. (2020). Another example are average marginal effects in large-T nonlinear panel data models, which have been traditionally defined conditioning on the in-sample fixed effects, see Higgins and Jochmans (2024) for a recent contribution.

Outline. Section 3.2 builds intuition and illustrates the results in a simple example under simple random sampling. Section 3.3 generalizes the framework and presents the main results on finite-population inference. Section 2.4 discusses a comprehensive simulation study and Section 2.5 contains the two empirical applications. Proofs can be found in Appendix A.1.

2.2 Simple example

I first illustrate the main points of the chapter in a simple example where interest is in a population average but the outcome of interest is contaminated with independent measurement noise. For reference, it might help to think of estimating average school quality in a particular district using average test scores for different cohorts.

Setup. Consider a population of size n. Unit i in the population is indexed by a fixed attribute θ_i, and we are interested in the population average of θ_i:

$$\beta_n = E_n\left[\theta_i\right] \equiv \frac{1}{n}\sum_{i=1}^{n}\theta_i.$$

The task of the researcher is to obtain an estimate $\hat{\beta}$ of β_n together with a quantification of estimation uncertainty, such as a standard error or a confidence interval. Randomness in $\hat{\beta}$ might arise from two sources, what I refer to as sampling-based and measurement-based uncertainty. First, we might only have access to a representative sample from the population of interest, which we indicate via the vector of inclusion indicators $R_{1:n} = (R_1, \ldots, R_n) \in \{0,1\}^n$, where $R_i = 1$ indicates that unit i is sampled. Second, even if unit i is sampled, we might only observe noisy measurements $Y_i = (Y_{i1}, \ldots, Y_{iT})'$ of θ_i, so that in a given sample the analyst has access to $\{R_i, R_iY_i\}_{i=1}^{n}$. Additionally, this requires an equation specifying the measurement system — the way Y_i relate to the underlying attributes of interest.

I formalize sampling and measurement in Assumptions 2.S1 and 2.S2, later generalized in Section 3.3.

Assumption 2.S1 (Simple random sampling).

$$P\left(R_{1:n} = r_{1:n}\right) = 1\bigg/\binom{n}{N},$$

for each n-vector $r_{1:n}$ *with* $E_n\left[r_i\right] = N/n$.

Assumption 2.S1 describes random sampling without replacement, which leads to a sample of size N from the target population. The sample-to-population fraction is thus N/n; the limit case where all population units are sampled corresponds to $N/n = 1$. Similarly, this formulation nests the "infinite" superpopulation framework where the sample represents a negligible fraction of the population if we let $n \to \infty$ for fixed N.

The attributes of interest for the sampled units are not directly observed. Instead, we have access to noisy measurements according to

$$Y_{it} = \theta_i + \varepsilon_{it}, \qquad \text{for } t = 1, \ldots, T. \tag{2.1}$$

We assume that $E\left[\varepsilon_{it}\right] = 0$, a reasonable requirement that ensures that Y_{it} are "good measurements" in the sense of being unbiased for θ_i for each unit. Note that while here we index measurements by t, these need not have a time ordering.

On top of this, we also assume below that there is limited dependence across measurement errors, a necessary condition in order to gauge the extent of measurement uncertainty.

Assumption 2.S2 (Weakly dependent measurements in the simple model). *The measurement errors* $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ *in* (2.1) *satisfy* $E\left[\varepsilon_i \varepsilon_i'\right] = I_T \sigma^2$; $\sigma^2 < \infty$.

Assumption 2.S2 implies that randomness around the attribute of interest is uncorrelated over measurements; that is, $E\left[\varepsilon_{it}\varepsilon_{is}\right] = 0$ for $s \neq t$. Throug the lens of the school quality example where Y_{it} are average test scores for the tth student cohort, this is a natural starting point: it implies that cohort-specific variation in test scores is uncorrelated across cohorts. Similar assumptions are common in measurement systems such as (2.1); see for instance the discussion in Gilraine et al. (2022) in the context of teacher value-added models. The homoskedasticity assumption is made for simplicity.

In Section 3.3, I generalize this assumption to allow for unrestricted heteroskedasticity and different forms of dependence across measurements such as moving average errors, and I discuss the extent to which weak dependence assumptions are testable. These notions are at the heart of measurement models, as they reflect the intrinsic trade-off between unobserved heterogeneity and persistence: in an error-component model such as (2.1), it is θ_i that induces strong dependence in Y_{it} across measurements. The suitability of a given set of restrictions should be discussed jointly with that of a given measurement model.

Estimation and estimation uncertainty. Let $\bar{Y}_i = T^{-1}\sum_{t=1}^{T} Y_{it}$ and $\bar{\varepsilon}_i = T^{-1}\sum_{t=1}^{T}\varepsilon_{it}$. A natural estimator for β_n is

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{n} R_i \bar{Y}_i = \frac{1}{N}\sum_{i=1}^{n} R_i \theta_i + \frac{1}{N}\sum_{i=1}^{n} R_i \bar{\varepsilon}_i,$$

where note that we average over all units, but only those with $R_i = 1$ effectively enter the sums. Using $E\left[R_i\right] = N/n$, it is easy to see that the estimator is unbiased: $E\left[\hat{\beta}\right] = \beta_n$.

The above expression also shows that the estimator decomposes into two different terms, which are at the core of much that follows. In particular, they represent orthogonal sources of estimation uncertainty: sampling and measurement. This can be read off directly from the variance of the estimator:

$$\operatorname{Var}\left(\hat{\beta}\right) = \operatorname{Var}\left(\frac{1}{N}\sum_{i=1}^{n} R_i\theta_i\right) + \operatorname{Var}\left(\frac{1}{N}\sum_{i=1}^{n} R_i\bar{\varepsilon}_i\right) = \left(1 - \frac{N}{n}\right)\frac{\operatorname{Var}_n\left(\theta_i\right)}{N} + \frac{\operatorname{Var}\left(\varepsilon_{it}\right)/T}{N}. \tag{2.2}$$

where $\text{Var}_n(\theta_i) = (n-1)^{-1}\sum_{i=1}^{n}(\theta_i - \beta_n)^2$ and where we have used Assumptions 2.S1 and 2.S2.[7] These two terms embed the notion of sampling and measurement uncertainty, respectively. The first term is the variance of the ideal estimator of β_n if θ_i were directly observable; sampling is the only source of randomness here. The second term captures uncertainty induced by the measurement problem.

For our purposes, the most important feature in the expression above is that sampling-based uncertainty is indexed by the sample-to-population fraction $f = N/n$. It is helpful to view $\text{Var}\left(\hat{\beta}\right)$ as a function of f, let it be denoted by $\text{V}(f)$. When $f = 1$, there is no sampling uncertainty: in the absence of a measurement problem, the ideal estimator of β_n would be β_n itself. On the other extreme, sampling uncertainty is largest when we regard the sample as a negligible fraction of the population. This is captured by $\text{V}(0) = \lim_{f\to 0}\text{V}(f)$.

At the same time, measurement uncertainty does not depend on f: our ability to obtain more accurate measurements of θ_i for each unit is not related to the sampling framework. The relative weight of these two components in estimation uncertainty is modulated by signal-to-noise in the data: sampling uncertainty is relatively larger the more dispersed the underlying attributes are in the population (signal) and the less noise there is in the measurement system, captured by the size of the measurement errors and the number of measurements.

Remark 2.1 (External validity). An advantage of the sampling–measurement framework is that it sheds light on the relevant sources of uncertainty for a given question of interest. One notion of external validity researchers might be concerned with is that of extrapolation beyond the specific circumstances that occurred during measurement. For instance, this might involve prediction exercises or "parallel universes" where a different sequence of shocks could have realized. Appropriately accounting for measurement-based uncertainty implies that β_n is directly informative for these questions.[8] Another notion of external validity in the literature is that of generalizability of results to an exchangeable population of interest; here it is adequately accounting for sampling-based uncertainty what guarantees external validity. When that population is the one over which β_n is defined, this follows from Assumption 2.S1. Alternatively, we can think of extrapolability of the results to new hypothetical units drawn from a superpopulation where the original and the new "target" units are exchangeable; say with size $\tilde{n} \geq n$. We then just need to redefine $\beta_{\tilde{n}}$ as the parameter of interest. Such conceptual exercises are often useful in meta-analyses (Meager, 2022). A more meaningful question for policy purposes is that of transferability of results to a new, different population. This requires additional tools that are independent of the sampling framework, see Jin and Rothenhäusler (2024) for a discussion in a finite population context.

Conventional inference. The conventional variance estimator for $\hat{\beta} = N^{-1}\sum_{i=1}^{n} R_i \bar{Y}_i$ would be

$$\hat{\text{V}}^{\text{cluster}} = \frac{1}{N(N-1)}\sum_{i=1}^{n} R_i\left(\bar{Y}_i - \hat{\beta}\right)^2;$$

[7] In particular, we use basic results on simple random sampling without replacement repeatedly. Assumption 2.S1 implies $E\left[R_i\right] = N/n$ and $E\left[R_i R_j\right] = N(N-1)/n(n-1)$ for $j \neq i$. It is this dependence across sampling indicators that induces the form of the first term in (2.2).

[8] This notion is often present in discussions about external validity when certain shocks are not accounted for; see for instance Hahn, Kuersteiner, and Mazzocco (2020) and Deeb and de Chaisemartin (2022) in the context of aggregate shocks.

a cluster-robust variance estimator (Liang and Zeger, 1986; Arellano, 1987). This is the natural choice here: clustering within units accounts for the presence of the persistent component θ_i in the measurement equation in (2.1); this is true regardless of the degree of dependence across measurement errors. In Appendix A.2.1, I show that

$$E\left[\hat{\mathrm{V}}^{\text{cluster}}\right] = \mathrm{V}(0) \geq \mathrm{V}(f),$$

for any sample-to-population fraction f. That is, the conventional variance estimator implicitly treats the sample as a random draw from a much larger population, and using $\hat{\mathrm{V}}^{\text{cluster}}$ for inference introduces superfluous sampling uncertainty when this is not the case.

By how much $\hat{\mathrm{V}}^{\text{cluster}}$ exaggerates estimation uncertainty is a matter of signal-to-noise. For instance, letting $\mathrm{Var}_n(\theta_i) = 1$ and $\mathrm{Var}(\varepsilon_{it})/T = 1$, the variance is on average twice as large as it should be when the sample is also the population. An exception is the limit case where $\theta_i = \theta$ for all units: since all underlying attributes are equal to each other, which population units are sampled and which ones are not is irrelevant.

Finite Population Corrections. It turns out that we can make progress when repeated measurements are available. In particular, the standard within-between variance decomposition gives

$$\widehat{\mathrm{Var}}(\varepsilon_{it}) = \frac{1}{N(T-1)} \sum_{i=1}^{n} R_i \sum_{t=1}^{T} (Y_{it} - \bar{Y}_i)^2,$$

$$\widehat{\mathrm{Var}_n}(\theta_i) = \frac{1}{N-1} \sum_{i=1}^{n} R_i \left(\bar{Y}_i - \hat{\beta}\right)^2 - \frac{\widehat{\mathrm{Var}}(\varepsilon_{it})}{T},$$

which rely on weak dependence for their validity (Assumption 2.S2). The finite-population variance estimator is then constructed via a simple adjustment to the conventional estimator — a Finite Population Correction:

$$\hat{\mathrm{V}}(f) = \hat{\mathrm{V}}^{\text{cluster}} - f\,\frac{\widehat{\mathrm{Var}_n}(\theta_i)}{N}.$$

It is not difficult to show that for any f we have

$$E\left[\hat{\mathrm{V}}(f)\right] = \mathrm{V}(f),$$

so that the finite-population variance estimator reflects the right amount of sampling and measurement uncertainty in the problem. Under the regularity conditions in Section 3.3, $\hat{\mathrm{V}}(f)$ can then be used to perform asymptotically correct inference for any f and a fixed number of measurements.

2.3 General case

In this section, I study estimation and inference for finite-population estimands in a general framework where sampling-based and measurement-based uncertainty coexist and provide Finite Population Corrections that guarantee non-conservative inference for any sample-to-population fraction.

I introduce the setup in Section 2.3.1 and characterize estimation uncertainty for moment-based estimators of the parameters of interest in Section 2.3.2. I introduce Finite Population Corrections and state the main result on non-conservative inference in Section 3.3.2. Proofs and derivations and relegated to Appendix A.1.

2.3.1 Setup

Consider a population of size n. Unit i in the population is characterized by a set of fixed attributes $\{\theta_i, W_i\}$, and the researcher is interested in a summary measure β_n of outcome θ_i, say, an average over the population or a coefficient on a regression involving characteristics W_i. Probability statements are understood to hold conditional on these fixed attributes. Population averages are denoted as $E_n\left[f(x_i)\right] \coloneqq n^{-1}\sum_{i=1}^n f(x_i)$ for a function f applied to an array $(x_i)_{i=1}^n$.

Given a sampling framework, we use $(R_1, \ldots, R_n) \in \{0,1\}^n$ to denote the vector of inclusion indicators; $R_i = 1$ indicates that unit i is sampled. The observed data for each sampled unit is a vector of noisy measurements $Y_i = (Y_{i1}, \ldots, Y_{iT})'$. In a given sample, the analyst has access to $\{R_i, R_iY_i, R_iW_i\}_{i=1}^n$. We now define the objects of interest and formalize each dimension of uncertainty.

Estimands

Let $\dim \theta_i = 1$ and $\dim \beta_n = p$. That θ_i are scalar outcomes is for simplicity and all results extend with minor modifications to the multivariate case; I will point those out throughout the exposition. The target objects β_n solve population moment conditions $h(\theta, W, b_n)$ affine in θ:

$$E_n\left[h_1(W_i;\beta_n)\left(\theta_i - h_0(W_i;\beta_n)\right)\right] = 0, \tag{2.3}$$

where h_0 (scalar-valued) and h_1 (of size $p\times 1$) are known functions continuously differentiable in β_n. I assume that $h_1(W_i;\beta_n) \neq 0$ for each unit in the population.[9]

The moment conditions in (2.3) define a broad class that includes moment-based methods such as linear regression models, IV-like estimands or nonlinear least squares. Setting $p = 1$, $h_1(W_i;\beta_n) = 1$ and $h_0(W_i;\beta_n) = \beta_n$ recovers $\beta_n = E_n\left[\theta_i\right]$ as in Section 3.2, a population average over heterogeneous attributes.

Remark 2.2 (Prediction example, revisited). In this application the population of interest comprises all local police agencies in the U.S., and θ_i is the baseline level of police violence associated to each department over the panel horizon. Montiel-Olea et al. (2021) are interested in observable determinants Z_i of θ_i such as state-level laws or demographics and specify an exponential model of the form

$$\theta_i = \exp\left(Z_i'\beta_n + \alpha_i\right), \tag{2.4}$$

where α_i is the unexplained component and the target coefficients β_n are defined via a nonlinear least squares problem. Through the lens of our framework, Z_i are observable attributes included in W_i, $h_1(W_i;\beta_n) = Z_i \exp\left(Z_i'\beta_n\right)$ and $h_0(W_i;\beta_n) = \exp\left(Z_i'\beta_n\right)$; equation (10) in Montiel-Olea et al. (2021) is then exactly (2.3). The ultimate objects of interest (predictions and counterfactuals) then involve known transformations of these β_n parameters.

[9] This is trivially satisfied by redefining the subpopulation to the set of units that satisfy this condition. In other words, this condition is definitional.

Remark 2.3 (Misallocation example, revisited). In the exercise in Section 2.5.2 we are interested in characterizing the extent of resource misallocation in the formal manufacturing sector in Indonesia, and θ_i are "wedges" that measure firm-level deviations from optimal allocation of labor. A popular approach is to explore whether wedges relate systematically to observable firm-level characteristics Z_i, such as measures of firm size. In this context, β_n are least-squares projection coefficients.

Equation (2.3), on the other hand, excludes nonlinear transformations of θ_i. Generally speaking, nonlinear transformations of unbiased measurements are not unbiased, and identification and estimation require additional assumptions. It is nonetheless possible to extend this framework to cover certain nonlinear transformations. Some of these are of great relevance in the empirical illustration in Section 2.5.2, where the extent of cross-sectional dispersion in "wedges" θ_i can be directly mapped to macroeconomic aggregates. I discuss extensions to this case in Remark 2.8 below and in the empirical application.

Measurement

I specify the following measurement equation for θ_i:

$$Y_i = g_0(X_i; \delta) + g_1(X_i; \delta)\theta_i + \varepsilon_i, \quad E\left[\varepsilon_i\right] = 0, \tag{2.5}$$

where $X_i = (X_{i1}', \ldots, X_{iT}')'$ is a $T \times \dim X_{it}$ matrix of fixed attributes contained in W_i,[10] $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ are independent but not necessarily identically distributed measurement errors and δ is a fixed k-vector of unknown parameters ($k \leq \dim X_{it}$). These might be of direct or auxiliary interest to the researcher and do not depend on population size n. Finally, g_0 and g_1 (of size $T \times 1$) are known functions continuously differentiable in δ. This formulation allows the model to be nonlinear in observable attributes X_i and common parameters and to be known only up to the latter; $T > 1$ is needed to estimate the system. Allowing for nonlinear terms in the measurement equation substantially broadens the applicability of the methods developed here; for instance, (2.5) covers factor models and multiplicative models with unobserved components such as the Poisson regression model in Section 2.5.1 (see the introduction section for examples). We also assume that $\det g_1(X_i; \delta)' g_1(X_i; \delta) \neq 0$ for all i, which essentially amounts to requiring that the data are informative of attribute θ_i for the population of interest.[11]

The zero mean assumption $E\left[\varepsilon_i\right] = 0$ ensures that the repeated measurements Y_{it} are unbiased for the nonstochastic component of the model. In a panel data context defined in a superpopulation where $\{W_i, \theta_i\}$ are treated as random quantities, $E\left[\varepsilon_i\right] = 0$ is a zero conditional mean assumption as in the random coefficients model in Chamberlain (1992) and Arellano and Bonhomme (2012). Assumption 2.1 below is the main necessary assumption of the chapter and formalizes the notion of limited dependence in measurements.

Assumption 2.1 (Weakly dependent errors).

[10] Attributes in W_i but not in X_i are excluded instruments from the point of view of the measurement equation (2.5); these might help describe θ_i and enter the moment condition for β_n.

[11] This condition is analogous to that on h_1 in equation (2.3), in that it implies that the results apply for the subpopulation of units that satisfy this requirement. For instance, for a difference-in-difference measurement model where $X_{it} \in \{0, 1\}$ and $T = 2$, the restriction that $X_{i1} + X_{i2} \neq 0$ redefines the population of interest to treated units and reflects the familiar result that $\beta_n = E_n\left[\theta_i\right]$ is the ATT rather than the ATE (in the absence of additional assumptions). See Graham and Powell (2012) for a discussion of irregular models where $\det g_1(X_i; \delta)' g_1(X_i; \delta)$ might be close to zero for some units.

Let $S_{(m)}$ be a $T^2 \times m$ full column rank selection matrix such that $E\left[\varepsilon_i \otimes \varepsilon_i\right] = S_{(m)}\omega_i$ for an m-vector of parameters ω_i and the measurement system defined in (2.5)*. Then*

$$m \leq \frac{T(T+1)}{2} - 1. \tag{2.6}$$

Assumption 2.1 rules out fully unrestricted covariance matrices, but allows for arbitrary patterns of dependence and heteroskedasticity in the non-restricted elements ω_i. Assumption 2.1 operationalizes the notion of weak dependence over repeated measurements via the choice of selection matrix $S_{(m)}$, which imposes linear restrictions on $\Omega_i = E\left[\varepsilon_i \varepsilon_i'\right]$.[12][13]

Limited dependence is particularly appealing in a repeated measurements context, where it is expected that randomness in those is (partly) non-systematic. When measurements are drawn in parallel, independence might be reasonable; when measurements have a natural time or spatial ordering, a stronger association might be expected between closer errors than between those far apart. Moving average processes are convenient implementations of this idea. Similar notions of weak dependence also underpin much of the work in time series econometrics.[14]

Remark 2.4 (Testable restrictions.). When the order condition (2.6) is strict, Assumption 2.1 is testable. A standard J-test can be constructed following the long-standing panel data literature on testing covariance structures (Abowd and Card, 1989; Arellano, 2003; Arellano and Bonhomme, 2012). The informative content of the data for Assumption 2.1 and its plausibility is evident in the empirical applications that are discussed below.

Sampling

Assumption 2.2 places assumptions on $R_1, \ldots, R_n$ and embeds the population into a sequence of finite populations of growing sizes $n \longrightarrow \infty$.

Assumption 2.2 (Random sampling).

(i) *Given the system in* (2.5)*, unit i is independently sampled with probability* $f_n > 0$.

[12]For instance, with $T = 2$ measurements, the restriction $E\left[\varepsilon_{i1}\varepsilon_{i2}\right] = 0$ can be represented as

$$S_{(2)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}'.$$

Note that this leaves $E\left[\varepsilon_{it}^2\right]$ for $t \in \{1, 2\}$ completely unrestricted.

[13]If $\dim \theta_i > 1$, (2.6) is modified to

$$m \leq \frac{T(T+1)}{2} - \frac{\dim \theta_i(\dim \theta_i + 1)}{2}.$$

This result is established in Arellano and Bonhomme (2012) and reflects a fundamental trade-off between unobserved heterogeneity and error persistence in panel data models with unit-level coefficients.

[14]Classical references are Hansen and Singleton (1982) and Newey and West (1987). The literature of repeated measurements has often favored moving average processes since they imply linear restrictions on Ω_i, see for instance Bonhomme and Robin (2009) and Bonhomme and Robin (2010) in the context of factor models and the discussion in Arellano (2003, Chapter 5). Autoregressive processes, on the other hand, are not covered by Assumption 2.1. Still, the methods in this chapter can be extended to cover such forms of dependence over measurements via quasi-differencing; see also Arellano and Bonhomme (2012, Section 3.2) for the counterpart to Assumption 2.1.

(ii) f_n *satisfies* $nf_n \to \infty$ *and* $f_n \to f \in [0,1]$.

Assumption 2.2(i) formalizes random sampling and rules out sample selection. It also implies that the sample size $N = \sum_{i=1}^{n} R_i$ is random, a convenience device to avoid dealing with dependence across inclusion indicators — moving beyond the exact results in Section 3.2. Note that $\hat{f} = N/n$ is a natural estimator of f_n; this is inconsequential for the large-sample results presented here as long as $\hat{f}/f_n \xrightarrow{p} 1$ as $n \to \infty$. Assumption 2.2(ii) ensures that as $n \to \infty$ the expected sample size nf_n also increases and that the sampling fraction f_n has a well-defined limit. In essence, this is a way of relying on large-sample approximations for finite populations while ensuring that these remain representative of the sampling framework; for instance, choosing a sequence such that $\lim f_n = 0$ allows us to capture the standard environment where the sample becomes negligible relative to the population.[15]

Estimator

Let $\gamma_n = (\delta', \beta_n')'$ be the $(k+p)$ parameters of interest, including both finite-population estimands and parameters of the measurement system. For a generic $\tilde{\gamma}$, let

$$u\left(Y_i, W_i, \tilde{\gamma}\right) = Y_i - g_0\left(X_i; \tilde{\delta}\right) - g_1\left(X_i; \tilde{\delta}\right) h_0\left(W_i; \tilde{\beta}\right) \tag{2.7}$$

and let $Q_i\left(\tilde{\delta}\right) = I_T - g_1\left(X_i; \tilde{\delta}\right) g_1\left(X_i; \tilde{\delta}\right)^{\dagger}$ denote the projection on the orthogonal of the span of $g_1\left(X_i; \tilde{\delta}\right)$.[16] Consider a (non-redundant) set of instruments $A\left(W_i, \tilde{\delta}\right)$ for δ, assumed to be continuously differentiable in δ. I consider a method-of-moments approach with moment function

$$\psi(Y_i, W_i, \tilde{\gamma}) = \begin{pmatrix} \psi_\delta(Y_i, W_i, \tilde{\gamma}) \\ \psi_\beta(Y_i, W_i, \tilde{\gamma}) \end{pmatrix} = \begin{pmatrix} A\left(W_i, \tilde{\delta}\right) Q_i\left(\tilde{\delta}\right) \\ h_1(W_i; \tilde{\beta}) g_1\left(X_i; \tilde{\delta}\right)^{\dagger} \end{pmatrix} u\left(Y_i, W_i, \tilde{\gamma}\right). \tag{2.8}$$

Some intuition is as follows. Unobserved attributes θ_i are incidental parameters from the point of view of estimation of the parameters of the measurement system (2.5). The role of $Q_i\left(\tilde{\delta}\right)$ is to induce a transformation of the system that does not depend on θ_i; note that $Q_i\left(\tilde{\delta}\right) g_1\left(X_i; \tilde{\delta}\right) = 0_T$. The moment conditions for β_n, on the other hand, rescale the system so that $g_1\left(X_i; \tilde{\delta}\right)^{\dagger} g_1\left(X_i; \tilde{\delta}\right) = 1$ and then bring in the population moment conditions for β_n defined in (2.3). The method-of-moments estimator $\hat{\gamma}$ solves:

$$\sum_{i=1}^{n} R_i \psi(Y_i, W_i, \hat{\gamma}) = 0. \tag{2.9}$$

[15] These embeddings are referred to as finite-population asymptotics in the literature, see Lehmann (1975), Aronow, Green, and Lee (2014), Li and Ding (2017), Abadie et al. (2020) and Xu (2021) for applications.

[16] For a matrix B, $B^{\dagger}$ denotes its Moore–Penrose pseudoinverse. In the context of panel data models, $g_1\left(X_i; \tilde{\delta}\right)^{\dagger}$ and $Q_i\left(\tilde{\delta}\right)$ are often referred to as generalized between- and within-group operators, respectively (see, for instance, Chamberlain, 1992; Arellano and Bonhomme, 2012).

2.3.2 Characterizing estimation uncertainty

Here, I study the large-sample properties of $\hat{\gamma}$ as an estimator of γ_n, characterize its asymptotic variance and discuss estimation uncertainty in finite populations.

First, note that γ_n solves

$$E_n\left[E\left[\psi(Y_i, W_i, \gamma_n)\right]\right] = 0. \tag{2.10}$$

This can be verified by noting that the moment function at γ_n satisfies:

$$\begin{aligned}\psi_\delta(Y_i, W_i, \gamma_n) &= A\left(W_i, \delta\right) Q_i\left(\delta\right)\varepsilon_i,\\ \psi_\beta(Y_i, W_i, \gamma_n) &= h_1(W_i, \beta_n)\left(\theta_i - h_0\left(W_i; \beta_n\right)\right) + h_1(W_i, \beta_n) g_1\left(X_i; \delta\right)^\dagger \varepsilon_i.\end{aligned}$$

The result follows from $E\left[\varepsilon_i\right] = 0$ and averaging over population attributes.[17] Note that setting $h_1 = 1$, $h_0 = \beta_n$ and $g_1 = 1_T$, $\psi_\beta(Y_i, W_i, \gamma_n)$ reduces to the estimation error in the simple example in Section 3.2. Let

$$\begin{aligned}\mathrm{V}_{\psi,n}(f_n) = &E_n\left[E\left[\psi(Y_i, W_i, \gamma_n)\psi(Y_i, W_i, \gamma_n)'\right]\right]\\ &- f_n E_n\left[E\left[\psi(Y_i, W_i, \gamma_n)\right] E\left[\psi(Y_i, W_i, \gamma_n)\right]'\right].\end{aligned} \tag{2.11}$$

Below we establish that the limit of $\mathrm{V}_{\psi,n}(f_n)$ as $n \to \infty$ is the inner term of the asymptotic variance; the second term in (2.11) is the Finite Population Correction. Finally, let $H_n = E_n\left[E\left[\nabla_{\tilde{\gamma}}\psi(Y_i, W_i, \gamma_n)\right]\right]$. Proposition 2.1 characterizes the asymptotic distribution of the (rescaled) estimation error; the following limits are assumed to exist as part of the regularity conditions.

Proposition 2.1 (Asymptotic distribution). *Under the measurement system in (2.5), Assumption 2.2 and the regularity conditions in Assumption A.1 in Appendix A.1, as $n \to \infty$ and for given $T > 1$:*

$$\sqrt{N}\left(\hat{\gamma} - \gamma_n\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, V(f)\right),$$

where

$$V(f) = \left(\lim_{n\to\infty} H_n\right)^{-1} \lim_{n\to\infty} V_{\psi,n}(f_n)\left(\lim_{n\to\infty} H_n'\right)^{-1}. \tag{2.12}$$

Proof. See Appendix A.1. □

Proposition 2.1 is the finite-population counterpart to standard results for problems with repeated measurements.[18] While the asymptotic variance has the usual "sandwich" form, it is indexed by f. In other words, estimation uncertainty depends on the sample-to-population fraction — even if the measurement problem does not. The reason for this is simple: these two sources of uncertainty are orthogonal to each other, and randomness in $\hat{\gamma}$ reflects both. As $f \to 1$, randomness due to sampling disappears and the

[17] I take as given that δ is identified from $E\left[\psi_\delta(Y_i, W_i, \gamma_n)\right] = 0$. A necessary requirement is that A contains $\dim \delta = k$ valid instruments.

[18] This can be established with our sampling assumption (Assumption 2.2) and regularity conditions for moment-based estimators (Newey and McFadden, 1994), but does not rely on assumptions of dependence across measurements. See also Xu (2021) for a similar result in a world with design-based uncertainty and no repeated measurements.

asymptotic variance of $\hat{\gamma}$ adjusts proportionally via the Finite Population Correction. Note that the FPC is positive-semidefinite; it follows that for $f' \geq f$, $\mathrm{V}(f') \leq \mathrm{V}(f)$ in the matrix sense.

Two particular cases are worth highlighting. First, when $f = 0$ the standard sandwich formula recovers. This is the basis for the standard, superpopulation-based approach to inference; let $\hat{\mathrm{V}}(0)$ denote any such estimator. It then follows that $\hat{\mathrm{V}}(0)$ is generally inconsistent for the finite-population variance when $f > 0$, and that conventional standard errors tend to exaggerate estimation uncertainty.

Second, the FPC is zero when $E\left[\psi(Y_i, W_i, \gamma_n)\right] = 0$, that is, when the population moment condition (2.10) holds *for each unit* i. When it holds only on average, the FPC is precisely equal to the variance of these heterogeneous unit-level moment conditions over the population, and is larger the more dispersed these are. In the simple example in Section 3.2, $E\left[\psi(Y_i, W_i, \gamma_n)\right] = \theta_i - \beta_n$ and the FPC equals $E_n\left[(\theta_i - \beta_n)^2\right]$, the variance over population heterogeneous responses. It is this insight and the connection to random coefficient models that I exploit to propose Finite Population Correction estimators.

One relevant case in which FPCs are zero are parameters of the measurement model, denoted here δ. This follows from $E\left[\psi_\delta(Y_i, W_i, \gamma_n)\right] = 0$; it can be verified that the upper-left $k \times k$ block of Σ_n is a zero matrix. This is intuitive: since the measurement problem is present regardless of how much of the population we sample, uncertainty about the measurement system itself should not depend on f. When population attributes $\{\theta_i\}_{i=1}^n$ are actually homogeneous, β_n becomes a common parameter. Through the lens of the causal inference literature, this is the classical result that conventional standard errors with randomized treatments are not conservative under constant treatment effects (Neyman, 1923/1990; Abadie et al., 2020).

Remark 2.5 (Perfect measurements). Suppose that outcome attributes are observed without error; for simplicity, set $Y_{it} = \theta_i$ and $T = 1$. Then the moment function ψ_β is nonstochastic and the variance of the moment condition in (2.11) adapts to reflect so:

$$\mathrm{V}_{\psi,n}(f_n) = (1 - f_n)E_n\left[\psi_\beta(Y_i, W_i, \gamma_n)\psi_\beta(Y_i, W_i, \gamma_n)'\right].$$

This is analogous to the classical finite-populations literature where sampling-based uncertainty is the only source of randomness in $\hat{\beta}$. The usual variance estimator (say, heteroskedasticity-robust) $\hat{\mathrm{V}}(0)$ is conservative, but an adjustment is here straightforward: $\hat{\mathrm{V}}(f) = (1 - f)\hat{\mathrm{V}}(0)$ will do.

Whichever the setup, the dominant paradigm in empirical work is to interpret uncertainty as-if derived from an infinite population. In the next section, I show that the discussion above is not only about the interpretation of uncertainty but has practical consequences: estimating FPCs is possible when repeated measurements are available.

2.3.3 Finite Population Corrections

Our goal in this section is to propose consistent finite-population standard errors and confidence intervals for $\hat{\gamma}$.

Let $\hat{u}_i \equiv u\left(Y_i, W_i, \hat{\gamma}\right)$, where $u\left(Y_i, W_i, \tilde{\gamma}\right)$ is defined in (2.7). This is a compound residual term that includes both types of unobservables in the measurement equation: attributes and measurement errors. In a nutshell, the idea is to mimic the approach in the simple example in Section 3.2 and propose variance estimators that weight these two elements according to some $\tilde{f} \in [0,1]$. Let $Q_i^*\left(\tilde{\delta}\right)$ denote the projection on the orthogonal of the span of $g_1\left(X_i;\tilde{\delta}\right) \otimes g_1\left(X_i;\tilde{\delta}\right)$.[19] The following are weighted unit-level contributions to the finite-population variance:[20]

$$\hat{\Lambda}_i\left(\tilde{f}\right) = \text{vec}^{-1}\left[\left(1-\tilde{f}\right)I_{T^2} + \tilde{f}S_{(m)}\left(Q_i^*\left(\hat{\delta}\right)S_{(m)}\right)^{\dagger}Q_i^*\left(\hat{\delta}\right)\right]\left(\hat{u}_i \otimes \hat{u}_i\right). \tag{2.13}$$

The cross-products $\hat{u}_i \otimes \hat{u}_i$ weighted by $\left(1-\tilde{f}\right)$ include both attributes and measurement errors; those weighted by $\tilde{f}$ include only measurement errors. The latter are constructed by imposing the covariance structure in Assumption 2.1 and then projecting out the part involving the attributes via $Q_i\left(\hat{\delta}\right)$. This estimator is based on the constructive identification proof in Arellano and Bonhomme (2012) for covariance structures in panel data random coefficient models. Importantly, this is the only modification that finite-population standard errors will require relative to conventional approaches: implementation only requires defining a selection matrix and a projection matrix.

Now, the finite-population variance estimator of the score is

$$\hat{V}_{\psi}\left(\tilde{f}\right) = \frac{1}{N}\sum_{i=1}^{n} R_i \begin{pmatrix} A\left(W_i,\hat{\delta}\right)Q_i\left(\hat{\delta}\right) \\ h_1(W_i;\hat{\beta})g_1\left(X_i;\hat{\delta}\right)^{\dagger}\end{pmatrix}\hat{\Lambda}_i\left(\tilde{f}\right)\begin{pmatrix} A\left(W_i,\hat{\delta}\right)Q_i\left(\hat{\delta}\right) \\ h_1(W_i;\hat{\beta})g_1\left(X_i;\hat{\delta}\right)^{\dagger}\end{pmatrix}',$$

where recall that $\hat{f} = N/n$. Note that using $\hat{\Lambda}_i(0)$ instead yields the conventional estimator of the variance of the scores for repeated measurement models, which averages over both dispersion in attributes and dispersion in measurement errors: the presence of the former reminds us that this is in the class of cluster-robust variance estimators.

Let $\hat{H} = N^{-1}\sum_{i=1}^{n} R_i \nabla_{\tilde{\gamma}}\psi(Y_i, W_i, \hat{\gamma})$. The estimator of the finite-population variance in (2.12) is given by

$$\hat{\mathrm{V}}\left(\hat{f}\right) = \hat{H}^{-1}\hat{\mathrm{V}}_{\psi}\left(\hat{f}\right)\hat{H}'^{-1}. \tag{2.14}$$

For $\hat{\mathrm{V}}\left(\hat{f}\right) \geq 0$ and an arbitrary column vector $\lambda \neq 0_{(k+p)\times 1}$, the finite-population standard error is $\hat{\sigma}_{\lambda}\left(\hat{f}\right) = \sqrt{\lambda'\hat{\mathrm{V}}\left(\hat{f}\right)\lambda/N}$. Finally, the $(1-\alpha)$ confidence interval for $\lambda'\gamma_n$ is

$$\hat{C}_{\lambda,\alpha}\left(\hat{f}\right) = \left[\lambda'\hat{\gamma} \pm z_{1-\alpha/2}\hat{\sigma}_{\lambda}\left(\hat{f}\right)\right], \tag{2.15}$$

where z_q is the q-quantile of the standard normal distribution. Proposition 2.2 below states that this leads

[19] This is the counterpart of $Q_i\left(\tilde{\delta}\right)$ for cross-products of the data:

$$Q_i^*\left(\tilde{\delta}\right) = I_{T^2} - g_1\left(X_i;\tilde{\delta}\right)g_1\left(X_i;\tilde{\delta}\right)^{\dagger} \otimes g_1\left(X_i;\tilde{\delta}\right)g_1\left(X_i;\tilde{\delta}\right)^{\dagger}.$$

[20] $\text{vec}_{m,n}^{-1} : \mathbb{R}^{mn} \to \mathbb{R}^{m\times n}$ is the inverse vec operator. For an $m \times n$ matrix B, we have $= \text{vec}^{-1}\text{vec}\, B = B$. I omit the subscripts in the text since I only use vec^{-1} here to reconstruct $T \times T$ matrices. This is readily available in commercial software, such as via `reshape` in MATLAB.

to non-conservative inference for any f_n that satisfies Assumption 2.2.

Proposition 2.2 (Asymptotically correct inference). *Under the measurement system in* (2.5), *Assumption 2.1, Assumption 2.2 and the regularity conditions in Assumption A.1 in Appendix A.1, if* $\operatorname{rank} Q_i^*(\delta)\, S_{(m)} = m$ *then as* $n \to \infty$ *and for given* $T > 1$

$$\lim_{n\to\infty} P\left(\lambda'\gamma_n \in \hat{C}_{\lambda,\alpha}\left(\hat{f}\right)\right) = 1 - \alpha.$$

Proof. See Appendix A.1. □

The most relevant assumption behind Proposition 2.2 is that of weak dependence across measurements. Assumption 2.1, however, is not sufficient. We also require the more primitive condition that $Q_i^*(\delta)\, S_{(m)}$ has linearly independent columns and thus the left inverse in (2.13) is well-defined. This rank condition rules out cases where it is not possible to distinguish attributes from dependence in measurement errors from the second-order moments of the data even if restrictions are such that there are sufficient free parameters in the "reduced-form" covariance matrix.[21]

Finally, note that all elements of $\hat{\mathrm{V}}\left(\hat{f}\right)$ are generally speaking a function of $\hat{f}$, despite our discussion following Proposition 2.1 that estimation uncertainty for common parameters is independent of the sampling fraction. Under Assumption 2.1, the proposed variance is nonetheless valid: as we move along $\tilde{f} \in [0,1]$, we are only changing the relative weight of the attributes component in the compound residual term, but the upper-left $k \times k$ block of $\hat{V}_{\psi}(\tilde{f})$ is constructed such that it projects out this component.[22] For $\tilde{f} = 1$ and from the point of view of common parameters, this can be seen as a generalization of the approach in Stock and Watson (2008).

Note that the variance estimator in (2.14) is not guaranteed to be positive semidefinite for all $\tilde{f} \in [0,1]$ as written, although there always exists some $\tilde{f}$ for which this is the case. A natural alternative is to use a conservative estimator, say $\hat{\mathrm{V}}(0)$.[23] This is not an issue neither in our empirical applications nor in the simulation evidence presented in Section 2.4.

Remark 2.6 (Finite Population Corrections). Note that (2.14) can be written as:

$$\hat{\mathrm{V}}\left(\hat{f}\right) = \hat{\mathrm{V}}(0) - \hat{f}\left(\hat{\mathrm{V}}(0) - \hat{\mathrm{V}}(1)\right),$$

which has the intuitive form "conventional estimator – FPC." The above representation is most useful when different conventional estimators might be available, such as when δ are not of direct interest and estimation

[21] One such example is provided by the panel data literature on distinguishing unobserved heterogeneity from genuine dependence, where a measurement model with uncorrelated measurements is observationally equivalent to one with common attributes ($\theta_i = \theta$ for all units) and serial correlation in very short panels ($T = 2$), see Arellano (2003, pp. 58–60) for additional details. More generally, the rank condition fails when covariance restrictions do not restrict dependence across measurements but only impose homogeneity assumptions such as equal diagonal entries.

[22] This follows from the definition of common parameters themselves, see again the discussion in Section 2.3.1. Of course, if Assumption 2.1 fails, only the estimator that sets $f = 0$ would be consistent.

[23] This is a common drawback of estimators that are constructed by subtracting terms, as it becomes clear in the remark below. As such, asymptotically valid estimators can also be constructed in a number of standard ways in the literature, such as rotating the eigenvalues in the eigendecomposition of $\hat{\mathrm{V}}\left(\hat{f}\right)$; see for instance the discussion for two-way clustering in Cameron, Gelbach, and Miller (2011).

proceeds in two steps. It is then common to resort to bootstrap methods for inference on $\hat{\beta}$; an example of this is the first of my empirical illustrations. Let us focus on the jth entry of $\hat{\beta}$ and denote by $\tilde{\mathrm{V}}_{\beta,j}(0)$ the bootstrap variance.[24] A finite-population variance estimator for $\hat{\beta}_j - \beta_{n,j}$ that is valid in the sense of Proposition 2.2 is then

$$\tilde{\mathrm{V}}_{\beta,j}(0) - \hat{f} e_j' \hat{H}_\beta^{-1} \left(\hat{V}_{\psi_\beta}(0) - \hat{V}_{\psi_\beta}(1) \right) \hat{H}_\beta'^{-1} e_j,$$

where e_j is the basis vector of size p, $\hat{H}_\beta$ indexes the corresponding $p \times p$ block of $\hat{H}$ and $\hat{V}_{\psi_\beta}\left(\hat{f}\right)$ is analogous to $\hat{V}_\psi\left(\hat{f}\right)$ but only involves the ψ_β. In other words, the analyst just needs an estimate of FPC, and $\tilde{\mathrm{V}}_{\beta,j}(0)$ automatically takes care of two-step uncertainty (see also Newey and McFadden, 1994, Chapter 6).

Remark 2.7 (Conservative finite-population inference). Asymptotically correct inference requires limited dependence in measurements. When Assumption 2.1 is not attractive, it is nonetheless possible to compute partial FPCs by leveraging the predictive content of covariates W_i for the attributes of interest. In particular, building on similar ideas from the causal inference literature (Fogarty, 2018; Abadie et al., 2020; Xu, 2021), one can regress scores $\psi_\beta\left(Y_i, W_i, \hat{\gamma}\right)$ on observable attributes and use the variance of the predicted values to form partial FPCs.

Remark 2.8 (Extensions to higher order moments). Model (2.5) is a measurement equation for θ_i. The results presented here extend to a nonlinear transformation of θ_i, say θ_i^2, if a measurement equation for θ_i^2 is available. A different question is whether this is also the case if we maintain (2.5) as a baseline equation. One possibility is as follows; suppose for simplicity that δ are known. Let $Y_i^* = \left(Y_i - g_0(X_i;\delta)\right) \otimes \left(Y_i - g_0(X_i;\delta)\right)$, $g_1^*(X_i;\delta) = g_1(X_i;\delta) \otimes g_1(X_i;\delta)$ and define

$$\tilde{Y}_i^* = \left[I_{T^2} - S_{(m)} \left(Q_i^*(\delta)\, S_{(m)} \right)^{\dagger} Q_i^*(\delta) \right] Y_i^*. \tag{2.16}$$

It can then be shown that:

$$\tilde{Y}_i^* = g_1^*(X_i;\delta)\theta_i^2 + \tilde{\varepsilon}_i^*, \quad E\left[\tilde{\varepsilon}_i^*\right] = 0,$$

which is a measurement equation for θ_i^2 of the form (2.5). We can then make progress by characterizing the covariance structure of $\tilde{\varepsilon}_i^*$ in parallel to the exposition above. A limitation of this approach is that it imposes more stringent conditions on the number of available measurements relative to those needed for estimation. Still, objects such as the population-level dispersion of θ_i are of great interest in the context of the application in Section 2.5.2. I take a slightly different route and propose there a non-conservative variance estimator based on higher-order cumulants that directly estimates the FPC. This also illustrates the discussion in Remark 2.6.

2.4 Simulation study

Here I discuss a simulation study intended to illustrate the discussion so far and verify the finite-sample properties of the inference procedures proposed in the previous section.

[24] Similar ot the cluster-robust case, it can be shown that nonparametric (block) bootstrap estimators are also only consistent for $\mathrm{V}(0)$, regardless of the sampling fraction f.

Design. The design here considers a relatively simple measurement system that is additive in a scalar attribute of interest θ_i, and thus in the spirit of Section 3.2 and the empirical illustration in Section 2.5.2. I augment it with some additional ingredients as follows.

First, we consider population outcome attributes $\{\theta_i\}_{i=1}^n$ that are drawn from a superpopulation $\theta_i \sim \mathrm{N}\left(1, \sigma_\theta^2\right)$; interest is on the population average $\beta_n = E_n\left[\theta_i\right]$. We also define a T-vector of observable attributes X_i such that

$$X_{i0} = \left(1 - 0.25\theta_i\right) + |\theta_i| U_{i0},$$
$$X_{it} = 0.8X_{i,t-1} + U_{it},$$

and $U_{it} \sim \mathrm{t}_{(\kappa)}$ independently for $t = 0, \ldots, T$. This allows for persistence and non-normal features in attributes X_i, and induces dependence or "fixed-effects endogeneity" in θ_i. The population is thus characterized by $\{\theta_i, X_i\}_{i=1}^n$. The measurement equation for θ_i is specified as

$$Y_{it} = \theta_i + \delta X_{it} + \varepsilon_{it}, \tag{2.17}$$

where $\varepsilon_{it} \sim \mathrm{N}\left(0, X_{it}^2\right)$ independently over measurements. This augments the simple measurement model with a common parameter δ, which has to be estimated in a first step, and heteroskedasticity in measurement errors.

The design sets (σ_θ, κ) to control signal-to-noise. That is, we consider different relative weights of sampling-to-measurement uncertainty in the variance of the estimator. Note that the presence of δ adds two-step uncertainty to the problem, which in practice is equivalent to measurement uncertainty in the sense that the measurement system is not fully known. I also consider relatively small sample sizes ($N = 200$) and $T = 3$ measurements, and vary population size according to a grid of sample-to-population fractions $f \in \{0, 0.1, \ldots, 0.9, 1\}$. For instance, $f = 0.1$ is associated to a population of $n = 2,000$ units. The results for $f = 0$ correspond to the superpopulation data generating process (that is, the estimand equals one).

Results. Figures 2.1, 2.2 and 2.3 report coverage and width of finite-population confidence intervals over different sample-to-population fractions and for three signal-to-noise regimes (low, moderate and large, respectively). I also report conservative (or superpopulation) confidence intervals Liang and Zeger (1986); Arellano (1987) that impose $f = 0$ regardless of the actual sampling framework. I use critical values based on t_{n-1} as recommended in (Hansen, 2007) for cluster-robust estimators.

The results suggest an excellent performance of finite-population inference even for relatively small sample sizes, maintaining coverage close to nominal for all sample-to-population fractions and signal-to-noise regimes considered. Similarly, the figures also illustrate the conservativeness of conventional estimators for $f > 0$. In particular, actual coverage increases monotonically as $f \to 1$, and is one or close to one for cases where the sample is a large fraction of the population.

The extent of conservativeness is better captured by looking at the relative width of confidence intervals, and so is the size of Finite Population Corrections as a result. In line with the discussion above, these are larger the more dispersed the underlying attributes are relative to the size of measurement errors. In particular, for the limit case $f = 1$ the relative width is around 0.85 in the low signal-to-noise regime, 0.6 in the moderate one and around 0.5 in the high signal-to-noise one. As expected and in parallel, the actual coverage probability tends to increase for conservative confidence intervals as signal dominates, while remaining close to 0.95 for finite-population ones.

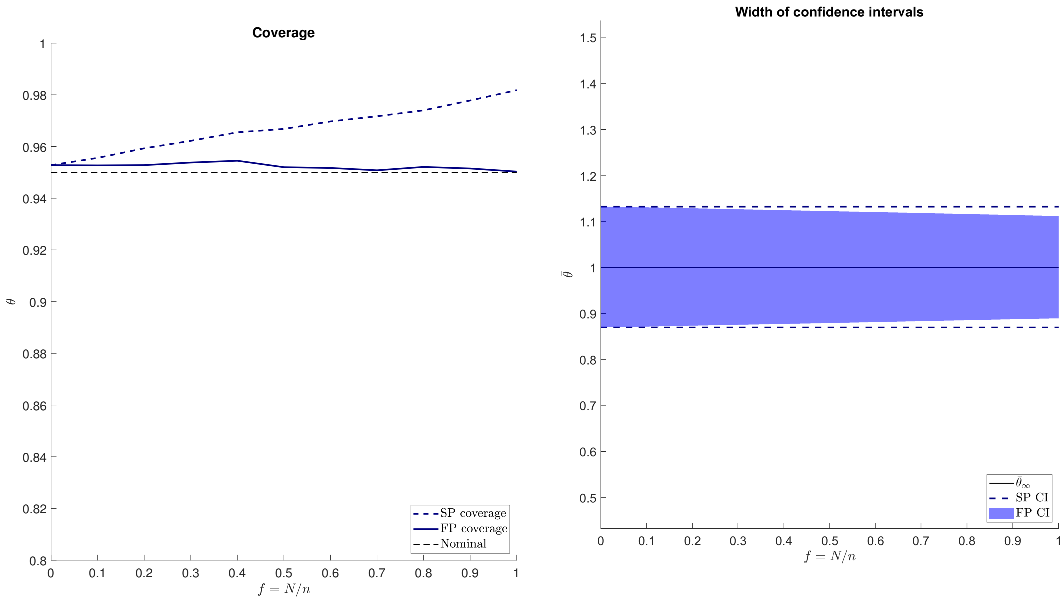


FIGURE 2.1. Results for β_n and the measurement model in equation (2.17): coverage (left) and width (right) of finite-population ("FP", solid lines) and superpopulation ("SP", dashed lines) confidence intervals. Nominal coverage is set to 0.95. Signal-to-noise ≈ 0.5.

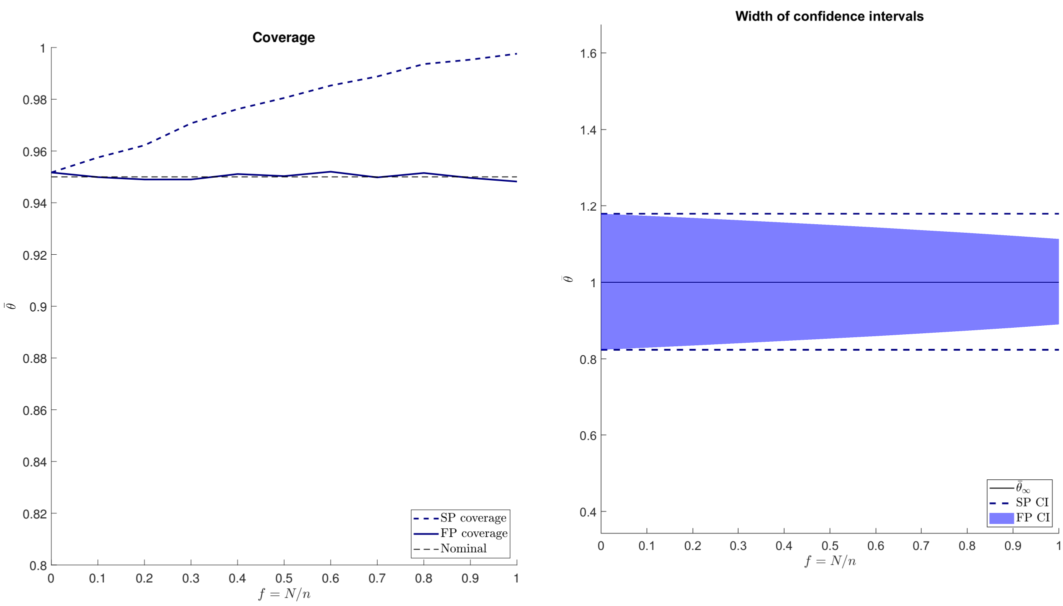


FIGURE 2.2. Results for β_n and the measurement model in equation (2.17): coverage (left) and width (right) of finite-population ("FP", solid lines) and superpopulation ("SP", dashed lines) confidence intervals. Nominal coverage is set to 0.95. Signal-to-noise ≈ 1.5.

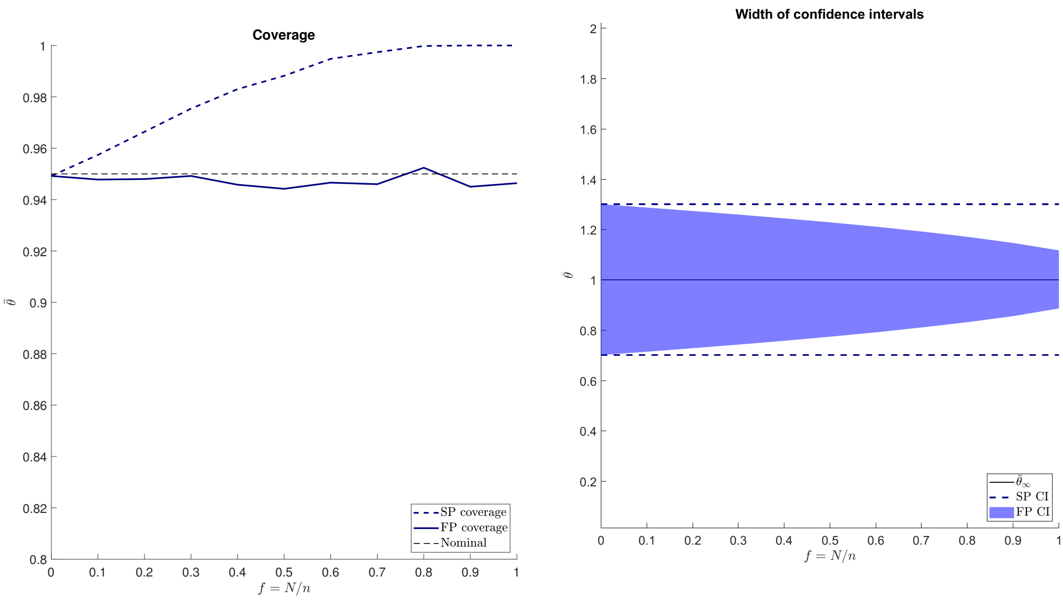


FIGURE 2.3. Results for β_n and the measurement model in equation (2.17): coverage (left) and width (right) of finite-population ("FP", solid lines) and superpopulation ("SP", dashed lines) confidence intervals. Nominal coverage is set to 0.95. Signal-to-noise ≈ 2.5.

2.5 Empirical illustrations

In this section, I illustrate how the methods in the previous sections allow for a systematic approach to uncertainty quantification in finite populations by considering two setups that span the wide range of empirical applications for which this chapter is relevant.

2.5.1 Predicting police violence

The first exercise is based on Montiel-Olea et al. (2021), which are interested in the determinants of the use of deadly force by police officers in the United States. More generally, it is aimed at illustrating finite-population inference in the microforecasting literature (??Giacomini et al., 2023), which is concerned with prediction of individual outcomes in short panels.

Data and background. The authors collect data on all local police departments in the United States, defined as those that serve a well-defined population. They use census records from the Law Enforcement Agency Identifiers Crosswalk dataset (LEAIC), and the final dataset contains $N = 7,585$ agencies.[25] The authors are interested in characterizing the determinants of police use of deadly force in the U.S. at the department level and aim at retrieving comprehensive records for all such departments: $f = 1$ is arguably a reasonable description of the sampling framework.

Here interest is a on a composite index θ_i for the agency-specific baseline level of lethal encounters, which is specified as an exponential model including observed, candidate determinants Z_i and unobserved, residual attributes α_i; see again equation (2.4). The measurement system is specified as a multiplicative (Poisson)

[25] The authors have made the data and code publicly available at https://github.com/jm4474/EmpiricalBayesCounterfactuals.

model for the number of lethal encounters, which can be recovered by setting $g_0 = 0_T$ and $g_1(X_i; \delta) = \exp(X_i\delta)$ in equation (2.5). In a slightly rearranged form, we have

$$Y_{it} = \alpha_i \exp\left(Z_i'\beta_n + X_{it}\delta\right) + \varepsilon_{it} \tag{2.18}$$

for $t = 1, \ldots, 6$, corresponding to yearly measurements over 2013–2018. In the main specification, $\dim X_{it} = 1$ and X_{it} are murders per 100,000 population served. Throughout the period, 1, 179 agencies have at least one lethal encounter, for a total number of 3, 504 homicides. The parameters of interest are $\gamma_n = (\delta, \beta_n')'$. The method-of-moments estimator $\hat{\gamma}$ solves a moment condition of the form (2.9).[26]

The ultimate objects of interest are counterfactual lethal encounters for agency i that would obtain if we were to replace some of its observed and unobserved intrinsic characteristics — encapsulated in Z_i and α_i, respectively — with those of agency j. Since the latter remain unobserved, the authors propose an estimator based on an Empirical Bayes approach, the Poisson model and the assumption of weakly dependent measurements. For our purposes, what matters is that this is a known mapping of estimated coefficients to the predicted number of lethal encounters:

$$\hat{Y}_t^*(i, j, z) = \frac{\bar{Y}_j + 1}{\exp\left\{Z_j'\hat{\beta}\right\} \sum_{t=1}^{T} \exp\left\{X_{jt}\hat{\delta}\right\}} \exp\left\{z'\hat{\beta} + X_{it}\hat{\delta}\right\}, \tag{2.19}$$

where (i, j, z) denotes a counterfactual for agency i if it had the unobserved characteristics of agency j and the observed characteristics z.[27] For instance, $\hat{Y}_t^*(i, i, z_j)$ are estimated counterfactuals for agency i with its own unobserved determinants (interpreted as selection and training practices and departmental culture) and with the observed characteristics of agency j.

Results. Table 2.1 reports the point estimates and standard errors for $\hat{\gamma}$; the first row corresponds to X_{it} and the subsequent entries correspond to the predictors Z_i. The second and third columns report conventional and finite-population standard errors, respectively. Conventional standard errors are based on standard method-of-moments variance estimators, which correspond to the diagonal entries of the square root of $\hat{\mathrm{V}}(0)$ in equation (2.14). Finite-population standard errors are constructed by calculating $\hat{\mathrm{V}}(1)$ assuming (conditionally) uncorrelated measurements, a maintained assumption in Montiel-Olea et al. (2021).

Table 2.1 shows that ignoring the finite-population dimension of the problem leads to standard errors that are between 1.2 and 2.5 times larger than the finite-population ones, a byproduct of introducing sampling-based uncertainty. For example, the standard error on the estimated coefficient associated to the poverty share goes from 0.003 to 0.007. Differences in the magnitude of the change can be traced back to how each particular predictor loads on signal-to-noise. Note that the standard errors on the coefficient associated to murders per 100,000 population served are unchanged: the Finite Population Correction is zero for common parameters, along the lines of our discussion in Section 3.3.

[26] In particular, we have discussed how to write β_n as a finite-population estimand in the sense of equation (2.5) in Remark 2.2. This leads to moment conditions for β_n (given δ) as in eq. (10) in Montiel-Olea et al. (2021). For the common parameter δ, I follow the authors and set $A(X_i, \delta) = X_i'$ in the moment function in (2.8); see eq. (8) in Montiel-Olea et al. (2021).

[27] Note that the confidence intervals discussed in this application are valid for the counterfactual Empirical Bayes estimand rather than the the infeasible one using the true θ_i, in the spirit of Ignatiadis and Wager (2022).

TABLE 2.1. Estimates of the parameters in equation (2.18).

	Coefficient	Conventional s.e.	FP s.e.
Murders per pop. (in hund. ths.)	0.005	0.003	0.003
Log of avg. pop. (in m.)	1.192	0.049	0.036
Officers per pop. (in ths.)	0.012	0.004	0.004
Gun death rate (%)	0.049	0.01	0.004
Share in poverty (%)	0.04	0.007	0.003
Share black (%)	-0.024	0.004	0.002
Garner	-0.031	0.127	0.102
LEOBR	-0.05	0.113	0.066
Land area (sq. km. per m.)	1.0231e-05	1.1511e-06	7.4896e-07

Notes: The first row corresponds to the time-varying variable X_{it} in equation (2.18); the rest are time-invariant predictors Z_i. "Garner" are dummy variables indicating the severity of state laws on the use of deadly force and "LEOBR" are dummy variables for state laws protecting police from misconduct allegations, see Montiel-Olea et al. (2021) for additional details. "*hund. ths.*" stands for "hundred thousands", "*m.*" stands for 'millions' and "*sq. km.*" for "square kilometers". The second and third columns report baseline standard errors (as in Montiel-Olea et al. (2021)) and finite-population standard errors for $f = 1$, respectively.

Importantly, this is just a first step towards computing counterfactuals. Statistical significance is not necessarily of interest here; instead, Table 2.1 is relevant in that uncertainty in the estimated counterfactuals in equation (2.19) stems directly from the covariance matrix of these estimated coefficients. The authors consider different types of counterfactuals; here we focus on those of the form $\hat{Y}_t^*(i, j, z_j)$, where both observed and unobserved characteristics of agency i are replaced with those of agency j.[28]

Table 2.2 reports the results for four of the ten largest departments according to population served (Phoenix, Chicago, Philadelphia and New York) and for these ten combined ("Totals"). A full list of counterfactuals is reported in Appendix A.3.1. Note that Table 2.2 directly reports prediction intervals rather than point estimates, calculated by drawing from the estimated asymptotic distribution. This is in line with the authors' emphasis on quantifying estimation uncertainty, something that makes this application particularly interesting for our purposes. Rows correspond to agency i and columns to agency j; the diagonal elements are the actual number of realized lethal encounters during the period. For instance, we might ask the following question:

"What would happen to the number of lethal encounters if all ten largest agencies had the department-specific attributes of the Chicago Police Department?"

We can read this off Table 2.2: the 90% finite-population prediction interval is $(565, 667)$, and the number of lethal encounters is thus expected to increase from a (realized) baseline of 548 encounters during the period. The answer to this question is however inconclusive if we were to calculate these prediction intervals as if the U.S. local police departments were a small subset of a much larger superpopulation: the 90% finite-population prediction interval increases to $[545, 700]$. Not only is the finite-population interval 34% smaller than the conventional one, it also leads to substantively different policy directions for the questions that the authors seek to answer.

[28] In particular, we consider the counterfactual values of *Officers per pop.*, *Gun death rate*, *Share in poverty*, *Garner* and *LEOBR* (see Table 2.1).

The discussion here illustrates that finite-population inference identifies the right source of estimation uncertainty for this problem — the fact that we only observe error-ridden measurements of agency-specific baseline police violence — and that Finite Population Corrections can lead to substantially more precise and meaningful uncertainty assessments.

TABLE 2.2. Counterfactual homicides: observed and unobserved determinants (selection of departments)

	Phoenix	Chicago	Philadelphia	New York
Phoenix	**93**	[28,32]	[21,30]	[5,9]
		(28,32)	(23,28)	(6,8)
Chicago	[189,216]	**63**	[46,64]	[12,19]
	(190,214)		(49,61)	(13,18)
Philadelphia	[89,130]	[29,40]	**28**	[7,9]
	(96,120)	(30,38)		(7,9)
New York	[568,1013]	[184,310]	[175,236]	**55**
	(643,870)	(204,275)	(180,228)	
Totals	[1689,2279]	[545,700]	[481,567]	[125,166]
(548)	(1791,2094)	(565,667)	(481,564)	(134,155)

Note: The agencies above are a selection of those in Table A.1 in Appendix A.3.1, which cover the ten largest departments by population served. Diagonal entries are observed lethal encounters (totalling 548 encounters for the top ten departments). Off-diagonal entries are 90% confidence intervals for counterfactual values of lethal encounters $\hat{Y}_t^*(i, j, z_j)$ in equation (2.19), which replace characteristics of agency i in the rows with those of agency j in the columns; see the text for additional details. Baseline prediction intervals (as in Montiel-Olea et al. (2021)) are reported in brackets and finite-population prediction intervals for $f = 1$ are reported in parenthesis.

2.5.2 Misallocation

The second exercise is motivated by the large literature on resource misallocation, which is based on the observation that differences in aggregate TFP might not be driven solely by technology but also by allocative efficiency. Following Hsieh and Klenow (2009), an extensive body of work has provided evidence of substantial heterogeneity in revenue productivity within industries, which under appropriate conditions can be used to quantify the extent of misallocation. See Restuccia and Rogerson (2017) for a review.

Exploring the sources of misallocation and obtaining aggregate summary statistics requires a combination of rich microdata and careful measurement, which makes this an appealing framework to illustrate the methods in this chapter. In remarkable contrast to the previous exercise, here the literature has often understated or ignored estimation uncertainty.

Data and background. For illustration, consider the monopolistic competition framework in Hsieh and Klenow (2009), where firms hire labor and capital in competitive markets, have Cobb-Douglas production functions and might face output, capital or labor distortions such as output subsidies, differential access to credit or labor market regulations. These create "wedges" relative to the efficient allocation, which manifests in heterogeneity in the marginal revenue product of capital and labor (MRPK and MRPL, respectively) within a given industry.

Measuring these firm-level wedges is challenging. Even if marginal revenue products can be measured in the data, within firm variation over short periods of time might reflect measurement errors, adjustment costs or transitory shocks. A popular approach is to focus on persistent–transitory decompositions such as firm fixed-effects in marginal revenue products as measures of these underlying wedges (David and Venkateswaran, 2019; Chen et al., 2022; Adamopoulos et al., 2022; Chen, Restuccia, and Santaeulàlia-Llopis, 2023; Nigmatulina, 2023).

For this exercise, I use data from the Statistik Industri, an annual census of all formal manufacturing firms in Indonesia with more than 20 employees. I follow Peters (2020), who focuses on firms that enter after 1990 and is interested in heterogeneous markups — a particular form of misallocation — to motivate a model of firm dynamics and market power.[29]

This leads to an unbalanced panel of about $N = 17,000$ firms, which also comprise the population of interest, for the period 1991–2000. Motivated by the literature above, I consider the following measurement model for log labor wedges:

$$\log \widetilde{\mathrm{MRPL}}_{it} = \theta_i + \varepsilon_{it}, \tag{2.20}$$

where $\log \widetilde{\mathrm{MRPL}}_{it}$ is log MRPL demeaned with respect to industry averages and where θ_i are firm-level wedges.[30] This is a natural formulation in a context where the distortions of interest are persistent market features such as frictions or regulations.

I then use this framework to compute popular misallocation statistics. First, I explore the relationship between labor distortions and firm size (labor force) in line with similar exercises in the literature (Gorodnichenko et al., 2021; Yeh et al., 2022). The finite-population estimands β_n are then least-squares coefficients from the projection of θ_i on firm size bins Z_i; I group firms into ventiles according to their position in the size distribution at entry.

Second, I calculate measures of allocative efficiency, or the aggregate TFP loss associated to the extent of misallocation. An often-used formula that has a closed form expression under normality (Hsieh and Klenow, 2009; Gorodnichenko et al., 2021) is

$$d \log \mathrm{TFP} = -\left(\frac{\alpha(1-\alpha)}{2} + \frac{(1-\alpha)^2 \sigma}{2} \right) \mathrm{Var}_n\left(\theta_i\right), \tag{2.21}$$

where $1-\alpha$ is the labor share and σ is the elasticity of substitution. (I follow Hsieh and Klenow (2009) and set $\sigma = 3$ and $\alpha = 0.33$.) The finite-population estimand here is $\beta_n = \mathrm{Var}_n\left(\theta_i\right)$, the dispersion of labor wedges across firms in the economy.

[29] Many firm surveys in developing countries have such a size-based/formal employer cutoff. This qualifies the population of interest and complicates measuring the extensive margin. Peters (2020) argues that a new firm in the census is also an entrant to the relevant product markets to the extent those are the ones formal firms compete in; see Section 3.1 in the chapter for additional discussion. The data and replication files are available online at https://onlinelibrary.wiley.com/doi/full/10.3982/ECTA15565.

[30] In particular, under Hsieh and Klenow (2009) marginal revenue products can be measured up to scale via average revenue products, which are directly available in most datasets. As usual, labor wedges are here identified up to a normalization with respect to other firm-level frictions. Here labor is measured via the wage bill instead of the number of employees and log MRPL is demeaned with respect to narrowly-defined industry indicators and time dummies, following Peters (2020). Finally, note that model (2.20) is a representative specification, but more general formulations are possible along the lines of Section 3.3.

As discussed in Remark 2.8, our baseline setup does not allow for such objects without further assumptions: while conceptually the problem is identical, the class of estimands considered rules out nonlinear transformations of the latent attributes. In Appendix A.2.2, I extend the framework to cover $\text{Var}_n(\theta_i)$ in the context of this application. The corresponding Finite Population Corrections rely on the same notion of weak dependence across measurements as in Assumption 2.1 and do not require additional measurements. We do need to limit the higher order dependence of measurement errors on latent attributes. This is not surprising: similar assumptions are needed in any deconvolution-like exercise when interest is in nonlinear features or higher-order moments, see Arellano and Bonhomme (2012) for further discussion.

Results. Figure 2.4 show the estimated relationship between labor wedges and firm size at entry, together with finite-population confidence intervals under the benchmark of conditionally uncorrelated measurements.[31] I also report the confidence intervals that would obtain if we were to treat the population of young formal Indonesian firms as a negligible fraction of some hypothetical superpopulation.

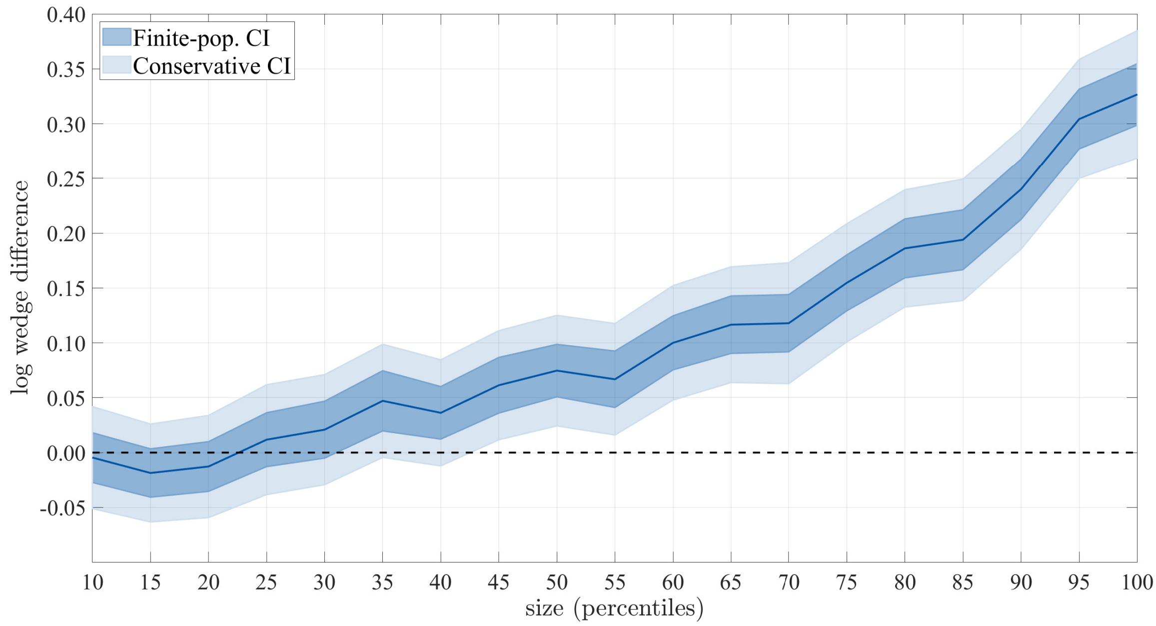


FIGURE 2.4. Labor wedges across the distribution of firm size at entry (relative to 5th percentile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates.

Overall, the results suggest a positive relationship between labor-related distortions and firm size, which might be indicative of size-dependent regulations that tend to distort the optimal allocation of labor (Guner et al., 2008). Estimation uncertainty is however not negligible: the relationship is quite noisy overall, and not statistically significant up to the 35% percentile. Ignoring uncertainty altogether would seem to suggest a stronger positive relationship; treating the population as a small sample from an infinite superpopulation would rule out much of a relationship in the bottom half of the distribution. Instead, finite-population in-

[31]This exercise also illustrates the applicability of our methods to unbalanced panels, with data missing at random. In particular, note that Assumption 2.1 applies unit-by-unit, and that the finite-population adjustments only appear in the unit-level weighted contributions to the variance in equation (2.13). As such, a simple modification to our framework allowing for unit-specific selection matrices $S_{i,(m)}$ according to the number of available measurements (and similarly for projection matrices) would do.

ference correctly identifies the nature of estimation uncertainty in this context — the measurement problem in model (2.20).

Consider now $\beta_n = \mathrm{Var}_n(\theta_i)$ and let $Y_{it} = \log \widetilde{\mathrm{MRPL}}_{it}$ and $\bar{Y} = N^{-1}\sum_{i=1}^{n} R_i \bar{Y}_i$. In empirical work, an often-reported object is the variance of the estimated firm fixed effects, which ignores the measurement problem:

$$\tilde{\beta} = \frac{1}{N}\sum_{i=1}^{n} R_i \left(T^{-1}\sum_{t=1}^{T} Y_{it} - \bar{Y}\right)^2 .$$

Consider the following alternative. Let $Q_i^* = I_{T^2} - T^{-2} 1_{T^2 \times T^2}$ and define $S_{(T)}$ as the selection matrix that has zeros everywhere but at positions $(1,1), (T+2,2), \ldots, (T^2, T)$. Letting $\hat{Y}_i^* = (Y_i - 1_T \bar{Y}) \otimes (Y_i - 1_T \bar{Y})$, a consistent estimator of the population-wide dispersion in θ_i is

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{n} R_i T^{-2} 1_{T^2}' \left[I_{T^2} - S_{(T)} \left(Q_i^* S_{(T)}\right)^{\dagger} Q_i^* \right] \hat{Y}_i^* . \tag{2.22}$$

Note that we are now imposing independence over measurements at the estimation step — a form of Assumption 2.1 (Arellano and Bonhomme, 2012). See again Appendix A.2.2 for additional details. Given this, an estimate of $d \log \mathrm{TFP}$ in equation (2.21) is readily available. I explore the evolution of this measure of allocative efficiency at entry over 1991–1999, analogous to similar exercises in empirical work (García-Santana, Moral-Benito, Pijoan-Mas, and Ramos, 2020; Bils, Klenow, and Ruane, 2021).[32] Through the lens of this framework, I characterize the second moments of a sequence of evolving finite populations.

Figure 2.5 shows the results for firms in the bottom quartile of the size distribution, a group for which labor-related distortions do not seem to differ systematically based on the number of employees.

The figure shows that the aggregate productivity losses from misallocation (if the economy-wide distortions were like those of entering firms) are of the order of 12–15%, with a slight upward trend over time. Importantly, this is a revised down estimate of around five percentage points in every cohort relative to the standard calculation that does not take the measurement problem into account (reported in gray in the figure). This is also lower than the observed dispersion in MRPL in the data (around 40% in terms of equation (2.21)). These magnitudes are broadly consistent with conventional wisdom that a large part of observed dispersion in total factor revenue productivity is in the firm fixed effect, but emphasize the role of measurement error and transitory shocks. Furthermore, these differences are statistically significant, but estimation uncertainty is here far from negligible: as an example, the change in TFP is within a confidence band of 14–20% for the 1999 cohort. In fact, the difference relative to the uncorrected estimates is only marginally significant if one were to add sampling uncertainty on top of the finite-population confidence bands.

In Appendix A.3.2, I report additional results allowing for dependence over measurements, with similar implications. If anything, finite-population confidence intervals tend to be wider. This stresses the importance of finite-population inference in this context — where sampling uncertainty is indeed small (even if one treats the population as a negligible fraction of some hypothetical superpopulation) while measurement

[32] Specifically, I calculate (2.22) for each entry cohort over this period. Note that we still use all measurements for each firm in estimation — this is what allows us to separate the persistent component from measurement errors. Finally, note that in order to calculate (2.22) at least two (independent) measurements are needed, which means that we cannot report results for firms in the 2000 cohort.

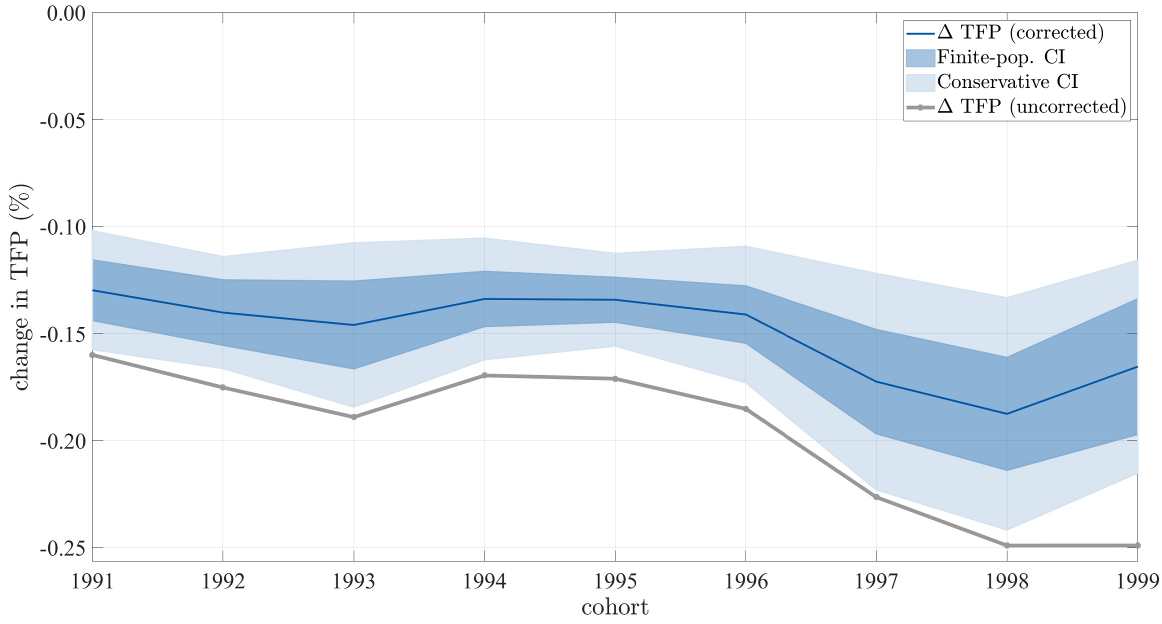


FIGURE 2.5. Evolution of allocative efficiency as in equation (2.21) for each cohort (within firms in the bottom size quartile). Cohorts refer to the year of entry but all repeated measurements for each firm are used in estimation. 95% confidence bands (finite-population and conservative) are displayed together with the point estimates.

uncertainty remains sizeable.

All in all, these results illustrate that the methods presented in this chapter provide guidance on the relevant sources of estimation uncertainty yet again — this time in a context where the contrast between sampling and measurement is particularly salient and where the conventional approach to inference has been to understate rather than exaggerate estimation uncertainty.

CHAPTER 3

MICRO RESPONSES TO MACRO SHOCKS

WITH MARTÍN ALMUZARA

3.1 Research context

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

A popular approach is to formulate estimating equations of the form

$$Y_{i,t+h} = \beta(h) s_i X_t + \text{controls} + v_{h,it}, \tag{3.1}$$

where Y_{it} is a *micro outcome* for unit i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, T$), such as household income or firm sales, and X_t an observed *macro shock* of interest, such as a monetary policy or oil supply shock. Shocks are often interacted with unit-level covariates s_i to document heterogeneity in transmission along observables. Estimates $\hat{\beta}(h)$ of the response at horizon h are then obtained via least squares; a panel local projections 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 it well: in our own survey of almost 50 recent papers, around half compute two-way clustered standard errors, one-third cluster within units only, and many others resort to Driscoll and Kraay (1998). 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 chapter, 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 signal-to-noise of macro shocks in the microdata. As a result, a very simple recipe for inference emerges: clustering standard errors at the time level and ex-ante including sufficient lags as controls. We refer to this strategy as *time-clustered lag-augmented* *heteroskedasticity-robust (t-LAHR) inference*.

We establish our results in a comprehensive setup that features observed and unobserved macro and micro shocks, cross-sectional heterogeneity in responses, general forms of serial dependence in outcomes, and unrestricted signal-to-noise. We first show that $\hat{\beta}(h)$ recovers the slope coefficient of a population linear projection of unit-specific impulse responses on characteristics s_i, thereby formalizing what practitioners have in mind when including interactions in equation (3.1). If $s_i = 1$, the estimand boils down to the average response in the population. Notably, since we place no restrictions on the underlying impulse-response heterogeneity or s_i, our characterization of the estimand is in effect nonparametric.[1]

Signal-to-noise. The degree of signal-to-noise of macro shocks in the microdata is a crucial parameter of the problem. Here, it is *common* shocks to all units that drives identification, and how sizable they are relative to micro shocks determines both the strength of identifying variation and the extent of unaccounted-for spatial dependence.[2] The notion of different signal regimes also reflects the scope of empirical work, which takes interest in atomistic and granular agents, administrative and narrow datasets, unit-specific and aggregate regression controls, etc.

Hence, one of our key contributions is to introduce a novel asymptotic framework where the signal value of aggregates may be arbitrarily low (or high) in the limit. We achieve this by indexing the relative standard deviation of macro to micro shocks by a parameter $\varkappa$ that can drift with the sample size. This device allows for a range of data generating processes (DGPs) in which estimation uncertainty is dominated by micro-level terms, a combination of micro and macro errors, or aggregate components only.[3] On the contrary, standard asymptotic plans where $\varkappa$ is fixed only capture the latter and ignore idiosyncratic shocks, potentially leading to poor approximations in small samples. It is clear then that the nature of estimation error depends on $\varkappa$ and the question is whether inference procedures are robust to different macro signal regimes. Our main result is that *t-LAHR inference is uniformly valid* over $\varkappa$, in other words, t-LAHR confidence intervals have correct asymptotic coverage for the (nonparametric) local projection estimand uniformly over $\varkappa$.

Inference. The key assumption in our framework is the availability of an observed macro shock X_t. Our notion of shocks is that of mean independent innovations with respect to both its own lags and leads and other shocks, in line with the time series literature on local projections inference (Stock and Watson, 2018; Montiel Olea and Plagborg-Møller, 2021). We first focus on the case where the shock of interest is observed — an assumption prevalent in most empirical applications — and then consider settings where the shock of interest is recoverable (spanned by X_t and its lags) or contaminated with measurement error but a proxy is available (as in local projection-instrumental variable estimators; LP-IV for short).[4]

[1]We discuss extensions to (exogenous) time-varying characteristics s_{it} in Section 3.3 (Remark 3.7).

[2]It is immediate that if $s_i = 1$ in equation (3.1), 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. In our exposition, we always allow for time indicators as controls when s_i displays cross-sectional variation.

[3]Our approach also resonates with the renewed interest on the potential for unit-level shocks to explain aggregate fluctuations, as in Gabaix (2011) and subsequent literature. Our device to obtain non-negligible micro errors is closer to Jovanovic (1987) in that we rely on scaling micro variation up rather than on fat-tailed distributions. However, we conjecture that similar inference results can be obtained in the latter under appropriate regularity conditions.

[4]Examples of popular identification methods include narrative approaches (as in Crouzet and Mehrotra, 2020, for monetary policy shocks), high-frequency identification (as in Känzig, 2021, for oil supply shocks) or a combination of Cholesky/structural

The *macro* and *shock* nature of X_t delivers the following observation which serves as a guiding principle throughout the chapter: panel local projections with macro shocks are equivalent to *synthetic* time series local projections with an appropriately aggregated dependent variable. This is true even if shocks interact with covariates s_i and if unit and time effects are included. Therefore, aggregating the microdata by collapsing the cross-sectional dimension of the panel and treating it as a time series delivers valid inference for any κ.[5] This is precisely what t-LAHR inference does, since time-level clustering in the panel problem and heteroskedasticity-robust inference in the synthetic time series problem are essentially equivalent.

The *macro* and *shock* nature of X_t also clarifies the role of lag augmentation. In a panel local projection that controls for p lags of s_iX_t, the regression scores (the product of shocks and residuals) are *nearly* uncorrelated even if residuals are not. Specifically, they are a moving average of order h where the first p autocovariances are zero and the remaining ones are independent of κ. This has two major implications. First, it confers a double layer of simplicity to inference: up to horizon $h \leq p$, there is no need for unit-level clustering or heteroskedasticity and autocorrelation robust (HAR) approaches to deal with serial dependence. Second, it explains why t-LAHR inference might have only small coverage distortions even for horizons exceeding p: these distortions depend on the size of the autocorrelation coefficients of the score, which are small in low-signal environments. In fact, if the DGP is well approximated by a low-order vector autoregression (VAR), we prove t-LAHR inference is uniformly valid over both κ and $h \propto T$, a result reminiscent of those in Montiel Olea and Plagborg-Møller (2021) for time series local projections.

We complement our theoretical results with simulations for realistic designs and sample sizes, allowing for moderately long horizons and substantial persistence in micro shocks. We study the performance of a battery of approaches, including an alternative to t-LAHR that substitutes lag augmentation with HAR inference, and incorporating small-sample refinements (Müller, 2004; Imbens and Kolesár, 2016; Lazarus, Lewis, Stock, and Watson, 2018). We find that t-LAHR inference shows remarkable performance relative to all other competitors, particularly in low-signal environments, in near non-stationary scenarios, and over moderate horizons even if we do not impose a VAR on outcomes.[6] In practice, we recommend to supplement t-LAHR inference (controlling for a reasonable number of lags of both outcome and shock variables) with the refinement proposed by Imbens and Kolesár (2016).

Empirical survey and illustration. We reviewed a large body of empirical research that precedes our work. The typical application uses administrative data for firms, tracks units at the quarterly or annual frequency for a limited number of periods, and estimates impulse responses to monetary policy shocks via local projections. Most applications include interactions of the form s_iX_t and both unit and time fixed effects, but vary widely in the number and nature of additional controls.[7]

VAR restrictions (as in Drechsel, 2023, for firm investment shocks). See Ramey (2016, Section 2.3) for a review of identification methods in macroeconometrics.

[5]This synthetic time series representation is also illustrative of the fact that the concentration rate of the estimation error is at most $T^{-1/2}$, even in situations where $N \gg T$. This suggests caution regarding the conventional wisdom in many empirical applications that a larger cross-sectional dimension somehow compensates for a shorter time series.

[6]It is known that ad-hoc parameter choices and small-sample biases in sample autocovariances contribute to the subpar relative performance of HAR estimators (Herbst and Johansenn, 2023).

[7]We reproduce the full list in Appendix B.3 which includes 47 empirical papers that run panel regressions with macro shocks. A few focus on the case $h = 0$ only, but the vast majority compute impulse responses over several horizons. The economic content of X_t is very diverse, including fiscal policy shocks, investment shocks, TFP and innovation shocks, carbon pricing shocks,

In otherwise comparable empirical designs, we document large dispersion in the way practitioners compute standard errors: among 47 different papers, 24 compute two-way clustered standard errors (within units and time), 15 cluster within units only, 7 use Driscoll and Kraay (1998) and 1 clusters within time only.

These choices reflect very different views on the dominant sources of statistical uncertainty, from ruling out serial dependence to ruling out spatial dependence; from a suggestion that both unit-level and aggregate dynamics need to be accounted for to an explicit stance that either of the two dominates. Often these choices are made with little discussion or citing previous work as justification.[8] Our framework allows us to revisit them. First, off-the-shelf autocorrelation consistent methods such as Driscoll and Kraay (1998) leave information on the table (the autocovariance function of the regression scores is known), which comes at a cost in small samples. Second, validity in the case where standard errors do not explicitly adjust for serial dependence (as in two-way clustering) boils down to whether a reasonable number of lags was included in estimation. Third, clustering within units is superfluous, even in low-signal regimes where the size of unit-level dynamics is comparable to that of aggregates. Fourth, clustering only within units breaks down even in the face of small amounts of spatial dependence induced by aggregate shocks (that is, moderate-signal environments). In fact, we offer a way to reinterpret these confidence intervals as providing valid inference for an estimand indexed to the actual realizations of aggregate shocks during the sample period.

Finally, we illustrate our methods in an empirical exercise inspired by a booming literature that investigates the role of financial frictions and firm heterogeneity in the transmission of monetary policy.[9] The exercise highlights the importance of the choice of inference method, and the value of the synthetic time series representation as a way to gain intuition about the source of identifying variation.

Related literature. This chapter contributes to various strands of the literature.

First, it relates to the time series literature on inference for local projections (Hansen and Hodrick, 1980; Jordà, 2005; Stock and Watson, 2018; Montiel Olea and Plagborg-Møller, 2021; Lusompa, 2023; Xu, 2023; Montiel Olea, Plagborg-Møller, Qian, and Wolf, 2024). Relative to this literature we are (to our knowledge) the first to deal with the panel data case with aggregate shocks.[10]

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 in the data and horizon h. They also postulate mean independent innovations, the same type of assumption

temperature shocks, etc. In these applications, the cross-sectional dimension is usually orders of magnitude larger than the effective time-series dimension. In our review we leave out empirical work with very small cross-sections where entities are meaningful and a unit-by-unit treatment is feasible. Nonetheless, when these units are pooled together, as in Fukui, Nakamura, and Steinsson (2023), our results still apply.

[8] The availability of a large cross-sectional dimension and the interaction of shocks with covariates s_i are also often argued as sources of large gains in statistical precision, also reflecting an implicit stance on the presence of macro shocks. We elaborate on the (im)plausibility of these notions in Remark 3.5.

[9] For instance, Crouzet and Mehrotra (2020), Ottonello and Winberry (2020), Anderson and Cesa-Bianchi (2024) and Jeenas and Lagos (2024) target impulse responses of firm investment to monetary policy shocks interacted with external covariates s_i such as firm size, default risk or stock turnover.

[10] Our results on limited serial dependence in regression scores relate to the earlier multi-step forecast literature (Hansen and Hodrick, 1980), which relied on including infinitely many lags to ensure that the forecast errors have a 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.

we impose on X_t. Our Proposition 3.2 (and, more generally, Section 3.3.3) can be interpreted as the panel version of their results. Nonetheless, our focus is on uniformity with respect to the macro signal-to-noise κ, which has no obvious counterpart in the time series setup, 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. Using stylized models, Hahn et al. (2020) bring attention to the drastic consequences of drawing inferences from short panels with aggregate uncertainty. Although our focus is on thought experiments where macro shocks are a key source of identification, we can connect to their results by reinterpreting confidence intervals that exploit independence across units as valid for an approach to inference that conditions on the path of aggregate shocks.

Recent additions to this literature study regional-exposure designs where the researcher has access to low-rank instruments of the form $s_i X_t$ (s_i are region-specific exposures to aggregate conditions) and so the reduced-form equation looks like (3.1) for $h = 0$. Arkhangelsky and Korovkin (2023) argue that exogenous variation comes from the time series shock X_t and focus on threats to instrument validity, whereas Majerovitz and Sastry (2023) consider either s_i or X_t as sources of identification and suggest that inference needs to take spatial dependence into account in the latter case. Our work extends these ideas by giving formal inference results that cover dynamic responses and different macro signal environments.

Third, this chapter relates to the cross-sectional dependence literature that studies models where the scores feature varying degrees of spatial dependence (Driscoll and Kraay, 1998; Andrews, 2005; Pesaran, 2006; Gonçalves, 2011; Pakel, 2019). Our framework falls 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). Moreover, our uniformity result (which translates into robustness to the degree of spatial dependence) is new to the literature.

Outline. Section 3.2 provides an overview of our results in a simple static model, illustrating the role of aggregate shocks and their signal relative to micro shocks. Section 3.3 presents our main inference result in a general, heterogeneous dynamic model. Section 3.4 discusses a comprehensive simulation study and Section 3.5 the empirical illustration. Proofs can be found in Section A.1 with additional details in the Appendix. A MATLAB code repository is available online.[11]

3.2 Simple model

We illustrate the main points of the chapter in a simple, static regression model with homogeneous responses. We keep the exposition simple and omit technical details with the goal of building insights. The more general setup is studied in Section 3.3.

[11] https://github.com/TinchoAlmuzara/PanelLocalProjections.

Model assumptions. We observe a micro outcome Y_{it} and a macro shock X_t for units $i = 1, \ldots, N$ and over periods $t = 1, \ldots, T$. They are related by

$$\begin{aligned} Y_{it} &= \beta_0 X_t + v_{it}, \\ v_{it} &= Z_t + \kappa u_{it}, \end{aligned} \tag{3.2}$$

where v_{it} is an error term including both aggregate and idiosyncratic unobservables, denoted Z_t and u_{it}, respectively. Here κ regulates their relative importance in the micro data, as explained below. The goal is to estimate and do inference on β_0.

This simple model is a stylized representation of an empirical setting where we are interested in the transmission of aggregate uncertainty to individual outcomes; the effect of X_t on Y_{it}. Examples of the former include changes in interest rates, tax regulations or oil prices, which might leave a mark on household consumption, worker's labor income or firm sales. In fact, one could entertain any combination of macro variables and micro outcomes in these examples. When interest centers around one aggregate variable — captured by X_t — it would be hard to ex-ante rule out the presence of any others — embedded in Z_t. This basic premise is at the core of the our results.

We now make two sets of assumptions, later generalized in Section 3.3 to allow for observable and unobservable heterogeneity, and more flexible dynamics.

Assumption 3.S3 (Stationarity and iidness in the simple model).

(i) $\{X_t, Z_t, \{u_{it}\}_{i=1}^N\}_{t=1}^T$ *is stationary.*

(ii) $\{\{u_{it}\}_{t=-\infty}^{\infty}\}_{i=1}^N$ *are i.i.d. over i conditional on* $\{X_t, Z_t\}_{t=1}^T$.

Assumption 3.S3(i) implies Y_{it} is stationary too. Assumption 3.S3(ii) simply assigns the role of inducing cross-sectional dependence in the error term v_{it} to Z_t.[12]

Assumption 3.S4 (Shocks and independence in the simple model).

(i) $E\left[X_t \middle| \{X_\tau\}_{\tau\neq t}, \left\{Z_\tau, \{u_{i\tau}\}_{i=1}^N\right\}_{\tau=1}^T\right] = 0.$

(ii) $E\left[Z_t \middle| \{Z_\tau\}_{\tau\neq t}, \left\{X_\tau, \{u_{i\tau}\}_{i=1}^N\right\}_{\tau=1}^T\right] = 0.$

(iii) $E\left[u_{it} \middle| \{u_{i\tau}\}_{\tau\neq t}, \left\{X_\tau, Z_\tau\right\}_{\tau=1}^T\right] = 0.$

Assumption 3.S4 implies X_t, Z_t and u_{it} are mutually unpredictable and serially uncorrelated. Assumption 3.S4(i) is ultimately an identification condition, whereas 3.S4(ii) and 3.S4(iii) are made for symmetry. Indeed, mutual unpredictability of macro shocks lies at the core of macroeconometrics and is typically necessary to give structural interpretation to impulse-response calculations (see, for instance, Ramey, 2016; Stock

[12] Both assumptions can be relaxed; we briefly discuss departures from 3.S3(i) in Section 3.3 and 3.4. Allowing for weak spatial dependence in u_{it} in place of 3.S3(ii) is also possible with minor modifications.

and Watson, 2016; Plagborg-Møller and Wolf, 2021).[13] Assumption 3.S4(i) is an empirically realistic starting point, since in the majority of applications X_t is the (perhaps noisy) measurement of a shock.

Remark 3.1 (Relaxing Assumption 3.S4). In practice, we might only observe a proxy shock X_t^*, which may be contaminated with measurement error or possess some residual autocorrelation structure, say $X_t^* = \sum_{\ell=1}^k \alpha_\ell X_{t-\ell}^* + X_t$ for known $k < \infty$. These cases can be handled by treating X_t^* as an instrument — a panel version of the LP-IV estimator (Stock and Watson, 2018, Section 1.3), which we study in Section 3.3.4 — or by including lags of X_t^* as controls, see also Section 3.3.3.

Estimation and inference. A natural estimator of β_0 is pooled least squares,

$$\hat{\beta} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_t Y_{it}}{\sum_{i=1}^N \sum_{t=1}^T X_t^2} = \frac{\sum_{t=1}^T X_t \left(N^{-1} \sum_{i=1}^N Y_{it}\right)}{\sum_{t=1}^T X_t^2},$$

which is also a panel local projection (LP) estimator at horizon $h = 0$ and the estimator in a time series regression involving the synthetic outcome $\hat{Y}_t = N^{-1} \sum_{i=1}^N Y_{it}$ and X_t. The double nature of $\hat{\beta}$ as panel and time series estimator arises naturally in the presence of macro shocks, as we further demonstrate in Section 3.3.

Denote the residual by $\hat{\xi}_{it} = Y_{it} - \hat{\beta} X_t$. A key takeaway from this chapter is that a reliable approach to inference uses the time-level cluster heteroskedasticity-robust standard error $\hat{\sigma}$, given by $\hat{\sigma}^2 = \hat{V}/T\hat{J}^2$ where $\hat{J} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T X_t^2 = T^{-1} \sum_{t=1}^T X_t^2$ is the least squares denominator and

$$\hat{V} = \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{N} \sum_{i=1}^N X_t \hat{\xi}_{it} \right)^2 .$$

Another sign of the duality between panel regressions with aggregate shocks and time series regression is that $\hat{\sigma}$ is also the usual Eicker–Huber–White standard error computed using the synthetic time series residuals $\hat{\xi}_t = N^{-1} \sum_{i=1}^N \hat{\xi}_{it}$.

As mentioned in the Introduction, two popular inferential choices in applications are based on one-way (unit-level) cluster and two-way (unit- and time-level) cluster standard errors, $\hat{\sigma}_{1W}$ and $\hat{\sigma}_{2W}$, given by $\hat{\sigma}_{1W}^2 = \hat{V}_{1W}/T\hat{J}^2$ and $\hat{\sigma}_{2W}^2 = \hat{V}_{2W}/T\hat{J}^2$ where

$$\hat{V}_{1W} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T X_t \hat{\xi}_{it} \right)^2, \qquad \hat{V}_{2W} = \hat{V} + \hat{V}_{1W} - \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N X_t^2 \hat{\xi}_{it}^2.$$

These standard errors reflect different concerns about the nature of estimation error or, more precisely, the correlation of the regression score $X_t v_{it}$ over units and time.

[13]Mean independence assumptions with respect to past and future innovations are a slight strengthening of the more standard martingale difference assumptions, and are convenient in representations where both leads and lags of the variable might enter the model, cf. Montiel Olea and Plagborg-Møller (2021, Assumption 1) in a similar context of local projection inference. This still allows for dynamics on the second- or higher-order moments given the paths of other shocks. It permits that, say, monetary, fiscal or oil supply shocks (X_t, Z_t) increase the variance of household-level income (Y_{it}) via higher order dynamics in u_{it}.

Substituting (3.2), the estimation error decomposes as

$$\hat{\beta}-\beta_0=\underbrace{\frac{\sum_{t=1}^T X_t Z_t}{\sum_{t=1}^T X_t^2}}_{O_p(1)}+\frac{\kappa}{\sqrt{N}}\underbrace{\left(\frac{1}{\sqrt{N}}\frac{\sum_{i=1}^N\sum_{t=1}^T X_t u_{it}}{\sum_{t=1}^T X_t^2}\right)}_{O_p(1)}, \tag{3.3}$$

i.e., as the sum of macro and micro components. The former induces cross-sectional correlation while the latter is uncorrelated across units and both have limited serial dependence — for $t \neq \tau$, $E\left[X_t Z_t \cdot X_\tau Z_\tau\right] = E\left[X_t u_{it} \cdot X_\tau u_{i\tau}\right] = 0$ by Assumption 3.S4(i) and iterated expectations.[14] This is a direct consequence of X_t being a shock.

The intuition for why $\hat{\sigma}$ gives valid inference is the following. If the macro term is not asymptotically small, $X_t v_{it}$ displays correlation over i but not over t, the type of situation for which $\hat{\sigma}$ is designed. If, on the other hand, the micro term dominates, $X_t v_{it}$ is uncorrelated over both i and t. Yet $\hat{\sigma}$ still works: while it does not impose that the cross-sectional covariances of $X_t v_{it}$ are zero, it will correctly estimate them to be zero. One may wish to switch to a non-clustered heteroskedasticity-robust standard error in that case, but we show both analytically (Proposition 3.1) and in simulations (Section 3.4) that there is no loss in simply using $\hat{\sigma}$.

Clearly, correlation over t at the unit-level is never a concern; that is why unit-level clustering either fails or is not needed. In fact, $\hat{\sigma}_{1W}$ is asymptotically equivalent to the non-clustered standard error, and the same holds for $\hat{\sigma}_{2W}$ and $\hat{\sigma}$.

Macro-micro signal-to-noise ratio. Which term dominates the decomposition (3.3) will depend upon $\kappa/\sqrt{N}$. We now provide another interpretation of this quantity. Consider the average outcome $\hat{Y}_t = N^{-1}\sum_{i=1}^N Y_{it}$ and, for the sake of illustration, suppose $\text{Var}\left(Z_t\right) = \text{Var}\left(u_{it}\right) = 1$. By Assumptions 3.S3 and 3.S4, the proportion of the variance of $\hat{Y}_t$ explained by the unobserved macro error can be measured as

$$\bar{R}^2(\kappa) = 1-\frac{\text{Var}\left(\hat{Y}_t | X_t, Z_t\right)}{\text{Var}\left(\hat{Y}_t | X_t\right)} = \frac{1}{1+\kappa^2/N}, \tag{3.4}$$

that is, the signal-to-noise ratio is $O(N/\kappa^2)$. It increases with N since cross-sectional averaging reduces the variance from idiosyncratic errors, but decreases with $|\kappa|$.

We will study estimation and inference in sequences of data generating processes (DGPs) where κ is allowed to grow as $T, N \to \infty$. This leads, in essence, to three regimes. If $\kappa/\sqrt{N} = o(1)$, (such as if κ is fixed), $\bar{R}^2(\kappa) \to 1$ and macro shocks are the only source of aggregate variation; we call this the asymptotically high-signal case. If $\kappa \propto \sqrt{N}$, $\bar{R}^2(\kappa)$ is bounded away from 0 and 1 in the limit and both macro and micro shocks matter for aggregate fluctuations; this is the asymptotically moderate-signal case. Finally, if $\kappa/\sqrt{N}$ diverges, $\bar{R}^2(\kappa) \to 0$, macro shocks are imperceptible and we are in the asymptotically low-signal case.[15]

[14] The lack of serial correlation would remain true even if Z_t and u_{it} were serially correlated.

[15] Of course, letting κ grow with the sample size should not be taken literally — it is simply a device to ensure our approximations suitably interpolate between high and low signal-noise environments. This type of embeddings are common in econometrics; an example which also has a low-signal interpretation is weak IV (Staiger and Stock, 1997).

The intuitive notion of κ-regimes has a natural counterpart in our asymptotic approximations, in that there is a close relation between the contribution of macro shocks to $\hat{Y}_t$ and the nature of estimation error for β_0, as illustrated by (3.2) and (3.4). In particular, the macro term dominates in the high-signal case, the micro term dominates in the low-signal case, and they are roughly balanced in the moderate-signal case. Moreover, it is not always possible to consistently detect what κ-regime applies. It is important then to derive inference procedures that are robust in the sense of uniform validity with respect to κ.[16]

Uniformity over κ. From the decomposition in (3.3), letting $N, T \to \infty$ and under regularity conditions specified in Section 3.3,

$$\sigma_0(\kappa)^{-1}\sqrt{T}\left(\hat{\beta} - \beta_0\right) \xrightarrow{\ \text{d}\ } N(0,1),$$

where

$$\left\{E\left[X_t^2\right]\right\}^2 \times \sigma_0(\kappa)^2 = \begin{cases} E\left[X_t^2 Z_t^2\right], & \text{if } \kappa/\sqrt{N} \to 0, \\ E\left[X_t^2\left(Z_t^2 + \bar{\kappa}^2 u_{it}^2\right)\right], & \text{if } \kappa/\sqrt{N} \to \bar{\kappa}, \\ \left(\kappa^2/N\right) E\left[X_t^2 u_{it}^2\right], & \text{if } \kappa/\sqrt{N} \to \infty, \end{cases}$$

This shows two things. First, the rate of concentration of the estimation error $\hat{\beta} - \beta_0$ is either $\sqrt{T}$ in the high- and moderate-signal cases or $\sqrt{NT}/\kappa$ (i.e., slower than $\sqrt{T}$ and possibly even zero, thus making $\hat{\beta}$ inconsistent) in the low-signal case. Second, the asymptotic distribution of $\hat{\beta}$ changes discontinuously across κ-regimes.

Despite the discontinuity, our main result is that the $(1-\alpha)$ confidence interval $\hat{C}_\alpha = \left[\hat{\beta} \pm z_{1-\alpha/2}\hat{\sigma}\right]$, where z_q is the q-quantile of the standard normal distribution, has correct coverage for β_0 *uniformly* over κ,

$$\lim_{T,N\to\infty} \sup_{\kappa} \left|P_\kappa\left(\beta_0 \in \hat{C}_\alpha\right) - (1-\alpha)\right| = 0. \tag{3.5}$$

where P_κ denotes probabilities for a DGP with a given κ. This is much stronger than pointwise validity, as it implies that the quality of the asymptotic approximation to the coverage probability of $\hat{C}_\alpha$ is itself robust to the κ-regime. Statement (3.5) also means that if sample information about macro shocks is extremely scarce and $\hat{\beta}$ is inconsistent, the length of $\hat{C}_\alpha$ adjusts as needed to reflect the weak macro signal.

One might wonder how much the static nature of (3.2) limits these results. The rest of the chapter will show that they extrapolate to a substantially more general and empirically realistic framework with rich forms of dynamics and heterogeneity.

Remark 3.2 (Inference conditional on aggregate shocks). Ignoring the unobservable macro component in (3.3) when doing inference is equivalent to conditioning on its realization. In that situation, $\hat{\sigma}_{1W}$ is a valid standard error for responses defined by moment restrictions that condition on the realized path of aggregate

[16] We will consider inference procedures that are invariant to rescaling. It follows that all of our results can be equivalently obtained in an embedding that scales down the macro component of the model in (3.2) by κ^{-1}. Put differently, what matters is the relative size of macro and micro components.

shocks during the sample period.[17] In general, this induces an internal/external validity trade-off whereby practitioners may be able to pin down certain parameters very precisely but these might lack generalizability to other contexts.

3.3 General case

In this section, we establish estimation and inference results for impulse responses to aggregate shocks in a general setup featuring observed and unobserved, macro and micro shocks, and unrestricted heterogeneity of individual responses.

We introduce the setup in Section 3.3.1 and state the main results in Section 3.3.2. We treat the important case of finite-order VAR DGPs in Section 3.3.3 and local projections with instrumental variables (LP-IV) in Section 3.3.4. Proofs are developed in Section A.1 with technical lemmas in Appendix B.1.

3.3.1 Setup

The researcher observes an outcome Y_{it}, an aggregate shock X_t and characteristics s_i for units $i = 1, \dots, N$ and over periods $t = 1, \dots, T$. Everything is scalar but it is straightforward to extend the results to the multivariate case. We assume

$$Y_{it} = \mu_i + \sum_{\ell=0}^{\infty} \beta_{i\ell} X_{t-\ell} + v_{it}, \tag{3.6}$$

$$v_{it} = \sum_{\ell=0}^{\infty} \gamma_{i\ell} Z_{t-\ell} + \kappa \sum_{\ell=0}^{\infty} \delta_{i\ell} u_{i,t-\ell}, \tag{3.7}$$

where Z_t and u_{it} are unobserved serially uncorrelated aggregate and idiosyncratic errors. We denote $\beta_i = \{\beta_{i\ell}\}_{\ell=0}^{\infty}, \gamma_i = \{\gamma_{i\ell}\}_{\ell=0}^{\infty}, \delta_i = \{\delta_{i\ell}\}_{\ell=0}^{\infty}$ and $\theta_i = \{\mu_i, \beta_i, \gamma_i, \delta_i\}$. These are draws from a cross-sectional distribution and below we specify conditions so that the infinite sums in (3.6)-(3.7) are well defined with probability one.

Here, θ_i traces out cross-sectionally heterogeneous responses to both aggregate and idiosyncratic shocks, and access to external variables s_i allows the researcher to study their transmission along unit-level observables. Our premise is that there is usually more heterogeneity in θ_i than can be explained by s_i alone and our goal is to characterize estimation and inference in that context.

As in Section 3.2, we consider a range of DGPs indexed by κ to cover different signal-to-noise environments. We also make the following assumptions:

Assumption 3.1 (Stationarity and iidness).

(i) $\{X_t, Z_t, \{u_{it}\}_{i=1}^{N}\}_{t=-\infty}^{\infty}$ *is stationary conditional on* $\{\theta_i, s_i\}_{i=1}^{N}$.

(ii) $\{\theta_i, s_i, \{u_{it}\}_{t=-\infty}^{\infty}\}_{i=1}^{N}$ *is i.i.d. over i conditional on* $\{X_t, Z_t\}_{t=-\infty}^{\infty}$.

[17] A proof and additional details are available upon request. As a practical example, we think of the responses of micro outcomes to monetary and fiscal policies during the COVID-19 pandemic.

Assumption 3.2 (Shocks and mean independence).

(i) $E\left[X_t \middle| \{X_\tau\}_{\tau\neq t}, \left\{Z_\tau, \{u_{i\tau}\}_{i=1}^N\right\}_{\tau=-\infty}^{\infty}, \{\theta_i, s_i\}_{i=1}^N\right] = 0.$

(ii) $E\left[Z_t \middle| \{Z_\tau\}_{\tau\neq t}, \left\{X_\tau, \{u_{i\tau}\}_{i=1}^N\right\}_{\tau=-\infty}^{\infty}, \{\theta_i, s_i\}_{i=1}^N\right] = 0.$

(iii) $E\left[u_{it} \middle| \{u_{i\tau}\}_{\tau\neq t}, \left\{X_\tau, Z_\tau\right\}_{\tau=-\infty}^{\infty}, \theta_i, s_i\right] = 0.$

Assumptions 3.1 and 3.2 generalize 3.S3 and 3.S4 to accommodate the presence of both unobserved heterogeneity and external covariates. Assumption 3.2 requires them to be strictly exogenous with respect to shocks. Importantly, the joint distribution of (θ_i, s_i) is left unrestricted, and so is that of $\{\theta_i, s_i\}_{i=1}^N$ conditional on $\{X_t\}_{t=-\infty}^{\infty}$, as in pure fixed effects approaches. For a discussion of all other components, we refer the reader to Section 3.2. Again, the crucial assumption is 3.2(i) on the availability of an observed macro shock satisfying certain orthogonality conditions. We consider alternatives to it in the form of mismeasurement with an instrument in Section 3.3.4.

Estimator and inference procedure

We now introduce the panel LP estimator and inference procedure. We denote by $W_{it} \in \mathbb{R}^d$ the vector of controls (d may change with the sample size). If W_{it} contains no time fixed effects, let $\hat{s}_i = s_i$ — this accommodates the case $s_i = 1$. Otherwise, let $\hat{s}_i = s_i - N^{-1}\sum_{j=1}^N s_j$ and note that if time fixed effects are included, local projections on $s_i X_t$ and $\hat{s}_i X_t$ produce numerically the same estimate $\hat{\beta}(h)$ below. In addition to unit and possibly time dummies, we consider below cases in which W_{it} contains lags of $s_i X_t$ or Y_{it} and we assume that W_{it} is observed for $t = 1, \ldots, T$.[18]

The fitted equation for the panel LP estimator $\hat{\beta}(h)$ is

$$Y_{i,t+h} = \hat{\beta}(h)\hat{s}_i X_t + \hat{\eta}(h)' W_{it} + \hat{\xi}_{it}(h),$$

where the residual $\hat{\xi}_{it}(h)$ is orthogonal to $\hat{s}_i X_t$ and W_{it}. To characterize $\hat{\beta}(h)$ we use Frisch–Waugh–Lovell. Consider the auxiliary regression of $\hat{s}_i X_t$ on W_{it},

$$\hat{s}_i X_t = \hat{\pi}(h)' W_{it} + \hat{x}_{it}(h), \tag{3.8}$$

where the residual $\hat{x}_{it}(h)$ is orthogonal to W_{it}. Then, an explicit formula for $\hat{\beta}(h)$ is

$$\hat{\beta}(h) = \frac{\sum_{t=1}^{T-h}\sum_{i=1}^N \hat{x}_{it}(h) Y_{i,t+h}}{\sum_{t=1}^{T-h}\sum_{i=1}^N \hat{x}_{it}(h)^2}. \tag{3.9}$$

The time-clustered heteroskedasticity-robust standard error is

$$\hat{\sigma}(h) = \sqrt{\frac{\hat{V}(h)}{(T-h)\hat{J}(h)^2}}, \tag{3.10}$$

[18] Since Y_{it}, X_t and s_i could be multivariate, this is without loss of generality. For example, a panel LP of Y_{it} on $s_i X_t$ controlling for X_t and lags of Y_{it} and another micro control $\tilde{Y}_{it}$ is covered by redefining Y_{it} to $(Y_{it}, \tilde{Y}_{it})$ and s_i to $(1, s_i)$. Also, note that if W_{it} includes lags of shocks or outcomes we assume we observe $s_i X_t$ or Y_{it} for $t < 1$.

with

$$\hat{J}(h) = \frac{1}{N(T-h)} \sum_{t=1}^{T-h} \sum_{i=1}^{N} \hat{x}_{it}(h)^2, \quad \hat{V}(h) = \frac{1}{(T-h)} \sum_{t=1}^{T-h} \left(\frac{1}{N} \sum_{i=1}^{N} \hat{x}_{it}(h) \hat{\xi}_{it}(h) \right)^2. \tag{3.11}$$

Finally, the $(1-\alpha)$ confidence interval is

$$\hat{C}_\alpha(h) = \left[\hat{\beta}(h) \pm z_{1-\alpha/2} \hat{\sigma}(h) \right], \tag{3.12}$$

where z_q is the q-quantile of the standard normal distribution.

Additional assumptions

To establish our uniform asymptotic approximations, we need the following:

Assumption 3.3 (Regularity conditions).

(i) There is a positive finite constant M_8 such that, almost surely,

$$E\left[X_t^8 \middle| \{\theta_i, s_i\}_{i=1}^N\right] \le M_8, \quad E\left[Z_t^8 \middle| \{\theta_i, s_i\}_{i=1}^N\right] \le M_8, \quad E\left[u_{it}^8 \middle| \theta_i, s_i\right] \le M_8.$$

(ii) There is a positive finite constant $\underline{M}$ such that, almost surely,

$$E\left[X_t^2 \middle| \{X_\tau\}_{\tau \neq t}, \{\theta_i, s_i\}_{i=1}^N\right] \ge \underline{M}, \quad E\left[Z_t^2 \middle| \{X_\tau\}, \{\theta_i, s_i\}_{i=1}^N\right] \ge \underline{M}, \quad E\left[u_{it}^2 \middle| \{X_\tau\}, \theta_i, s_i\right] \ge \underline{M}.$$

(iii) The conditional cumulants up to fourth-order of $\operatorname{vec}\left\{(X_t, Z_t, u_{it})(X_t, Z_t, u_{it})'\right\}$ *given $\{\theta_i, s_i\}_{i=1}^N$ are almost surely absolutely summable.*

(iv) There are positive finite constants C_ℓ such that $C = \sum_{\ell=0}^\infty C_\ell < \infty$ and, almost surely,

$$|\beta_{i\ell}| \le C_\ell, \quad |\gamma_{i\ell}| \le C_\ell, \quad |\delta_{i\ell}| \le C_\ell, \quad |s_i| < C.$$

(v) There is a positive finite constant $\underline{C}$ such that, almost surely,

$$\sum_{\ell=0}^\infty \left(N^{-1} \sum_{i=1}^N \hat{s}_i \beta_{i\ell} \right)^2 \ge \underline{C}, \quad \sum_{\ell=0}^\infty \left(N^{-1} \sum_{i=1}^N \hat{s}_i \gamma_{i\ell} \right)^2 \ge \underline{C}, \quad N^{-1} \sum_{\ell=0}^\infty \sum_{i=1}^N \hat{s}_i^2 \delta_{i\ell}^2 \ge \underline{C}.$$

Our model interprets θ_i as unit-specific parameters and $\{X_t, Z_t, u_{it}\}$ as sources of uncertainty. This calls for making time series assumptions on the uncertainty given parameters (parts (i), (ii) and (iii)) while requiring that parameters ensure sufficient regularity for all units in the cross-sectional population (parts (iv) and (v)).

Parts (i), (ii) and (iii) are standard in the time series context (see, for instance, Assumption 2 in Montiel Olea and Plagborg-Møller (2021)). They put limits on the tails of the distributions of shocks, as well as the predictability and dependence of their second moments. Part (iv), on the other hand, guarantees that infinite moving averages, such as $\sum_{\ell=0}^{\infty}\beta_{i\ell}X_{t-\ell}$, are well defined for all units. Absolute summability rules out unit roots but still allows for rich persistence patterns — such as those from stationary ARMA and other short-memory processes.[19]

Lastly, part (v) requires non-zero variability given $\{\theta_i, s_i\}_{i=1}^{N}$ of $N^{-1}\sum_{i=1}^{N}\hat{s}_i\sum_{\ell=0}^{\infty}\beta_{i\ell}X_{t-\ell}$, $N^{-1}\sum_{i=1}^{N}\hat{s}_i\sum_{\ell=0}^{\infty}\gamma_{i\ell}Z_{t-\ell}$ and $N^{-1/2}\sum_{i=1}^{N}\hat{s}_i\sum_{\ell=0}^{\infty}\delta_{i\ell}u_{i,t-\ell}$. It is mostly a technical condition to prevent trivial cases in which the regression score has zero variance. Nevertheless, it is compatible with, say, a non-negligible fraction of units having zero exposure to macro or micro shocks. It also places no restriction on the relative importance of macro versus micro shocks which is governed by κ.

3.3.2 Main result

The main contribution of the chapter is to characterize the large-sample properties of $\hat{\beta}(h)$, $\hat{\sigma}(h)$ and $\hat{C}_\alpha(h)$. In the asymptotic plan, we take $T, N \to \infty$ and we are interested in uniform approximations with respect to κ. The key result is Proposition 3.1 which states that $\hat{C}_\alpha(h)$ delivers uniformly valid inference for the coefficient in a regression of β_{ih} on $\hat{s}_i$ if enough lags of $\hat{s}_i X_t$ are used as controls.

We describe first the estimand and then the uniform inference result. We use P_κ to indicate probabilities under a DGP associated to a given value of κ and we omit the subindex from objects whose probabilities (or expectations) do not depend on κ (such as those in Assumptions 3.2 and 3.3).

Estimand. If s_i is not a constant and time fixed effects are included, the population object targeted by the panel LP is

$$\beta(h) = \frac{\operatorname{Cov}\left(s_i, \beta_{ih}\right)}{\operatorname{Var}\left(s_i\right)}. \tag{3.13}$$

In other words, panel LPs estimate the slope in a population linear projection of β_{ih} on characteristics s_i including an intercept. Similarly, if $s_i = 1$, the estimand becomes the mean impulse response $\beta(h) = E\left[\beta_{ih}\right]$. Note that omitting either X_t or time dummies as controls in a panel LP has the effect of forcing the regression of β_{ih} on s_i through the origin, leading to the estimand $\beta(h) = (E\left[s_i^2\right])^{-1}E\left[s_i\beta_{ih}\right]$. In order to obtain a rich summary of the heterogeneity in β_{ih}, therefore, the researcher will typically need to explore different choices of s_i or allow s_i to be a vector.[20]

Under the conditions of Proposition 3.1, $\hat{\beta}(h) = \beta(h) + o_{P_\kappa}(1)$ for any DGP sequence P_κ such that $\kappa/\sqrt{TN} = o(1)$: that is, if the panel LP estimator converges, it is to $\beta(h)$.

This clarifies the sense in which panel LPs can be interpreted when the underlying population of interest features unrestricted heterogeneity in responses to shocks, as in (3.6). Precisely because we place virtually no restriction on the joint distribution of (θ_i, s_i), the characterization of the estimand is of a nonparametric nature.

[19]We conjecture, however, that many of our results remain valid at moderate horizons in the presence of near unit roots and our simulation evidence supports this claim. See Section 3.3.3 for further discussion.

[20]For example, the best linear approximation $E^*\left[\beta_{ih}|s_i\right] = E\left[\beta_{ih}\right] + \left(\operatorname{Cov}\left(s_i, \beta_{ih}\right)/\operatorname{Var}\left(s_i\right)\right)\left(s_i - E\left[s_i\right]\right)$ requires both estimands or, alternatively, the interaction of X_t with $(1, s_i)$ rather than s_i alone (omitting time effects). If s_i is multivariate, a confidence region constructed on the basis of a time-clustered heteroskedasticity-robust variance estimate enjoys the same uniform validity property of Proposition 3.1. We illustrate this in our empirical calculations in Section 3.5.

Uniformly valid inference. Let p be the number of lags of $\hat{s}_i X_t$ included in the controls W_{it}. Both p and h are fixed as $T, N \to \infty$ while $T/N \to 0$.[21] Our main result is that $\hat{C}_\alpha(h)$ has correct coverage for $\beta(h)$ uniformly over κ so long as $h \leq p$:

Proposition 3.1. *Under Assumptions 3.1, 3.2 and 3.3, for $h \leq p$,*

$$\lim_{T,N\to\infty} \sup_{\kappa} \left| P_\kappa \left(\beta(h) \in \hat{C}_\alpha(h) \right) - (1-\alpha) \right| = 0. \tag{3.14}$$

Proof. See Section A.1. □

Proposition 3.1 states that valid inference results from clustering standard errors at the time level, which accounts for cross-sectional dependence induced by omitted aggregate shocks, and from ex-ante including lags of $\hat{s}_i X_t$ as controls, which renders the regression scores uncorrelated. We refer to this strategy as time-clustered lag-augmented heteroskedasticity-robust (t-LAHR) inference. As in Section 3.2 and as explained below, it is closely linked to inference in time series LPs.

Despite the general error dynamics in (3.6)–(3.7), the regression score $\sum_{i=1}^N X_t \hat{s}_i \xi_{it}(h,\kappa)$, with $\xi_{it}(h,\kappa)$ the population counterpart to $\hat{\xi}_{it}(h)$ defined in (3.19), has limited serial correlation. It is an MA(h) process with the first p autocovariances set to zero. Thus, it becomes uncorrelated when $p \geq h$ which is why t-LAHR works. Besides, when $p < h$, the autocovariances stem only from leftover leads of X_t and not from the unobserved macro error Z_t or micro error u_{it}. In fact, they will tend to be small compared to the variance of the score in low-signal (large κ) DGPs or if $\beta_{i\ell}$ decays quickly. We therefore expect t-LAHR inference to have small coverage distortions even for $p < h$; we provide affirmative evidence via simulations in Section 3.4.

A striking implication of Proposition 3.1 is that t-LAHR inference remains valid even in the low-signal setting $\kappa/\sqrt{N} \to \infty$ where there is scarcity of information about aggregate shocks in the sample and $\hat{\beta}(h)$ is inconsistent. The uniformity over DGPs with different macro-micro signal-noise obviates the need to take a stand on the κ-regime, which is important because κ is not always consistently estimable.

In contrast, inference based on unit-level clustering of the regression score is not uniformly valid as it tends to severely undercover $\beta(h)$ in high- and moderate-signal regimes. Similarly to Section 3.2, provided lags of $\hat{s}_i X_t$ are included, unit-level clustering is asymptotically equivalent to not clustering at all, whereas two-way clustering is equivalent to time-level clustering. That is, unit-level clustering is neither necessary nor sufficient for valid inference — yet another implication of X_t being a shock that has no counterpart in a more generic time series setup.

Remark 3.3 (Proof steps). To establish (3.14), we decompose the problem into showing (A) asymptotic normality of the score, (B) consistency of the standard error, and (C) negligibility of some remainder terms. We obtain uniformity via the drifting parameter sequence approach (see Andrews, Cheng, and Guggenberger (2020)).

In (A), although the regression score is serially uncorrelated, it contains leads and lags of macro and micro errors. This makes the reverse martingale technique of Montiel Olea and Plagborg-Møller (2021)

[21] We regard $T/N \to 0$ as a mild requirement for the empirical applications of reference. It follows from the proof of Proposition 3.1 that if T/N is not asymptotically negligible (as if taking N as fixed), (3.14) holds with $\beta(h)$ replaced by the *finite-population* estimand $\tilde{\beta}(h) = (\sum_{i=1}^N \hat{s}_i^2)^{-1} \sum_{i=1}^N \hat{s}_i \beta_{ih}$.

inapplicable. Instead, using a similar insight to that of Xu (2023), we produce a martingale approximation by rearranging the score so that the leads at time t become the lags at a time in the future of t. See Lemma B.1 in Appendix B.1 for the details.

In (B) and (C), we rely on direct calculation of uniform bounds. The presence of heterogeneity poses a challenge with no parallel in the time series case. Because of Assumption 3.3, we can derive many of the bounds by first conditioning on $\{\theta_i, s_i\}_{i=1}^N$, exploiting the connection between conditional and unconditional convergence.

Remark 3.4 (Synthetic time series). A useful device to interpret panel LPs is the following representation. The residual $\hat{x}_{it}(h)$ in (3.8) can be written as $\hat{x}_{it}(h) = \hat{s}_i\hat{X}_t(h)$, where $\hat{X}_t(h)$ is the residual from regressing X_t on $X_{t-1}, \dots, X_{t-p}$ and an intercept (on $T - h$ observations).[22]

Then, the panel LP estimator in (3.9) can be written as

$$\hat{\beta}(h) = \frac{\sum_{t=1}^{T-h}\sum_{i=1}^N \hat{s}_i\hat{X}_t(h)Y_{i,t+h}}{\sum_{t=1}^{T-h}\sum_{i=1}^N \hat{s}_i\hat{X}_t(h)^2} = \frac{\sum_{t=1}^{T-h}\hat{X}_t(h)\hat{Y}_{t+h}}{\sum_{t=1}^{T-h}\hat{X}_t(h)^2},$$

i.e., the time series LP estimator that regresses cross-sectional regression 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 $X_{t-1}, \dots, X_{t-p}$ and an intercept. The standard error $\hat{\sigma}(h)$ in (3.10) 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, t-LAHR inference for panel LPs and lag-augmented heteroskedasticity-robust inference for time series LPs are intimately related.

Remark 3.5 (s_i and precision). This representation is also useful to illuminate the fact that estimation error is of order $T^{-1/2}$ in environments with $\kappa \propto \sqrt{N}$, despite what otherwise looks like a standard panel regression with potentially very rich micro data. We can give interpretable conditions under which variation in s_i affords faster convergence rates. These are akin to s_i being a cross-sectional instrument: we require s_i to correlate with β_{ih} — that is, be relevant for heterogeneity in transmission of X_t at horizon h — but to be orthogonal to all other exposures to aggregate shocks, $(\{\beta_{i\ell}\}_{\ell\neq h}, \gamma_i)$. These conditions seem particularly hard to meet: for each horizon h, a source of variation that is orthogonal to responses at all other horizons is required. (Assumption 3.3(v) rules this out in our formulation.) In some sense, this reveals an intrinsic trade-off between documenting interesting transmission mechanisms and finding valid instruments for precision.

Remark 3.6 (t-HAR). In principle, time-clustered HAR inference is a valid alternative to t-LAHR. An analogue to Proposition 3.1 can be established for a confidence interval that replaces $\hat{V}(h)$ in (3.11) with the Hansen and Hodrick (1980) variance estimator $\hat{V}(h) + 2\sum_{\ell=p+1}^{h}\tilde{V}_\ell(h)$ where

$$\tilde{V}_\ell(h) = \frac{1}{(T-h)}\sum_{t=\ell+1}^{T-h}\left(\frac{1}{N}\sum_{i=1}^N \hat{x}_{it}(h)\hat{\xi}_{it}(h)\right)\left(\frac{1}{N}\sum_{i=1}^N \hat{x}_{i,t-\ell}(h)\hat{\xi}_{i,t-\ell}(h)\right),$$

This boils down to $\hat{V}(h)$ for $p \geq h$. Unlike $\hat{V}(h)$, this alternative variance estimator is not guaranteed to

[22] To see this, note that $\hat{x}_{it}(h)$ is $\hat{s}_iX_t$ minus a linear combination of $\hat{s}_iX_{t-1}, \dots, \hat{s}_iX_{t-p}$ and unit and possibly time indicators which is orthogonal to all of the latter. When W_{it} includes additional controls, the synthetic time series representation is asymptotically but not numerically equivalent.

be positive semidefinite. Also, t-LAHR inference is simpler to implement and refine, remains tractable over moderate horizons under VAR DGPs (Section 3.3.3), and performs better in small samples (Section 3.4).

Remark 3.7 (State-dependence). In some applications, interest is in the differential pass-through of shocks to responses along an observable (time-varying) state, denoted now s_{it}. Formalizing this requires extending (3.6)–(3.7) to allow for time-varying impulse responses:

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

Letting $\hat{s}_{it} = s_{it} - N^{-1} \sum_{j=1}^{N} s_{jt}$, the corresponding panel LP estimator on $\hat{s}_{it} X_t$ retains its interpretation as the slope coefficient of the linear projection $E^* \left[\beta_{ith} \middle| s_{it}\right]$ as long as s_{it} and impulse responses are exogenous with respect to X_t. Although a more detailed exploration is beyond the scope of our chapter, the treatment of s_{it} is analogous to that of s_i, and all the results above carry over with little modification. We revisit this in simulations in Section 3.4 and in our empirical illustration in Section 3.5.[23]

3.3.3 Panel VAR model

It is not uncommon in applications that the researcher is interested in responses at an horizon h which is a non-negligible fraction of T. Proposition 3.1 guarantees exact coverage for short horizons depending on the number of lags of the outcome and shock used as controls. There is, however, one important class of DGPs for which our uniformity result extends to $h \propto T$: the VAR class.

We now assume a panel VAR(p) model (with $p < \infty$):

$$Y_{it} = m_i + \sum_{\ell=1}^{p} A_\ell Y_{i,t-\ell} + \sum_{\ell=0}^{p} B_{i\ell} X_{t-\ell} + C_{i0} Z_t + \kappa D_{i0} u_{it}. \tag{3.15}$$

If $\sum_{\ell=1}^{p} A_\ell < 1$, as implied by Assumption 3.3(iv), we can recover the unit-specific parameters μ_i, $\{\beta_{i\ell}\}$, $\{\gamma_{i\ell}\}$, $\{\delta_{i\ell}\}$ from m_i, $\{A_\ell\}$, $\{B_{i\ell}\}$, C_{i0}, D_{i0} by inverting the lag polynomial $A(L) = 1 - \sum_{\ell=1}^{p} A_\ell L^\ell$. That is, VAR model (3.15) is a special case of (3.6)–(3.7).

Assuming that p is known and that W_{it} contains p lags of Y_{it} and $s_i X_t$, the t-LAHR confidence interval $\hat{C}_\alpha(h)$ defined in (3.12) has uniform validity even for moderately long horizons h exceeding p:

Proposition 3.2. *Under Assumptions 3.1, 3.2 and 3.3, for some positive constant $\phi < 1$,*

$$\lim_{T,N\to\infty} \sup_{0\le h\le \phi T} \sup_{\kappa} \left| P_\kappa \left(\beta(h) \in \hat{C}_\alpha(h) \right) - (1-\alpha) \right| = 0. \tag{3.16}$$

Proof. See Section A.1. □

The intuition and proof for Proposition 3.2 mirror that of Proposition 3.1. Under VAR model (3.15) the regression score $\sum_{i=1}^{N} X_t \hat{s}_i \xi_{it}(h,\kappa)$, with $\xi_{it}(h,\kappa)$ now defined in (3.20), is serially uncorrelated not just for

[23] Rambachan and Shephard (2021, Section 3.4) offer a nonparametric characterization of local projection estimands when states are endogenous in a time-series potential outcomes framework; see also Gonçalves, Herrera, Kilian, and Pesavento (forthcoming) for the case where $s_t = \mathbb{1}\{X_t > c\}$.

$h \leq p$ but for any h. The basic consequence is that if a low-order VAR model is a reasonable approximation, the t-LAHR inference approach that relies on controlling for a small number of lags of the outcome and shock is robust over long horizons and regardless of the amount of micro noise.[24]

Remark 3.8 (LP inference when the shock is not observable). Proposition 3.2 can be read as the panel data counterpart to the result in Montiel Olea and Plagborg-Møller (2021) under stationarity when the shock is directly observable. That parallel implies that if X_t is unavailable but instead we observe $X_t^* = \sum_{\ell=1}^{p-1} \alpha_\ell X_{t-\ell}^* + X_t$ and we run a local projection of $Y_{i,t+h}$ on $s_i X_t^*$ including p lags of Y_{it} and $s_i X_t^*$ in the control vector W_{it}, t-LAHR inference is uniformly valid over h and κ.[25]

Remark 3.9 (Heterogeneity in VAR coefficients). Model (3.15) assumes homogeneous coefficients $\{A_\ell\}$. This is common in the microeconometric literature on panel VARs (Arellano, 2003, Chapter 6) but it is not necessary for (3.16). For example, we can establish Proposition 3.2 in a moderate heterogeneity environment that replaces A_ℓ with $A_{i\ell}$ where $\sup_{1\leq i \leq N} |A_{i\ell} - A_\ell| = O_p\left(T^{-1/2}\right)$. Proposition 3.2 can also be established (under slightly different regularity conditions) if we allow for heterogeneity in $\{A_\ell\}$ but we include p *unit-specific* lags of Y_{it} as controls in W_{it}.

3.3.4 Panel LP-IV and proxy shocks

The most common implementation of panel LPs in empirical work treats the shock of interest as observed. Nevertheless, it is sometimes more realistic to assume there is measurement error in the shock elicitation process. This creates an endogeneity problem that can be dealt with by using the shock measures as instruments for the actual underlying shock (Ramey, 2016; Stock and Watson, 2018).

The researcher observes the outcome Y_{it} and characteristics s_i, but instead of the actual shock X_t she observes an endogenous aggregate state variable $\tilde{X}_t$ and a proxy shock X_t^*. In addition to (3.6)–(3.7), we assume

$$\tilde{X}_t = \sum_{\ell=0}^{\infty} b_\ell X_{t-\ell} + \sum_{\ell=0}^{\infty} c_\ell Z_{t-\ell}, \tag{3.17}$$

$$X_t^* = a_0 X_t + \nu_t, \tag{3.18}$$

where ν_t is measurement error. We normalize $b_0 = 1$ to fix the scale of the estimand as only relative impulse responses are identified.[26] We also adopt the following:

Assumption 3.4 (LP-IV).

[24] The results in Montiel Olea et al. (2024) suggest that for a *fixed* horizon h, t-LAHR inference would also remain valid if the VAR model (3.15) were contaminated by moving averages of Z_t and u_{it} in a $T^{-1/4}$-neighborhood of zero — that is, if the VAR model holds only approximately. The simulation evidence in Section 3.4 based on DGPs which are not VARs is consistent with this idea.

[25] Lag-augmentation means including at least one more lag than the autoregressive order of X_t^* which is $p-1$. The connection with Montiel Olea and Plagborg-Møller (2021) also suggests that $\hat{C}_\alpha(h)$ is uniformly valid over the VAR parameter space (including unit roots) if a certain condition on uniform non-singularity of the least squares denominator matrix (Assumption 3 in their paper) holds.

[26] It is straightforward to include intercepts in both (3.17) and (3.18). Additionally, as in Section 3.3.3, we can derive uniformity results with respect to the horizon h by assuming a VAR model in (3.6), (3.7) and (3.17).

(i) $a_0 \neq 0$.

(ii) Assumptions 3.1, 3.2 and 3.3 hold with Z_t *replaced by* (Z_t, ν_t).

(iii) For the same constants C_ℓ *and* $\underline{C}$ *of Assumption 3.3,*

$$|b_\ell| \leq C_\ell, \quad |c_\ell| \leq C_\ell, \quad \sum_{\ell=0}^{\infty} b_\ell^2 \geq \underline{C}, \quad \sum_{\ell=0}^{\infty} c_\ell^2 \geq \underline{C}.$$

Assumption 3.4(i) is needed for instrument relevance, and we restrict our attention to the strong instrument case where we keep a_0 fixed as $N, T \to \infty$. On the other hand, Assumption 3.4(ii) implies that ν_t is orthogonal to $\{X_\tau, Z_\tau\}$. This embodies the key lead-lag exogeneity condition requiring X_t^* to be contemporaneously correlated only with X_t, a well-known condition in the time series LP-IV context.[27] Finally, Assumption 3.4(iii) imposes regularity on the endogenous variable $\tilde{X}_t$.

LP-IV estimation and inference. LP-IV regresses $Y_{i,t+h}$ on $\tilde{\boldsymbol{X}}_t = (\tilde{X}_t, \tilde{X}_{t-1}, \ldots, \tilde{X}_{t-p})'$ using $\boldsymbol{X}_t^* = (X_t^*, X_{t-1}^*, \ldots, X_{t-p}^*)'$ as instruments (both interacted with s_i), controlling for unit and time effects (W_{it} denotes controls). The residualized instrument is

$$\hat{\boldsymbol{x}}_{it}(h) = \hat{s}_i \boldsymbol{X}_t^* - \hat{\boldsymbol{\pi}}(h)' W_{it} = \hat{s}_i \hat{\boldsymbol{X}}_t^*(h),$$

where $\hat{\boldsymbol{X}}_{\boldsymbol{t}}^{\boldsymbol{*}}(h) = \boldsymbol{X}_t^* - (T-h)^{-1} \sum_{t=1}^{T-h} \boldsymbol{X}_t^*$. The panel LP-IV estimator $\hat{\boldsymbol{\beta}}^{\text{IV}}(h)$ is then

$$\hat{\boldsymbol{\beta}}^{\text{IV}}(h) = \left(\sum_{t=1}^{T-h} \sum_{i=1}^{N} \hat{\boldsymbol{x}}_{it}(h) \hat{s}_i \tilde{\boldsymbol{X}}_t' \right)^{-1} \sum_{t=1}^{T-h} \sum_{i=1}^{N} \hat{\boldsymbol{x}}_{it}(h) Y_{i,t+h} = \left(\sum_{t=1}^{T-h} \hat{\boldsymbol{X}}_t^*(h) \tilde{\boldsymbol{X}}_t' \right)^{-1} \sum_{t=1}^{T-h} \hat{\boldsymbol{X}}_t^*(h) \hat{Y}_{i,t+h},$$

where $\hat{Y}_{i,t+h}$ is the synthetic outcome defined in Remark 3.4. Put another way, panel LP-IV admits a synthetic time series LP-IV representation.

The only entry of $\hat{\boldsymbol{\beta}}^{\text{IV}}(h)$ that has interpretation as an estimate of a relative impulse response is $\hat{\beta}_0^{\text{IV}}(h) = e_1' \hat{\boldsymbol{\beta}}^{\text{IV}}(h)$ where e_1 is the first column of I_{p+1}. The remaining entries are necessary for t-LAHR inference to be valid. Given residuals

$$\hat{\xi}_{it}^{\text{IV}}(h) = Y_{i,t+h} - \hat{s}_i \tilde{\boldsymbol{X}}_t' \hat{\boldsymbol{\beta}}^{\text{IV}}(h) - \hat{\eta}^{\text{IV}}(h)' W_{it},$$

we define

$$\hat{\boldsymbol{J}}^{\text{IV}}(h) = \frac{1}{N(T-h)} \sum_{t=1}^{T-h} \sum_{i=1}^{N} \hat{\boldsymbol{x}}_{it}(h) \hat{s}_i \tilde{\boldsymbol{X}}_t', \quad \hat{\boldsymbol{V}}^{\text{IV}}(h) = \frac{1}{(T-h)} \sum_{t=1}^{T-h} \left(\frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{x}}_{it}(h) \hat{\xi}_{it}^{\text{IV}}(h) \right)^2.$$

The time-clustered heteroskedasticity-robust standard error for $\hat{\beta}_0^{\text{IV}}(h)$ is

$$\hat{\sigma}_0^{\text{IV}}(h) = \left[\frac{1}{(T-h)} \cdot \left(e_1' \hat{\boldsymbol{J}}^{\text{IV}}(h)^{-1} \right) \hat{\boldsymbol{V}}^{\text{IV}}(h) \left(e_1' \hat{\boldsymbol{J}}^{\text{IV}}(h)^{-1} \right)' \right]^{1/2}$$

[27] See, for instance, Stock and Watson (2018, p. 924) and Plagborg-Møller and Wolf (2021, p. 970). The setup can be extended to allow ν_t to be serially correlated and to the case where X_t^* is valid only after conditioning on a set of controls.

and the $(1-\alpha)$ confidence interval, $\hat{C}^{\mathrm{IV}}_{\alpha}(h) = \left[\hat{\beta}^{\mathrm{IV}}_0(h) \pm z_{1-\alpha/2}\hat{\sigma}^{\mathrm{IV}}_0(h)\right]$. Then:

Proposition 3.3. *Under Assumption 3.4, for $h \leq p$,*

$$\lim_{T,N\to\infty} \sup_{\kappa} \left| P_{\kappa}\left(\beta(h) \in \hat{C}^{IV}_{\alpha}(h)\right) - (1-\alpha) \right| = 0.$$

Proof. See Section A.1. □

Remark 3.10 (Absence of first-stage heterogeneity). The LP-IV estimand coincides (under the normalization $b_0 = 1$) with the LP estimand (3.13) despite the presence of heterogeneity. This is far from obvious: under treatment effect heterogeneity, IV estimands are generally (weighted averages of) local average treatment effects (Angrist and Imbens, 1995; Angrist, Imbens, and Graddy, 2000). It is the aggregate-only nature of the first-stage model that underlies this result. This is yet another illustration of the unique setting that we study in this chapter.

3.4 Simulation study

We ran a comprehensive simulation study to verify the finite-sample robustness of the inference procedures analyzed in Section 3.3. Here we provide a summary and defer additional detail and results to Appendix B.2.

Designs. Our study relies on two different DGPs. The first is the general setup (3.6)–(3.7) supplemented with (3.17)–(3.18) 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 B.2. We calibrate the distribution of $\{\beta_{i\ell}, \gamma_{i\ell}, \delta_{i\ell}\}$ and the value of $\{b_\ell, c_\ell\}$ 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 3.7.

The second DGP is the VAR model (3.15). Again we generate shocks as i.i.d. $N(0,1)$ and we simulate the heterogeneity as detailed in Appendix B.2. 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.[28]

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 $N = 1,000$ (although we also considered experiments with larger N) and we let κ take values consistent with $\bar{R}^2(\kappa) \in \{0.99, 0.66, 0.33\}$ as defined in (3.4). As a reference, $\bar{R}^2(\kappa) = 0.66$ corresponds to the one-third of aggregate fluctuations explained by micro shocks suggested by Gabaix (2011) for GDP growth, which we take as moderate signal-to-noise.

[28] We also considered experiments where (a) in the first DGP shocks are conditionally heteroskedastic, and (b) in the VAR DGP we have unit-specific VAR parameters $\{A_{i\ell}\}$. We did not find any major difference with what we report here.

Inference procedures. We compare t-LAHR inference with one-way (1W), two-way (2W), and Driscoll-Kraay (DK98) inferences. These are implemented without lag augmentation, as is common practice. For illustrative purposes, we also include t-HR (the non-lag-augmented counterpart to t-LAHR) and t-HAR alternatives.

For t-LAHR inference we use the simple lag selection rule $p = \min\{h, (T-h)^{1/3}\}$ (except in the VAR DGP where p is known) and we apply the finite-sample refinement advocated by Imbens and Kolesár (2016). The lag selection rule is motivated by Xu (2023, Section 3.3) for fixed h and provides fairly generous lag augmentation. For t-HAR inference we use the equally-weighted cosine approach (Müller, 2004) with the choice of tuning parameter recommended in Lazarus et al. (2018).

Results. In Figure 3.1, we report pointwise coverage rates for horizons $0 \leq h \leq 0.25T$ with $T = 100$. 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.

Figure 3.1 suggests four takeaways. First, t-LAHR performs best in all scenarios, with coverage close to the nominal rate even in low-signal cases and for horizons h well beyond p. Its mean absolute coverage distortion never exceeds 2%, whereas it is between 4% and 7% for the second best option (t-HAR) under high signal.

Second, estimating the long-run variance of the score (instead of lag augmenting) can be challenging with small T. This is particularly true for DK98 which relies on Newey–West. Interestingly, these approaches do better in low-signal DGPs where, as mentioned before, there is less to gain from doing HAC.

Third, one-way clustering is very sensitive to $\bar{R}^2(\kappa)$, suffering severe distortions in intermediate- and high-$\bar{R}^2(\kappa)$ cases. What is more, it is outperformed by t-LAHR even if micro shocks explain the majority of aggregate variation. This is consistent with the view that 1W guards against the wrong type of correlation in the score.

Finally, two-way clustering 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 alone. This is possibly due to the non-standard behavior of variance estimators when there are micro (near) unit roots.

Identical takeaways emerge in experiments where we substitute s_i with either 1 or s_{it} (Appendix B.2), and with a sample size $T = 30$ (Figure 3.2).

In sum, the small-sample evidence reinforces many of our theoretical results. It shows that the large-sample approximations of Section 3.3 provide reliable guidance for understanding estimation and inference with aggregate shocks. Furthermore, it illustrates the practical relevance of achieving uniformity with respect to κ, and it delivers a clear methodological prescription: t-LAHR inference.

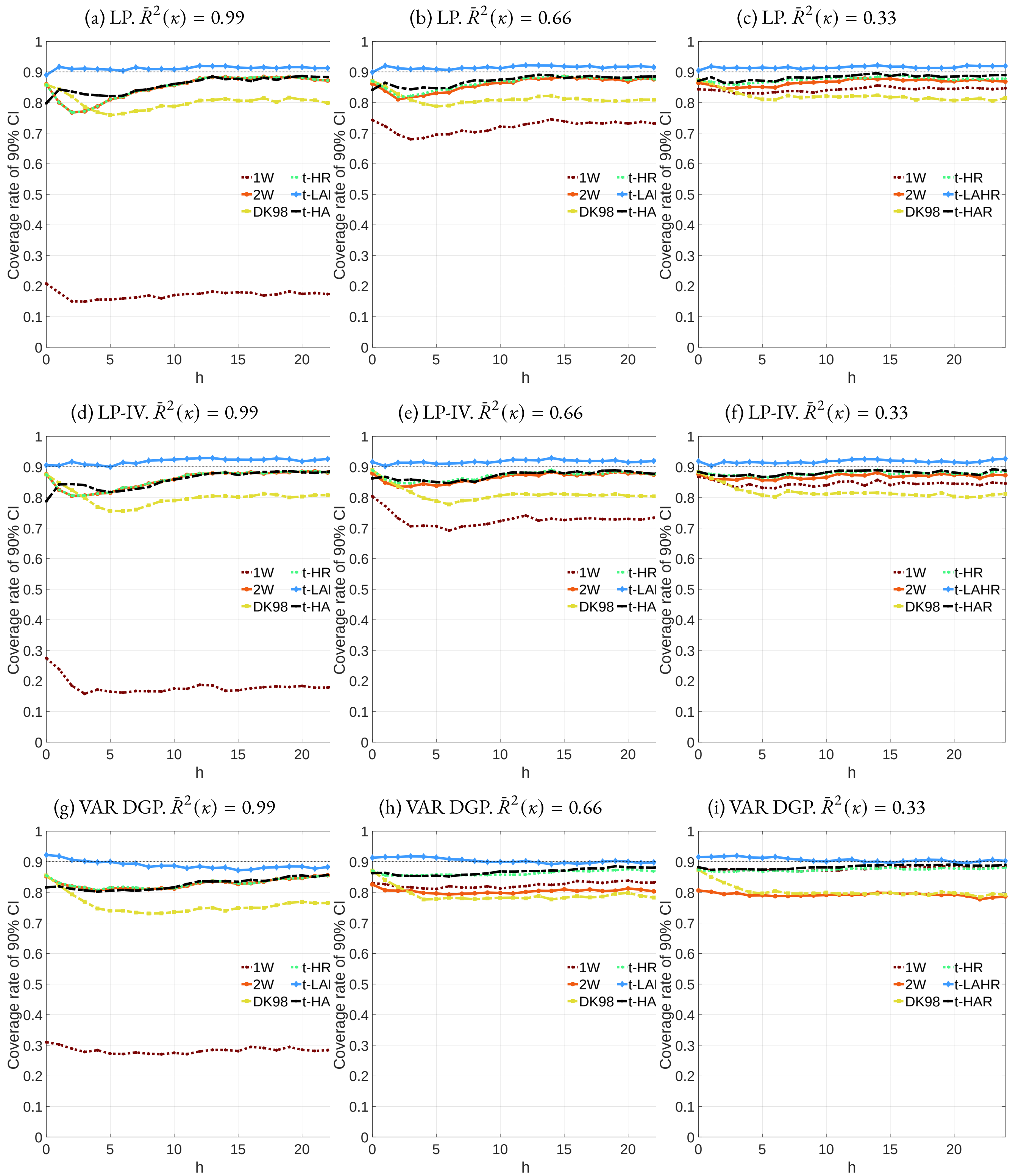


FIGURE 3.1. Coverage rates of 90% confidence intervals for $T = 100$.

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

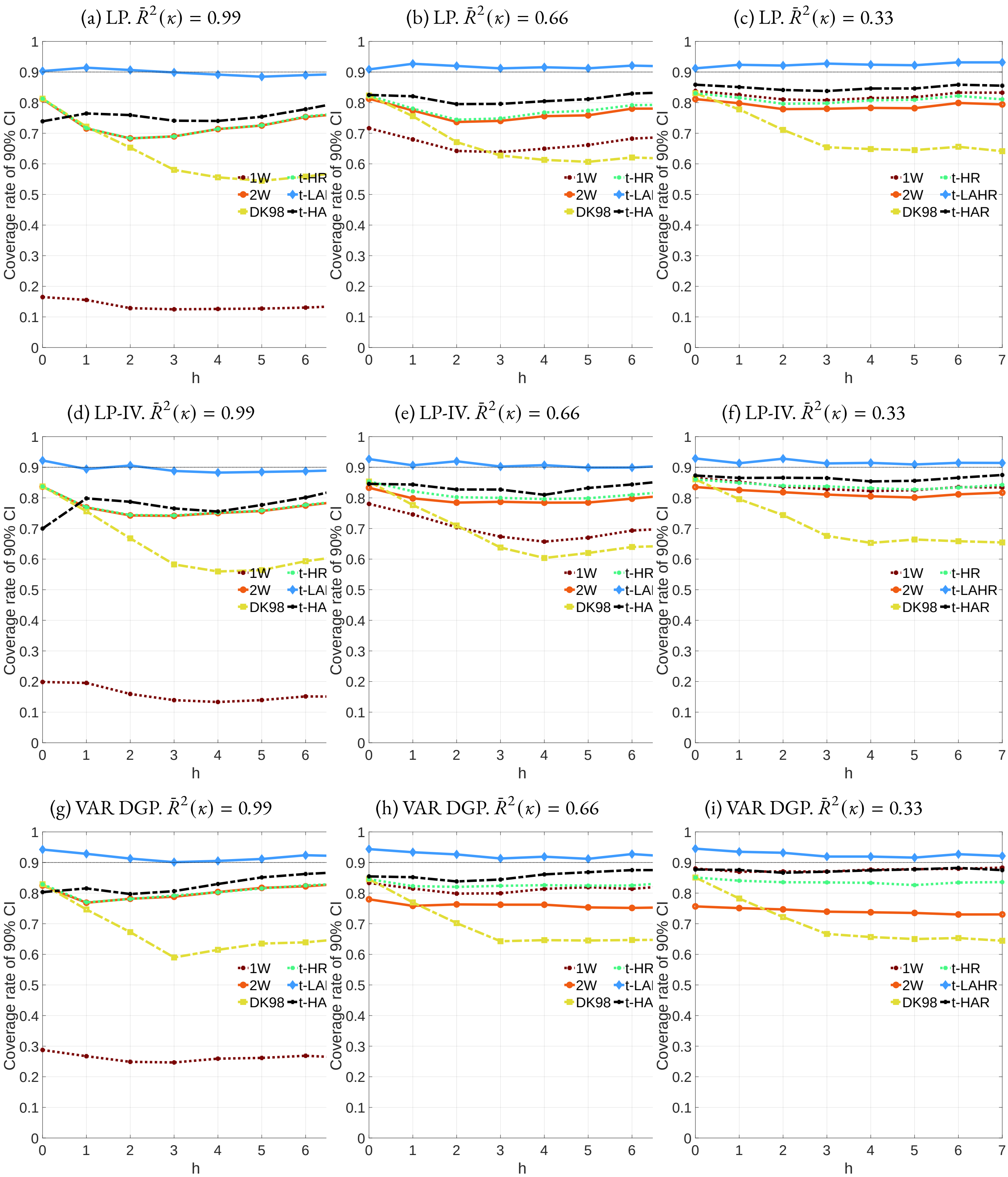


FIGURE 3.2. Coverage rates of 90% confidence intervals for $T = 30$.

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

3.5 Empirical illustration

We now discuss an empirical exercise that demonstrates the applicability of our methods in a setup featuring time-varying s_{it} and unbalanced panels, and compares our practical recommendation to popular alternatives. The exercise is motivated by the burgeoning literature on the role played by firm heterogeneity and financial frictions in the propagation of monetary policy.

Data and background. Quantifying firm-level responses to exogenous changes in policy is a key empirical goal as it is informative on the mechanisms through which monetary policy operates. For instance, Crouzet and Mehrotra (2020) focus on the role of firm size for investment response heterogeneity, finding larger (albeit noisy) responses for smaller firms; Ottonello and Winberry (2020) instead focus on default risk, finding larger responses for less risky companies.

For our empirical analysis, we construct a dataset similar to the latter based on Compustat and high-frequency identified monetary policy shocks (Gürkaynak, Sack, and Swanson, 2005; Gorodnichenko and Weber, 2016). This results in an unbalanced panel for the period 1990Q1–2010Q4 with observations on firm-level investment, size, and leverage.[29] In total, there are $T = 80$ quarters and $N = 4,187$ individual companies which, net of missing data, amount to 235,233 observations.

We consider regressions of cumulative investment changes $Y_{i,t+h} = \log(k_{i,t+h}/k_{i,t-1})$ (k_{it} being the capital stock) on policy shocks X_t interacted with s_{it}, a vector containing size, leverage, and their product. From Section 3.3, we know that under unrestricted heterogeneity the population counterpart is the linear projection of firm-level impulse responses on s_{it}. Thus, including size and leverage together (as well as their interaction) in s_{it} is a way to enrich the linear approximation.

Synthetic time series representation. A fundamental insight of this chapter is that the synthetic time series form of the microdata is a sufficient statistic for the panel LP; a low dimensional representation of a highly complex, unbalanced dataset.[30]

Figure 3.3 displays it for the three components of s_{it}. It is clear that movements in synthetic outcomes concurrent with surprise cuts in policy rates, mostly around recessions, are the main source of identification. There is also substantial variation in synthetic outcomes unrelated to X_t, indicating the presence of omitted aggregate or non-negligible idiosyncratic shocks — the central premises of this chapter.

[29] We use the paper's replication code to build the data and we verify that we can replicate the original results, with minor numerical differences that can be attributed to revisions in input data. Firm size is measured by the value of total assets held by a company while leverage is its debt-to-assets ratio. We have also tried the distance-to-default measure in Ottonello and Winberry (2020) with qualitatively similar results.

[30] Remark 3.4 generalizes as follows. Let $d_{it} = 0$ indicate a missing observation with $d_{it} = 1$ otherwise. Abstracting from controls, the panel local projection estimator with a time-varying s_{it} is

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

where $\omega_t = \sum_{i=1}^{N} d_{it}s_{it}^2$ and $\hat{Y}_{t+h} = (\sum_{i=1}^{N} s_{it}^2)^{-1}\sum_{i=1}^{N} s_{it}Y_{i,t+h}$. 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.

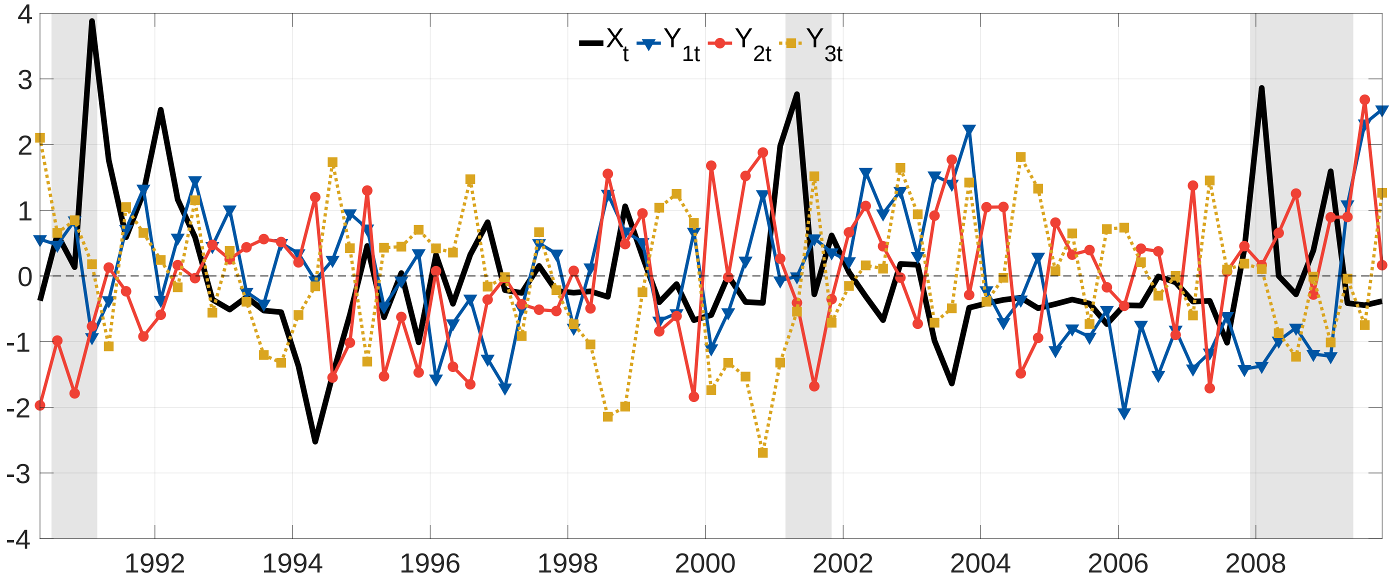


FIGURE 3.3. Synthetic time series representations.

Note: Grey areas are NBER-dated recessions. s_1 is size, s_2 is leverage and s_3 is the interaction. X_t and $\hat{Y}_t$ are standardized to zero mean and unit variance; $X_t > 0$ indicates a surprise cut in the Fed Funds rate.

Estimation and inference method comparison. Figure 3.4 reports point estimates and 90% confidence intervals for the coefficient on each entry of $s_{it}X_t$ at different horizons.[31] According to the t-LAHR intervals, the evidence favors the hypothesis that larger and less indebted firms respond less to monetary policy shocks, with the size effect more persistent and not much interaction between the two.

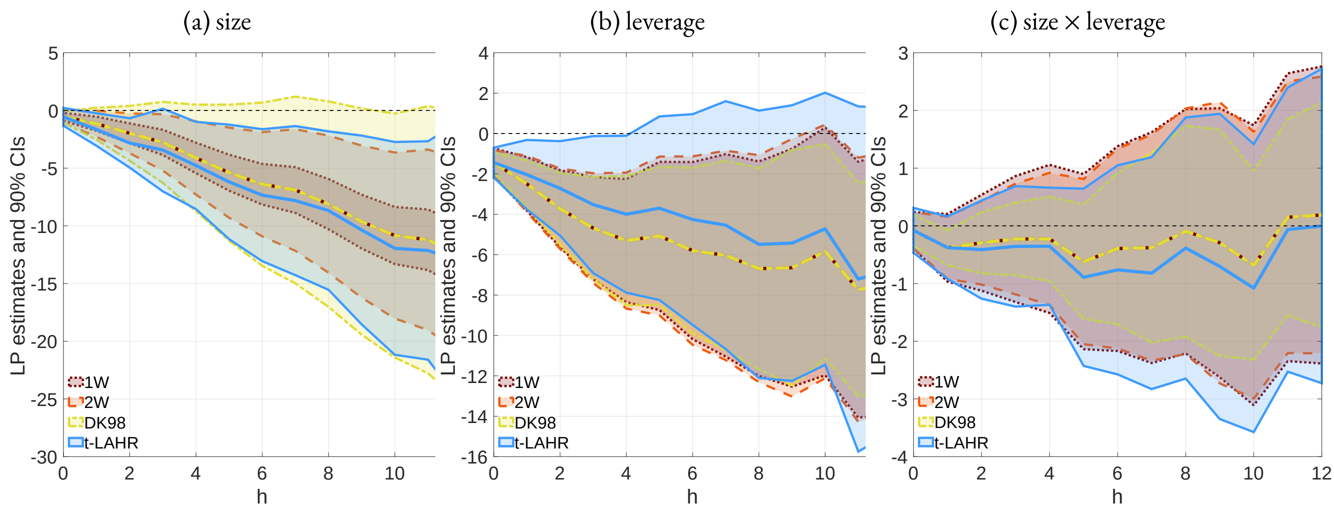


FIGURE 3.4. Point estimates and 90% confidence intervals.

Note: The procedures are one-way unit-level clustering (1W, dotted line), two-way clustering (2W, dashed line), Driscoll-Kraay (DK98, dash dotted line), and t-LAHR (solid line).

From an applied point of view, the main message is that popular methods can deviate significantly from the (asymptotically robust) t-LAHR method. For example, one-way clustering produces intervals that are

[31] The panel local projections include as controls unit and time effects, lagged firm-level sales growth, and both lagged GDP growth and lagged unemployment interacted with s_{it}. For t-LAHR inference we include p lags of Y_{it} and $s_{it}X_t$. We limit $p = \min\{h, 2\}$ to discipline the number of regressors in view of the dimension of s_{it}. One-way, two-way and Driscoll-Kraay are implemented without lag-augmentation.

too short (panel (a)) and too wide (panel (c)).[32] Two-way clustering is close to t-LAHR but produces different conclusions in panel (b) and may be unreliable in high-persistence, low-signal setups. Finally, Driscoll-Kraay intervals can be misleading with $T = 80$. In fact, they lead to exactly the opposite conclusions about the role of size and leverage.

3.6 Proofs

Proposition 3.1

Let $\tilde{\beta}(h) = \left(\sum_{i=1}^N \hat{s}_i^2\right)^{-1} \sum_{i=1}^N \hat{s}_i \beta_{ih}$ be the coefficient in the (infeasible) regression of β_{ih} on $\hat{s}_i$ — the finite-population counterpart to $\beta(h)$. Also, define

$$\xi_{it}(h,\kappa) = \sum_{\ell=0}^{\infty} \left(\iota_\ell(h)\beta_{i\ell}X_{t+h-\ell} + \gamma_{i\ell}Z_{t+h-\ell} + \kappa\delta_{i\ell}u_{i,t+h-\ell}\right), \tag{3.19}$$

$$\xi_t(h,\kappa) = \frac{1}{N}\sum_{i=1}^N \hat{s}_i\xi_{it}(h,\kappa) = \sum_{\ell=0}^{\infty}\left(\iota_\ell(h)\bar{\beta}_\ell X_{t+h-\ell} + \bar{\gamma}_\ell Z_{t+h-\ell} + \frac{\kappa}{N}\sum_{i=1}^N \hat{s}_i\delta_{i\ell}u_{i,t+h-\ell}\right)$$

where $\iota_\ell(h) = 1 - \mathbb{1}\{h \le \ell \le h+p\}$, $\bar{\beta}_\ell = N^{-1}\sum_{i=1}^N \hat{s}_i\beta_{i\ell}$ and $\bar{\gamma}_\ell = N^{-1}\sum_{i=1}^N \hat{s}_i\gamma_{i\ell}$. Finally, let $V(h,\kappa) = \mathrm{Var}_\kappa\left(X_t\xi_t(h,\kappa)\middle|\{\theta_i, s_i\}_{i=1}^N\right)$.

Proof of Propositions 3.1. Let $\sum_{i,t}$ denote summation over $1 \le t \le T-h$ and $1 \le i \le N$. For any $\psi \in \mathbb{R}^d$,

$$\begin{aligned}\left(\sum_{i,t}\hat{x}_{it}(h)^2\right)\left(\hat{\beta}(h) - \tilde{\beta}(h)\right) &= \sum_{i,t}\hat{x}_{it}(h)\left(Y_{i,t+h} - \tilde{\beta}(h)\hat{s}_iX_t - \psi' W_{it}\right)\\ &= \sum_{i,t}\hat{s}_iX_t\left(Y_{i,t+h} - \beta_{ih}X_t - \psi' W_{it}\right)\\ &\quad - \sum_{i,t}(\hat{\pi}(h)'W_{it})\left(Y_{i,t+h} - \tilde{\beta}(h)\hat{s}_iX_t - \psi' W_{it}\right).\end{aligned}$$

The first line uses $\sum_{i,t}\hat{x}_{it}(h)^2 = \sum_{i,t}\hat{x}_{it}(h)\hat{s}_iX_t$ and $\sum_{i,t}\hat{x}_{it}(h)W_{it} = 0_{d\times 1}$ (to introduce ψ). The second line uses $\hat{x}_{it}(h) = \hat{s}_iX_t - \hat{\pi}(h)'W_{it}$ and $\sum_{i,t}\hat{s}_iX_t(\tilde{\beta}(h)\hat{s}_iX_t - \beta_{ih}X_t) = 0$.

We can choose ψ so that

$$\sum_{i,t}\hat{s}_iX_t\left(Y_{i,t+h} - \beta_{ih}X_t - \psi' W_{it}\right) = \sum_{i,t}\hat{s}_iX_t\xi_{it}(h,\kappa) = N\sum_{t=1}^{T-h}X_t\xi_t(h,\kappa).$$

Here, W_{it} consists of p lags of $\hat{s}_iX_t$, unit indicators, and (possibly) time indicators (so that $d = p+N+T$). To choose ψ, we set the coefficient on $\hat{s}_iX_{t-\ell}$ to $\tilde{\beta}(h+\ell) = \left(\sum_{i=1}^N \hat{s}_i^2\right)^{-1}\sum_{i=1}^N \hat{s}_i\beta_{i,h+\ell}$, the coefficient on the unit-i indicator to μ_i, and the coefficients on time indicators to zero. Moreover, $\hat{\pi}(h)'W_{it} = \hat{s}_i(X_t - \hat{X}_t(h))$

[32] This can happen even in the same exercise because the estimation errors of different coefficients load differently on the macro and micro components of the regression score. Figure 3.4 suggests the size coefficient is driven by the macro component and the other coefficients by the micro component.

with $\hat{X}_t(h)$ the residual from a regression of X_t on $X_{t-1}, \ldots, X_{t-p}$ and an intercept. Then,

$$\sum_{i,t}(\hat{\pi}(h)'W_{it})\left(Y_{i,t+h} - \tilde{\beta}(h)\hat{s}_i X_t - \psi' W_{it}\right) = \sum_{i,t}(\hat{\pi}(h)'W_{it})\xi_{it}(h,\kappa).$$

It follows that the standardized estimation error can be written as

$$\begin{aligned}\frac{\hat{\beta}(h)-\tilde{\beta}(h)}{\hat{\sigma}(h)} &= \frac{\sum_{t=1}^{T-h}\sum_{i=1}^{N}\hat{x}_{it}(h)(Y_{i,t+h}-\tilde{\beta}(h)\hat{x}_{it}(h))}{N\sqrt{(T-h)\hat{V}(h)}} \\ &= \sqrt{\frac{V(h,\kappa)}{\hat{V}(h)}} \times \left(\frac{\sum_{t=1}^{T-h}X_t\xi_t(h,\kappa)}{\sqrt{(T-h)V(h,\kappa)}} + R_T(h,\kappa)\right)\end{aligned}$$

where the remainder term is

$$R_T(h,\kappa) = -\frac{\sum_{t=1}^{T-h}\sum_{i=1}^{N}\left(\hat{\pi}(h)'W_{it}\right)\xi_{it}(h,\kappa)}{N\sqrt{(T-h)V(h,\kappa)}}.$$

To establish our uniform approximation we exploit drifting parameter sequences (see Andrews et al. (2020) for formal results connecting the two). For simplicity we index everything to T, including $N = N_T$. We show that for any $\{\kappa_T\}$, as $T \to \infty$,

(A) $\left\{(T-h)V(h,\kappa_T)\right\}^{-1/2}\sum_{t=1}^{T-h}X_t\xi_t(h,\kappa_T) \xrightarrow[P_{\kappa_T}]{\text{d}} N(0,1)$,

(B) $\hat{V}(h)/V(h,\kappa_T) \xrightarrow[P_{\kappa_T}]{\text{p}} 1$,

(C) $R_T(h,\kappa_T) \xrightarrow[P_{\kappa_T}]{\text{p}} 0$.

Hence, for any such $\{\kappa_T\}$,

$$\frac{\hat{\beta}(h)-\tilde{\beta}(h)}{\hat{\sigma}(h)} \xrightarrow[P_{\kappa_T}]{\text{d}} N(0,1).$$

We establish (A), (B) and (C) in Lemmas B.1, B.2 and B.3 in Appendix B.1. Now, Assumptions 3.1(ii) and 3.3(iv) imply $\tilde{\beta}(h) - \beta(h) = O_{P_{\kappa_T}}\left(N^{-1/2}\right)$ whereas Lemma B.2 implies $\min\{1, \kappa_T^{-1}\}\hat{\sigma}(h) = O_{P_{\kappa_T}}\left((T-h)^{-1/2}\right)$. Since $T/N \to 0$,

$$\frac{(\hat{\beta}(h)-\beta(h))}{\hat{\sigma}(h)} = \frac{(\hat{\beta}(h)-\tilde{\beta}(h))}{\hat{\sigma}(h)} + o_{P_{\kappa_T}}(1)$$

and the result follows. □

Proposition 3.2

Define

$$\xi_{it}(h,\kappa) = \sum_{\ell=0}^{h} \left(\iota_\ell(h)\beta_{i\ell}X_{t+h-\ell} + \gamma_{i\ell}Z_{t+h-\ell} + \kappa\delta_{i\ell}u_{i,t+h-\ell}\right), \tag{3.20}$$

$$\xi_t(h,\kappa) = \frac{1}{N}\sum_{i=1}^{N}\hat{s}_i\xi_{it}(h,\kappa) = \sum_{\ell=0}^{h}\left(\iota_\ell(h)\bar{\beta}_\ell X_{t+h-\ell} + \bar{\gamma}_\ell Z_{t+h-\ell} + \frac{\kappa}{N}\sum_{i=1}^{N}\hat{s}_i\delta_{i\ell}u_{i,t+h-\ell}\right),$$

and, as before, let $V(h,\kappa) = \mathrm{Var}_\kappa\left(X_t\xi_t(h,\kappa)\middle|\{\theta_i, s_i\}_{i=1}^N\right)$. By recursive substitution,

$$Y_{i,t+h} = m_i(h) + \sum_{\ell=1}^{p}(A_\ell(h)Y_{i,t-\ell} + B_{i\ell}(h)X_{t-\ell}) + \beta_{ih}X_t + \xi_{it}(h,\kappa),$$

for some $m_i(h)$, $\{A_\ell(h)\}$, $\{B_{i\ell}(h)\}$ that depend on the VAR parameters m_i, $\{A_\ell\}$, $\{B_{i\ell}\}$.

Proof of Proposition 3.2. We follow exactly the same steps as for Proposition 3.1. The control vector W_{it} includes p lags of Y_{it} and $\hat{s}_iX_t$ in addition to unit and time effects. In the step where we choose ψ, we set the coefficient on $Y_{i,t-\ell}$ to $A_\ell(h)$, the coefficient on $\hat{s}_iX_{t-\ell}$ to $\tilde{B}_\ell(h) = \left(\sum_{i=1}^N \hat{s}_i^2\right)^{-1}\sum_{i=1}^N \hat{s}_iB_{i\ell}(h)$, the coefficient on the unit-i indicator to $m_i(h)$, and the coefficients on time indicators to zero.

The standardized estimation error can then be written as

$$\frac{\hat{\beta}(h)-\tilde{\beta}(h)}{\hat{\sigma}(h)} = \sqrt{\frac{V(h,\kappa)}{\hat{V}(h)}} \times \left(\frac{\sum_{t=1}^{T-h}X_t\xi_t(h,\kappa)}{\sqrt{(T-h)V(h,\kappa)}} + R_T(h,\kappa)\right)$$

where the remainder term is now

$$R_T(h,\kappa) = -\frac{\sum_{t=1}^{T-h}\sum_{i=1}^{N}\left(\hat{\pi}(h)'W_{it}\right)\left[(\beta_{ih}-\tilde{\beta}(h)\hat{s}_i)X_t + \sum_{\ell=1}^{p}(B_{i\ell}(h)-\tilde{B}_\ell(h)\hat{s}_i)X_{t-\ell} + \xi_{it}(h,\kappa)\right]}{N\sqrt{(T-h)V(h,\kappa)}}.$$

Let $\phi < 1$. In contrast to Proposition 3.1, instead of a single drifting parameter we now have two. We show that for any $\{h_T, \kappa_T\}$ such that $h_T \le \phi T$,

(A) $\left\{(T-h_T)V(h_T,\kappa_T)\right\}^{-1/2}\sum_{t=1}^{T-h_T}X_t\xi_t(h_T,\kappa_T) \xrightarrow[P_{\kappa_T}]{\mathrm{d}} N(0,1)$,

(B) $\hat{V}(h_T)/V(h_T,\kappa_T) \xrightarrow[P_{\kappa_T}]{\mathrm{p}} 1$,

(C) $R_T(h_T,\kappa_T) \xrightarrow[P_{\kappa_T}]{\mathrm{p}} 0$.

We prove (A), (B) and (C) in Lemmas B.8, B.9 and B.10 in Appendix B.1. The rest of the argument is identical to that of Proposition 3.1. □

Proposition 3.3

Using (3.17), substitute $\tilde{X}_t, \tilde{X}_{t-1}, \ldots, \tilde{X}_{t-p}$ in succession into (3.6)–(3.7) to obtain

$$Y_{i,t+h} = \mu_i + \beta_{ih}\tilde{X}_t + \sum_{\ell=1}^{p} \tilde{\eta}_{i\ell}\tilde{X}_{t-\ell} + \xi_{it}(h,\kappa),$$

$$\xi_{it}(h,\kappa) = \sum_{\ell=0}^{\infty} \left(\iota_\ell(h)\tilde{\beta}_{i\ell}X_{t+h-\ell} + \tilde{\gamma}_{i\ell}Z_{t+h-\ell} + \kappa\delta_{i\ell}u_{i,t+h-\ell} \right),$$

for some coefficients $\{\tilde{\eta}_{i\ell}\}$, $\{\tilde{\beta}_{i\ell}\}$, $\{\tilde{\gamma}_{i\ell}\}$ that depend on $\{\beta_{i\ell}\}$, $\{\gamma_{i\ell}\}$, $\{b_\ell\}$, $\{c_\ell\}$ and satisfy the bound conditions in Assumption 3.3 for a suitable choice of C_ℓ and $\underline{C}$. Also define $\tilde{\boldsymbol{\beta}}(h) = \left(\sum_{i=1}^N \hat{s}_i^2\right)^{-1} \sum_{i=1}^N \hat{s}_i \boldsymbol{\beta}_{ih}$ with $\boldsymbol{\beta}_{ih} = (\beta_{ih}, \tilde{\eta}_{i1}, \ldots, \tilde{\eta}_{ip})'$, $\xi_t(h,\kappa) = N^{-1}\sum_{i=1}^N \hat{s}_i \xi_{it}(h,\kappa)$ and $\boldsymbol{V}(h,\kappa) = \mathrm{Var}_\kappa \left(\boldsymbol{X}_t^* \xi_t(h,\kappa) \middle| \{\theta_i, s_i\}_{i=1}^N \right)$.

Proof of Proposition 3.3. Following similar steps to the derivation in Proposition 3.1, let $\sum_{i,t}$ denote summation over $1 \le t \le T-h$ and $1 \le i \le N$. For any ψ,

$$\begin{aligned} \left(\sum_{i,t} \hat{\boldsymbol{x}}_{it}(h)\hat{s}_i \tilde{\boldsymbol{X}}_t' \right) \left(\hat{\boldsymbol{\beta}}^{\mathrm{IV}}(h) - \tilde{\boldsymbol{\beta}}(h) \right) &= \sum_{i,t} \hat{s}_i \boldsymbol{X}_t^* \left(Y_{i,t+h} - \tilde{\boldsymbol{X}}_t' \tilde{\boldsymbol{\beta}}_{ih} - \psi' W_{it} \right) \\ &\quad - \left(\frac{\sum_{t=1}^{T-h} \boldsymbol{X}_t^*}{(T-h)} \right) \sum_{i,t} \hat{s}_i \left(Y_{i,t+h} - \hat{s}_i \tilde{\boldsymbol{X}}_t' \tilde{\boldsymbol{\beta}}(h) - \psi' W_{it} \right). \end{aligned}$$

Note W_{it} includes unit and (possibly) time effects. To choose ψ, set the coefficient on the unit-i indicator to μ_i and the coefficients on time indicators to zero, so that

$$\sum_{i,t} \hat{s}_i \boldsymbol{X}_t^* \left(Y_{i,t+h} - \tilde{\boldsymbol{X}}_t' \tilde{\boldsymbol{\beta}}_{ih} - \psi' W_{it} \right) = N \sum_{t=1}^{T-h} \boldsymbol{X}_t^* \xi_t(h,\kappa),$$

$$\left(\frac{\sum_{t=1}^{T-h} \boldsymbol{X}_t^*}{(T-h)} \right) \sum_{i,t} \hat{s}_i \left(Y_{i,t+h} - \hat{s}_i \tilde{\boldsymbol{X}}_t' \tilde{\boldsymbol{\beta}}(h) - \psi' W_{it} \right) = \left(\frac{\sum_{t=1}^{T-h} \boldsymbol{X}_t^*}{(T-h)} \right) \sum_{t=1}^{T-h} \xi_t(h,\kappa).$$

Thus, the standardized estimation error can be written as

$$\begin{aligned} \frac{\hat{\beta}_0^{\mathrm{IV}}(h) - \tilde{\beta}(h)}{\hat{\sigma}_0^{\mathrm{IV}}(h)} &= \sqrt{\frac{\left(e_1'\boldsymbol{J}^{-1}\right) \boldsymbol{V}(h,\kappa) \left(e_1'\boldsymbol{J}^{-1}\right)'}{\left(e_1'\hat{\boldsymbol{J}}^{\mathrm{IV}}(h)^{-1}\right) \hat{\boldsymbol{V}}^{\mathrm{IV}}(h) \left(e_1'\hat{\boldsymbol{J}}^{\mathrm{IV}}(h)^{-1}\right)'}} \\ &\quad \times \left(\frac{\left(e_1'\hat{\boldsymbol{J}}^{\mathrm{IV}}(h)^{-1}\right) \sum_{t=1}^{T-h} \boldsymbol{X}_t^* \xi_t(h,\kappa)}{\sqrt{(T-h)\left(e_1'\boldsymbol{J}^{-1}\right) \boldsymbol{V}(h,\kappa) \left(e_1'\boldsymbol{J}^{-1}\right)'}} + R_T(h,\kappa) \right) \end{aligned}$$

where $\boldsymbol{J} = (N^{-1}\sum_{i=1}^{N}\hat{s}_i^2)E\left[\boldsymbol{X}_t^*\tilde{\boldsymbol{X}}_t'\right]$ and the remainder term is

$$R_T(h,\kappa) = -\frac{\left\{(T-h)^{-1}\left(e_1'\hat{\boldsymbol{J}}^{\text{IV}}(h)^{-1}\right)\sum_{t=1}^{T-h}\boldsymbol{X}_t^*\right\}\sum_{t=1}^{T-h}\xi_t(h,\kappa)}{\sqrt{(T-h)\left(e_1'\boldsymbol{J}^{-1}\right)\boldsymbol{V}(h,\kappa)\left(e_1'\boldsymbol{J}^{-1}\right)'}}.$$

As in Proposition 3.1, we show that for any $\{\kappa_T\}$ and $\boldsymbol{\lambda} \neq 0_{(p+1)\times 1}$

(A) $\left\{(T-h)\boldsymbol{\lambda}'\boldsymbol{V}(h,\kappa_T)\boldsymbol{\lambda}\right\}^{-1/2}\sum_{t=1}^{T-h}\boldsymbol{\lambda}'\boldsymbol{X}_t^*\xi_t(h,\kappa_T) \xrightarrow[P_{\kappa_T}]{\text{d}} N(0,1)$,

(B) $\left(\boldsymbol{\lambda}'\hat{\boldsymbol{V}}^{\text{IV}}(h)\boldsymbol{\lambda}\right)/(\boldsymbol{\lambda}'\boldsymbol{V}(h,\kappa_T)\boldsymbol{\lambda}) \xrightarrow[P_{\kappa_T}]{\text{p}} 1$ and $\hat{\boldsymbol{J}}^{\text{IV}}(h) \xrightarrow[P_{\kappa_T}]{\text{p}} \boldsymbol{J}$,

(C) $R_T(h,\kappa_T) \xrightarrow[P_{\kappa_T}]{\text{p}} 0$.

The technical steps for (A), (B), and (C) are stated in Lemmas B.11, B.12 and B.13 in Appendix B.1. The rest of the argument is as in Proposition 3.1. □

CHAPTER 4

ESTIMATING FLEXIBLE INCOME PROCESSES FROM SUBJECTIVE EXPECTATIONS DATA: EVIDENCE FROM INDIA AND COLOMBIA

WITH MANUEL ARELLANO, ORAZIO ATTANASIO AND SAM CROSSMAN

4.1 Research context

Households allocate current income between consumption and savings taking into account the uncertainty about their future income *as they perceive it*. How persistent they perceive their future income flows to be, their dispersion or asymmetry are all subjective features of households' income uncertainty that critically impact their spending plans. Those perceptions and their heterogeneity across households are also key determinants of economy-wide consumption inequality.

Conventional approaches to identify the stochastic process of uncertain variables are *indirect*, relying on statistical models of the dynamics of the realized variables and/or models of choice.[1] Under rational expectations, income dynamics as perceived by households may coincide with the dynamics of realized income, but this needs not be the case. An alternative *direct* approach is to rely on subjective probabilistic expectation questions from surveys. In this chapter, we develop a methodology for modeling household income processes using subjective expectations of future income. Our approach is flexible enough to assess the extent of nonlinear persistence and non-Gaussian distributional features in households' perceptions. We then take our methods to subjective expectations data elicited within two surveys that were conducted in Colom-

[1] See surveys of the literature on earnings dynamics in Meghir and Pistaferri (2011), Arellano (2014), and Altonji and Vidangos (2023).

bia and India. Learning about the nature of income uncertainty is particularly important in developing economies where there tends to be more volatility.

Subjective expectations have been around for a while. For a long time they were received with skepticism by some, but the current evidence is that individuals are able to respond to probabilistic questions about variables that matter to them in a meaningful way (Manski, 2004, 2018; Delavande, Giné, and McKenzie, 2011). Specifically, for developing countries a lot of progress has been made in understanding the implications of different methods of eliciting expectations (Attanasio, 2009; Delavande, 2023).

Some of the early work on subjective income expectations is due to Dominitz and Manski (1997b). They used responses to the probability questions in their survey to fit respondent-specific parametric distributions, which they compared with those implied by the income processes used in Hall and Mishkin (1982). They found that subjective dispersion measures varied across households and were not proportional to subjective medians (see also Dominitz (1998, 2001)). Attanasio and Augsburg (2016) were the first to use the subjective expectations data in the survey of Indian households in combination with current income to estimate an income process.

We are also motivated by recent work on flexible income processes. A recent literature has uncovered significant non Gaussian nonlinearities in the dynamics of realized incomes (Arellano, Blundell, and Bonhomme, 2017; Guvenen, Karahan, Özkan, and Song, 2021). These nonlinearities are potentially relevant for individual behavior and policy design, like saving choices (De Nardi and Paz-Pardo, 2020) or optimal income taxation (Golosov and Tsyvinski, 2015). It is of great interest to find out if these nonlinearities are also present in the subjective expectations of poor households in developing contexts.

Our first contribution is to show how to identify and estimate a standard (log) linear dynamic model for household income, with and without fixed effects, using data on subjective expectations and current income. Our approach is to map the model directly to individual subjective probabilities, and in particular to the *observed* log odd ratios, which we regard as noisy measures of the model counterparts, subject to an additive elicitation error. Fixed effects estimation of the model parameters is robust to un-modelled distributional heterogeneity of elicitation errors and, contrary to indirect approaches based on realized income, does not suffer from Nickell bias (Nickell, 1981). This is a convenient feature of subjective expectation models, since unobserved disturbances do not contain future shocks but only measurement errors in the elicited probabilities.

We use the log-linear model as starting point that conveys the main ideas of our approach. We then propose a generalized estimation framework to deal with nonlinear income processes with unobserved heterogeneity. We consider a sieve approach with a sequence of flexibly parameterized predictive conditional distributions. Despite their generality, these distributions can be cast as static fixed effects models that can be estimated by within-group methods. We also explore extensions with more general patterns of unobserved heterogeneity where log odd ratios can vary differentially with individual effects. Our approach allows us to estimate subjective measures of risk and persistence that may differ across observed income levels, the size of income shocks, and individual effects.

In our empirical analysis, we use two waves from both the Colombian and Indian surveys, combining expectations with actual income data. In fact, the combination of the two is essential to our approach. In both surveys, income expectations were collected using Dominitz and Manski (1997a,b) elicitation method, alongside realized income and other indicators of the nature and sources of earnings. Respondents are asked to provide a relevant range of variation for their future income. Next, they are asked to report the probability that their future income will exceed each of three equally-spaced points within their selected range. These elicited probabilities are the individual-level outcomes in our models.

We reject the standard linear model in the data on subjective expectations from the two surveys in favor of more flexible models. Subjective income distributions exhibit nonlinear persistence, along with dispersion and skewness that vary with current income levels and unobserved heterogeneity. Interestingly, we find a negative association between conditional dispersion and current income, and between conditional skewness and current income.

Estimated persistence plummets for poorer households experiencing large positive shocks, but not for relatively affluent households experiencing negative shocks. Those findings for the perceived risks of households in developing economies are partially consistent with the results found in Arellano et al. (2017) for the realized incomes of US households from the Panel Study of Income Dynamics (PSID). Essentially, we find low persistence for large positive shocks at the bottom of the income distribution as they do, but not for large negative shocks at the top. In interpreting the results, we argue that the nonlinear persistence we find for the poorest households is consistent with a poverty trap interpretation. We also find that unobserved heterogeneity matters, and is composed of household specific and village level factors. Households with large fixed effects have more persistent histories overall and less variability in persistence with current income and shock size. The pattern of nonlinear persistence is robust to allowing for more general forms of unobserved heterogeneity, although quantitatively its importance is reduced.

The chapter is structured as follows. In the next section, we discuss how probabilistic subjective expectations are elicited in the two surveys that we use. Section 4.3 lays out our modeling and estimation framework, first for a linear process with fixed effects and then for more flexible models. Section 4.4 describes the two survey data sets that we use for the analysis. In Section 4.5, we present our empirical results. Finally, we conclude in Section **??**. The appendices contain additional results and technical material.

4.2 Eliciting subjective expectations

Designing subjective expectation questionnaires to elicit information about respondents' perceived probability distribution of future variables is challenging. Several open questions remain in the growing literature on the topic, ranging from the establishment of a metric for the variables of interest to the way conditional and unconditional probability measures are elicited. As a consequence, important choices need to be made throughout the process.

In this section, we first briefly discuss some of the outstanding issues in the literature and then describe the approach used to elicit subjective expectations in our two surveys, which employed similar methods and questionnaire designs. While not particularly novel, it is useful to describe and relate them to possible alternatives.

4.2.1 Anchoring subjective expectations

A first issue in the design of subjective expectation questions is establishing an anchor and a metric for the variable whose probability distribution is being elicited.

In eliciting the probability distribution of future income, two different approaches have been used. In some surveys, the current value of income is used as an implicit anchor, and respondents are asked the probability of a number of possible percentage changes of future income relative to current income. In other contexts, respondents are asked to provide a range of possible values, often the *minimum* and *maximum* for future income. These values are then used to define a number of intervals and respondents are asked the probability that future income will fall in each of these intervals. This approach was developed by Dominitz and Manski (1996, 1997a,b) and has been widely adopted in surveys across both developed and developing economies, including the Indian and Colombian datasets we use. The precise formulation of these questions is described in an elaborate script, which is reported in Appendix C.5.

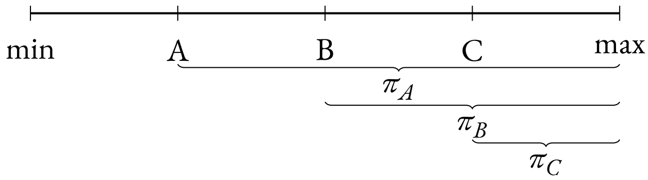
Morgan and Henrion (1990) point out that asking first for the minimum and maximum of possible income realizations may help reduce two common problems in the elicitation of expectations. The first is *overconfidence*, wherein respondents focus too much on central tendencies and therefore understate the true uncertainty that they face — asking about the minimum and maximum first helps to prime respondents to think about the full range of probable realisations. The second is *anchoring* or *framing*, whereby figures provided by the interviewer might influence the responses provided: if the chosen *cdf* support points are specified by the interviewer, respondents may be inclined to think these points are salient for one reason or another, and therefore likely to restrict their answers around those values.

Using current income as an anchor for future income might be problematic when the former is unusually low or high, in that it is unclear whether the elicited probability distribution around that value is particularly informative. This is avoided by the minimum/maximum approach as respondents can *re-center* their answers around the information they have. On the other hand, it is not obvious whether the elicited minimum/maximum values are really what they are labeled to be or rather some arbitrary low or high percentile of the subjective future income distribution.

4.2.2 Eliciting subjective probability distributions

Having registered the minimum and maximum, the interviewers elicit from respondents information on several intermediate points on the subjective cumulative distribution function (*cdf*) of future total household income. The minimum and maximum are used to compute, using a simple pre-specified algorithm, a set of J points within the support of the distribution of future income $\{y_{k1}, y_{k_2}, \ldots, y_{k_J}\}$. Respondents are then asked about the probabilities $\{\pi_{k1}, \pi_{k_2}, \ldots, \pi_{k_J}\}$ they assign to their next period income being larger than these points. In both survey data from India and Colombia, $J = 3$ was used. Furthermore, as shown in Figure 4.1, each sub-interval is of equal size.

A possible objection to this method is that respondents, especially from disadvantaged backgrounds, might be unfamiliar with the concept of probability and with its translation into numerical values. These issues might be particularly relevant in the context of developing countries, where respondents often have no or very limited formal education. In such a situation, the use of preliminary priming questions that could



Note. The figure illustrates the way of eliciting three points on the subjective income distribution. Respondents are asked to provide $y_{\min}$ and $y_{\min}$. The interviewer then computes $y_B = (y_{\min} + y_{\max})/2$, $y_A = (y_{\min} + y_B)/2$ and $y_C = (y_B + y_{\max})/2$, and proceeds to elicit probabilities π_A, π_B and π_C.

FIGURE 4.1. A description of the elicitation process.

familiarize respondents with the theoretical constructs that researchers want to elicit, as well as the use of visual aids, may be advisable (Delavande et al., 2011).

For this reason, in both surveys, respondents were first *primed* in the use of the concept of probability and conditional probability through specific examples about future uncertain events. In particular, a sequence of questions was asked about the likelihood of rain, designed to ensure that probabilities are non-decreasing. For instance, the question "What is the probability that it will rain tomorrow?" is followed by "What is the probability that it rains in the next seven days?", pointing out that the latter should be no smaller than the former.

As a form of visual aid for the probability questions, a *ruler* graded from 0 to 100 was used. Respondents were instructed to point to 0 to express the certainty that an event will not happen, to 100 for the certainty that it will, and to intermediate points to express uncertainty. While this approach seems to have worked in this context, it is not a silver bullet: different methods might be necessary in different contexts.[2]

During the data collection, interviewers were instructed to correct respondents' inconsistent answers during the training phase but not during the actual subjective expectations questions about future income. Respondents might, therefore, provide inconsistent answers, thereby flagging possible quality problems in the data. We explore this in depth in Section 4.4.

When designing expectations questions, an important choice is the number of intervals into which the range of values identified by the minimum and maximum is divided and the placement of cutoff points y_{k_j} for $j = 1, \dots, J$. While a high number of cutoff points would increase the information on the *cdf*, improving the ability to fit flexible and possibly complex subjective *cdf*'s, such high values of J might impose an excessive burden on respondents and jeopardize the quality of the data. In a variety of contexts, J has been set at 1 or 3.

Another issue is whether the minimum and maximum should be treated as the genuine minimum and maximum of future income or as values where the *cdf* takes values relatively close to 0 or 1, respectively. We avoided committing to a specific interpretation of the minimum and maximum values, so that they play no role in our formal analysis beyond providing a range of respondent-specific points at which the *cdf* is elicited.

[2]Other surveys have asked respondents to allocate stones, balls or other items available in the local context, into a number of bins, see, for instance, Delavande and Rohwedder (2008).

4.3 Mapping income processes to subjective expectations

In this section, we show how to use data on subjective expectations to estimate models of household income dynamics, as perceived by respondents. We discuss the econometric approach we take to this problem, and show that the use of subjective expectations data poses inference problems that are conceptually different from those present when estimating dynamic models using actual income realisations. We start our discussion with a relatively simple (log) linear model, which is particularly useful in conveying the main ideas of our approach. We then generalize our approach and show that these data can also be used to estimate more complex and flexible income processes.

We interpret the models of the income processes as representing the conditional subjective probability distribution respondents hold, given the information available to them, including current income and other conditioning variables. To estimate the parameters of these models we then match the answers respondents give to the subjective expectations questions to the corresponding quantities implied by the statistical model we specify. As we discuss below, the identification of the structural parameters of the statistical models we consider relies on a number of assumptions. However, we argue that such assumptions are different and weaker than those used when estimating such models with actual income realisations. This approach allows us to use a wide class of estimators without incurring the biases that would affect such estimators when using actual income realizations.

4.3.1 Modeling approach

We take advantage of the availability of subjective expectations data to fit a model for the conditional *cdf* directly to the observed individual subjective probabilities. Let a household's subjective (conditional) cumulative probability distribution of log future income $y_{i,t+1}$ be denoted as

$$F_{it}(r) = P\left(y_{i,t+1} \leq r \middle| I_{it}\right), \tag{4.1}$$

where I_{it} denotes the information set available to household i in period t. As discussed in Section 4.2, the survey elicitation process employed in the datasets we use yields noisy measurements p_{jit} of $F_{it}\left(r_{jit}\right)$ for $r_{jit} = r_{it}^{\min} + \left(r_{it}^{\max} - r_{it}^{\min}\right) j/4$ $\left(j = 1, 2, 3\right)$, or equivalently of the subjective cumulative odds

$$\ell_{jit}^{*} = \text{logit}\left[F_{it}\left(r_{jit}\right)\right], \tag{4.2}$$

where $\text{logit}(p) = \ln\left[p/(1-p)\right]$. We model the log of the subjective cumulative odds, so that the outcomes we consider have an unlimited range of variation.[3] We also allow for a survey elicitation error ε_{jit}, which is plausibly assumed to be additive in the log of cumulative odds, so that the observed cumulative odds $\ell_{jit} = \text{logit}\left(p_{jit}\right)$ are given by

$$\ell_{jit} = \ell_{jit}^{*} + \varepsilon_{jit}. \tag{4.3}$$

[3]Notice that this transformation rules out observations with $p_{jit} = 0$ or $p_{jit} = 1$. In Appendix C.4, we explore an alternative to (4.2) that introduces an adjustment to the logit function that can be interpreted as proportional to the accuracy of the elicitation process. This approach allows us to retain these observations and verify that our empirical results are largely unchanged.

We assume that ε_{jit} is a classical measurement error, in the sense that it is mean independent of I_{it}, the information set in equation (4.1). Apart from that, we allow for dependence in ε_{jit} across j and t. Moreover, the variance or other moments of ε_{jit} may change with j and t, and may also depend on variables in the information set. Modeling the distribution of elicitation errors is of separate interest, but the linearity assumption allows us to leave this distribution unmodeled while being robust to a variety of elicitation error configurations.

We only observe three points of F_{it} for each unit, but many different points across units. The general idea is to learn by combining data for all units; as long as there is sufficient variability in r_{jit} and common features in the probability distributions across units, they are potentially nonparametrically identifiable.

Information set. The information set is assumed to be Markovian in the sense that given the current values of the relevant variables, values from earlier periods cannot reduce subjective prediction uncertainty. In our analysis, the Markov property is assumed to hold conditionally given unobserved heterogeneity.

The set I_{it} consists of time-varying and time-invariant characteristics. The time-varying variables include observable current income y_{it} and indicators x_{it} of the nature and sources of income, such as the number of earners in the household, as well as household demographics. As for the time-invariant characteristics, we adopt a latent variable approach, assuming that they can be captured by an unobservable individual effect α_i. This effect is intended to encompass both household-level characteristics and geographical (say, village-level) characteristics. We therefore assume that $I_{it} = \left(y_{it}, x_{it}, \alpha_i\right)$. The individual effect α_i may be correlated with $\left(r_{jit}, y_{it}, x_{it}\right)$.

A system of equations. The relationship between the information set available to individual households (part of which might be unobservable to the econometrician) and the elicited conditional *cdf*s depends on the specific model being considered. Here we define such a relationship as a function g, so to obtain:

$$\ell_{jit}^{*} = g\left(r_{jit}, y_{it}, x_{it}, \alpha_i\right) \qquad \left(i = 1, \ldots, n; j = 1, 2, 3; t = 1, 2\right) \tag{4.4}$$

where g is a non-decreasing function in its first argument. We specify below the function g corresponding to different models of income dynamics. Thus, our econometric model consists of a system of six equations for n households with the addition of measurement errors in elicited probabilities.[4]

Identification of nonlinear panel data models with continuous outcomes and unobserved heterogeneity has been discussed in Evdokimov (2010), Arellano and Bonhomme (2016), Hu (2017), and Schennach (2022), amongst others. Nonparametric identification of the response function g and the conditional distribution of α_i in model (4.3) and (4.4) can be established using the arguments in Evdokimov (2010), under the assumption that the additively separable disturbance terms are conditionally independent over time, and independent of the individual effect α_i.

Next, we consider alternative specifications of the income model, starting with the simplest version, which assumes linearity and corresponds to models that have been widely used in the literature on income

[4]Note that an additive α_i could be reflecting both persistent elicitation differences (that is, part of measurement error) or heterogeneity in income risk. This distinction will matter for interpretation when documenting the effect of "heterogeneity" in nonlinear models.

processes.

Income processes. In a life-cycle model of income and consumption choices, a popular specification decomposes household income into the product of a deterministic (or profile) component, which might include a fixed effect, and a stochastic component, often in the form of persistent shocks with autoregressive dynamics (sometimes also including transitory shocks). A log-linear model of this kind with no transitory shocks can be written as

$$Y_{i,t+1} = Y_{it}^{\rho} V_{i,t+1} \exp\left(p_{i,t+1} + \alpha_i\right),$$

where Y_{it} is the level of income for household i at time t, $V_{i,t}$ innovations to income, $p_{i,t}$ captures household age and demographic variables, and α_i represents the fixed effect. In both of our data sets, these fixed effects can be decomposed into village-level and purely idiosyncratic effects. We omit this distinction here for notational simplicity. Taking logs, we have

$$y_{i,t+1} = \rho y_{it} + p_{i,t+1} + \alpha_i + v_{i,t+1}, \tag{4.5}$$

where $y_{it} = \ln Y_{it}$ and $v_{it} = \ln V_{it}$.

One could consider decomposing the stochastic part of income into persistent and transitory components. In a standard persistent/transitory model, consumers are assumed to observe the values of the two components as separate state variables, whereas they remain unobserved to the modeler. However, when conditioning on current income as we do, this situation introduces a measurement-error problem that can be dealt with instrumental variables. Such an approach cannot be used in a two-wave panel like ours. While we do not consider this possibility in our application, we do estimate multiple-state processes that include indicators of the nature and sources of income, whose motivation is not entirely different from that behind unobservable income component models.

4.3.2 Predictive distributions for linear income processes

We first illustrate how our approach allows us to estimate conditional distributions of subjective income risk when the underlying income process is a standard log-linear autoregressive model. This is a convenient benchmark which provides a simple framework to illustrate identification and estimation issues, while highlighting the benefits of using subjective expectations data relative to a more standard approach using income realizations.

Considering a first-order autoregressive process with fixed effects,[5] we can rewrite equation (4.5) as follows:

$$y_{i,t+1} = \alpha_i + \rho y_{it} + \sigma v_{i,t+1}, \tag{4.6}$$

where $v_{i,t+1}$ are assumed to have a logistic distribution independent of y_{it} and α_i. The corresponding condi-

[5] We discuss specifications with time-varying characteristics below; see Section 4.3.3. On a similar note, we include time (wave) effects in all models that we consider, but omit them from explicit formulas for notational simplicity.

tional *cdf* is then

$$P\left(y_{i,t+1} \leq r \middle| y_{it}, \alpha_i\right) = P\left(v_{i,t+1} \leq \frac{r - \alpha_i - \rho y_{it}}{\sigma} \middle| y_{it}, \alpha_i\right)$$
$$= \Lambda\left(\frac{r - \alpha_i - \rho y_{it}}{\sigma}\right),$$

where $\Lambda(x) = \left(1 + \exp(-x)\right)^{-1}$ is the standard logistic *cdf*. Applying the logit transformation, it follows that in this case g in (4.4) is linear, since we can write

$$\ell_{jit} = \ell^*_{jit} + \varepsilon_{jit} = \beta_0 r_{jit} + \beta_1 y_{it} + \eta_i + \varepsilon_{jit}, \tag{4.7}$$

where $\beta_0 = 1/\sigma$, $\beta_1 = -\rho/\sigma$ and $\eta_i = -\alpha_i/\sigma$. The logit transformation in (4.2), therefore, allows us to map the "structural" parameters ρ and σ to the "reduced-form" estimation parameters in equation (4.7).[6] Equation (4.7) is a linear panel model with fixed effects and strictly exogenous regressors, and a standard within group estimator yields consistent estimates of the parameters.

As discussed in Section 4.4 below, in both surveys we use, the observations are clustered in villages. Therefore, when considering fixed effects, we allow for village-specific means via the following decomposition:

$$\eta_i = \bar{\eta}_{v(i)} + \tilde{\eta}_{i,v(i)},$$

where the subscript $v(i)$ indicates the village of household i, $\bar{\eta}_{v(i)}$ is a village-specific mean, and $\tilde{\eta}_{i,v(i)}$ denotes the deviation of the individual fixed effect from the village average. Regardless of whether the variance of the purely idiosyncratic fixed effects is constant across villages or not, we can estimate the variance of the village fixed effects and the unconditional variance of the purely idiosyncratic fixed effects.

Subjective expectations and income realizations. Despite superficial similarities there are profound differences between the subjective expectation and observed income approaches. First, with subjective expectations data, an AR(1) model without fixed effects can be estimated on a *single cross-section*, as information on *expected future income* (on the left-hand side) is provided by subjective expectations. If fixed-effects are included, the variance of the shock (a measure of risk) can still be estimated on a single cross-section, and the full model would require two waves of data — whereas the observational approach would need at least three.

Second, estimates from subjective expectations represent the perceptions individual households have of their own income, even if they do not have rational expectations. Such an object is what is relevant for household consumption and saving decisions.

Finally, estimation of the model using subjective expectations data does not suffer from the so-called Nickell bias, and so there is no need to use instrumental variable techniques, despite the small time dimension. This is typically not the case when using only income realizations. The reason is that outcomes are not future incomes but rather points in the predictive distribution; therefore, the error term does not contain future shocks but only measurement error in predictive probabilities.

[6] We would obtain a similar mapping if we assumed, for instance, that $\ell_{jit} = \text{probit}\left(p_{jit}\right)$ and $v_{i,t+1} \sim \mathrm{N}(0,1)$, independent of y_{it} and α_i.

While the linear model is useful to illustrate how subjective expectations can be used to recover the parameters of a standard income process, it still imposes a number of tight restrictions regardless of whether subjective expectations or realized income data are used in estimation. For example, persistence ρ and dispersion σ are common to all households in equation (4.7), a restriction that we relax next.

4.3.3 Enlarging the state space

The existing literature has mainly focused on single-state processes in which current income (or a persistent/transitory decomposition of income) is a sufficient statistic for the information set in a household's predictive distribution of future income. However, it is possible that indicators of the nature and sources of income and/or the occurrence of specific shocks help predict future income over and above total current income. If so, consumption decisions might depend on the joint probability distribution of a vector of future variables. Multivariate models of income dynamics are beyond the scope of this chapter, but it is still of interest to find out if our subjective probabilities of future income depend on a larger state space than current income.

Thus, additional flexibility can be added by including relevant time-varying household characteristics x_{it} in the conditioning set, which extends (4.7) to

$$\ell_{jit} = \beta_0 r_{jit} + \beta_1 y_{it} + \delta_0' x_{it} + \delta_1' x_{it} y_{it} + \eta_i + \varepsilon_{jit}. \tag{4.8}$$

In our empirical analysis, we also provide results for models of this type.

4.3.4 Flexible income processes

We now generalize the linear model in equation (4.7) to the following specification:

$$\ell_{jit} = \beta_0(r_{jit}) + \beta_1(r_{jit})\psi\left(y_{it}\right) + \beta_2(r_{jit})\eta_i + \varepsilon_{jit}, \tag{4.9}$$

where $\beta_s(r_{jit})$ for $s = \{0, 1, 2\}$ and $\psi\left(y_{it}\right)$ are functions such as splines or orthogonal polynomials and, again, we omit time-varying observables x_{it} for simplicity. Models with additive fixed effects correspond to setting $\beta_2(r_{jit}) = 1$, whereas the full generality of (4.9) allows for interactive effects.[7] The linear model (4.7) is a special case of (4.9) with linear $\beta_0(\cdot)$ and $\psi(\cdot)$ and constant $\beta_1(\cdot)$ and $\beta_2(\cdot)$.

This model is reminiscent of distribution regression, but the empirical setup is rather different. In distribution regression, one would use realized data on $y_{i,t+1}$ and estimate a sequence of logit or probit regressions for binary outcomes defined as $\mathcal{I}(y_{it} < r)$ to get estimates of $\beta_k(r)$ for different chosen values of r.[8] In our context, we observe $P\left(y_{i,t+1} \leq r \middle| y_{it}, \alpha_i\right)$ for $r = r_{jit}$, so that we can fit these observed probabilities to the specific model we consider. To perform such an exercise, the functions $\beta_s(r)$ and $\psi(\cdot)$ need to be parameterized. Implementation and estimation details are discussed in Section 4.3.5 below.

[7] In principle, interactions between η_i and y_{it} could add even greater flexibility, but we did not explicitly include them to preserve the simplicity of estimation given the characteristics of our samples. Still, the growth-rate form of the model that we discuss in Section 4.3.5 effectively incorporates such interactions.

[8] See Foresi and Peracchi (1995), and Chernozhukov, Fernández-Val, and Melly (2013).

Measuring dispersion, skewness, and persistence. The coefficients of the splines and polynomials in equation (4.9) may not have a straightforward or meaningful interpretation on their own. Instead, we use them to compute quantile-based measures of dispersion, skewness and persistence, which characterize some of the properties of the nonlinear models of interest. To do so, we need to calculate the implied quantiles from our conditional *cdf* model. Let $q_{it}(\tau)$ be the τ quantile from the model for some $\tau \in (0,1)$, which is the value of r that solves the equation

$$g\left(r, y_{it}, x_{it}, \alpha_i\right) = \text{logit}\left(\tau\right). \tag{4.10}$$

For example, for the linear autoregressive model in equation (4.6), the conditional quantile is defined as

$$q_{it}\left(\tau\right) = \rho y_{it} + \alpha_i + \sigma\,\text{logit}\left(\tau\right). \tag{4.11}$$

More generally, the solution can be found numerically using bracketing or interpolation methods.

A standard measure of dispersion is the interquantile range:

$$\text{IR}_{it}(\tau) = q_{it}\left(\tau\right) - q_{it}\left(1-\tau\right), \tag{4.12}$$

where usually $\tau = 0.75$ or $\tau = 0.90$. For example, for the linear AR(1) model we have

$$\text{IR}_{it}(\tau) = \sigma \times 2\,\text{logit}\left(\tau\right).$$

Dependence of IR on y_{it} and/or η_i indicates heteroskedasticity. Similarly, the Bowley-Kelley measure of skewness for some $\tau > 0.5$ is given by:

$$SK_{it}(\tau) = \frac{\left[q_{it}\left(\tau\right) - q_{it}\left(0.5\right)\right] - \left[q_{it}\left(0.5\right) - q_{it}\left(1-\tau\right)\right]}{q_{it}\left(\tau\right) - q_{it}\left(1-\tau\right)}. \tag{4.13}$$

Finally, in nonlinear models we use the measure of persistence proposed in Arellano et al. (2017), which is defined as :

$$\rho_{it}\left(\tau\right) = \frac{\partial q_{it}\left(\tau\right)}{\partial y_{it}}. \tag{4.14}$$

Using the chain rule, $\rho_{it}(\tau)$ can be written as a scaled derivative effect of realized income in our model for the cumulative distribution:

$$\rho_{it}\left(\tau\right) = -\frac{\partial g\left(q_{it}\left(\tau\right), y_{it}, x_{it}, \alpha_i\right)}{\partial y_{it}} \Bigg/ \frac{\partial g\left(q_{it}\left(\tau\right), y_{it}, x_{it}, \alpha_i\right)}{\partial r}. \tag{4.15}$$

For the linear AR(1) model we simply have $\rho_{it}(\tau) = \rho$. In general, the persistence of the process will depend on the position of a household in the distribution of current income, fixed effects, and the value of τ.

Note that an equation such as (4.6) relates the *realized* shock $v_{i,t+1}$ with rank $\Lambda(v_{i,t+1})$ to the *realized* outcome $y_{i,t+1}$ given $y_{i,t}$. However, we can also consider hypothetical shocks and their corresponding hypothetical outcomes. For example, we can ask what would be the $t+1$ outcome if the $t+1$ shock was one with rank $\tau \in (0,1)$. This is precisely the information provided by the conditional quantile function (4.11).

In the nonlinear generalization, the persistence measure (4.14) provides the weight of current income in the function that produces future income when a household is hit by a shock of rank τ.

In our analysis we do not rely on realized future outcomes and realized future shocks, but on the subjective conditional probability distribution of future outcomes, which allows us to speak about the impact of *potential (subjective) shocks* on potential future outcomes.

4.3.5 Implementation and estimation

We now discuss the specification of the various functions that enter the flexible income model in equation (4.9) and the estimation of the relevant parameters.

Specification

The functions $\beta(\cdot)$ and $\psi(\cdot)$ in (4.9) need to be parameterized. We first reformulate the model we are considering as a predictive distribution for income growth, since departures from linearity may be better captured for income changes than for levels. This change is immaterial for the linear model, but leads to different approximating models for nonlinear specifications.

Predictive distributions for growth rates. The elicited probabilites p_{jit}, which are noisy measures of $F_{it}\left(r_{jit}\right) = P\left(y_{i,t+1} < r_{jit} \middle| I_{it}\right)$, also measure $F_{it}^{\Delta}\left(s_{jit}\right) = P\left(\Delta y_{i,t+1} < s_{jit} \middle| I_{it}\right)$ for $s_{jit} = r_{jit} - y_{it}$. This is so because y_{it} is part of the information set. The function $F_{it}^{\Delta}\left(s_{jit}\right)$ is the predictive distribution of future income growth and is connected to $F_{it}\left(r_{jit}\right)$ by a simple translation of its argument: $F_{it}(r) = F_{it}^{\Delta}(r - y_{it})$. Thus, in a non-parametric sense, there is no difference between estimating one function or the other. However, in practice it may be better to estimate flexible models for $F_{it}^{\Delta}(s_{it})$ instead of $F_{it}(r)$, even if the interest is in $F_{it}(r)$.

If the true process is a random walk, $F_{it}^{\Delta}(s)$ will be a constant *cdf*, which does not depend on y_{it}, so that modeling $F_{it}^{\Delta}(s)$ is equivalent to modeling departures from a random walk. More generally, it can be expected that standardizing the range of variation in the argument by subtracting y_{it} will help modeling. Another consideration is that in the *cdf* of $\Delta y_{i,t+1}$, the nonlinearities considered in Arellano et al. (2017) are close to tail departures from linearity in a single-index logit or probit, but require translations of the index in the *cdf* of $y_{i,t+1}$.[9]

In the linear case, the model to be estimated remains essentially unchanged by targeting log income changes instead of log levels.[10] However, for nonlinear specifications, the actual flexible model to be esti-

[9] This observation can be made precise using the simple switching income process with nonlinear persistence in Arellano et al. (2017, equations (S6) and (S7)), where the predictive probit of income growth for a household around median income is a straight line independent of income. For a low (high) income household, the line jumps upwards (downwards) at right (left) tail values of income changes, but remains a straight line for most of the range of variation. In contrast, the predictive probit of log income will change with the income level across the entire income distribution, compounded with additional nonlinear variation in the tails.

[10] The linear reparameterized equation is

$$\ell_{jit} = \beta_0^{\dagger} s_{jit} + \beta_1^{\dagger} y_{it} + \eta_i + \varepsilon_{jit},$$

where $\beta_0^{\dagger} = \beta_0, \beta_1^{\dagger} = \beta_1 + \beta_0 = (1 - \rho) / \sigma$ and $s_{jit} = r_j - y_{it}$.

mated will be different:

$$\ell_{jit} = \beta_0^\dagger(s_{jit}) + \beta_1^\dagger(s_{jit})\psi(y_{it}) + \beta_2^\dagger(s_{jit})\eta_i + \varepsilon_{jit}. \tag{4.16}$$

For a given level of complexity, the functions $\beta_k^\dagger(s_{jit})$ may be better approximators to a class of models of interest than $\beta_k(r_{jit})$.

Implementation. Our empirical specification will be based on (4.16), taking the components of $\psi(\cdot)$ in a polynomial basis of functions. We specified $\psi(\cdot)$ as a vector of low-order Hermite polynomials in standardized current income. The functional coefficients $\beta_k^\dagger(\cdot)$ are taken as natural cubic splines on s_{jit}, also entering in standardized form in those functions. In general, fitting a natural cubic spline with $L \geq 2$ knots requires estimating L parameters. Further details are provided in Appendix C.2.

Note that a flexible specification of the $\beta_k^\dagger(\cdot)$ coefficients can undo the possible restrictiveness of the logistic transformation that we use. For example, if the income process is a random walk with non-logistic shocks, for a sufficiently flexible specification of the intercept term $\beta_0^\dagger(\cdot)$, the formulation $P\left(\Delta y_{i,t+1} < s_{jit} \middle| I_{it}\right) = \Lambda(\beta_0^\dagger(s_{jit}))$ will capture a broad class of *cdf*s regardless of $\Lambda(\cdot)$.

Estimation

In specifications with $\beta_2^\dagger(s_{jit}) = 1$, the model is a *static* fixed effects regression that can be consistently estimated using the within-group estimator. Our estimation approach allows for the introduction of a ridge penalty $\lambda > 0$ on the higher-order coefficients of the spline to control overfitting in the more flexible specifications, although the results reported in the chapter set $\lambda = 0$.

Rearrangement. Since monotonicity of g is not imposed, the estimated curve may be non-monotone. To address this issue we follow the method proposed in Chernozhukov, Fernández-Val, and Galichon (2010), which consists in sorting the original estimated curve into a monotone rearranged curve.

Substantial violations of monotonicity may signal misspecification. When using flexible specifications in growth-rate form, the rearranged and non-rearranged estimated probability distribution functions that we obtain are virtually identical.

Estimating models with interacted fixed effects. In specifications where $\beta_2^\dagger(s_{jit})$ depends on unknown parameters, there is an incidental parameters problem in the fixed-effects approach. Specifically, the least-squares estimator of the model's common parameters based on their joint estimation with $(\eta_1, \ldots, \eta_n)$ suffers from an errors-in-variables bias and is not consistent in a short panel. The difficulty owes to the fact that $\hat{\eta}_i$, which is used as a regressor, is a noisy estimator of η_i.

To obtain consistent estimates that take into account small-T errors in the estimated fixed effects, one can resort to either method-of-moments or pseudo maximum-likelihood approaches. A method-of-moments approach to estimating a random coefficients model for panel data is developed in Chamberlain (1992). Here we use a simple extension of the linear model in (4.7) to illustrate how to construct a linear instrumental-variable estimator of the kind that we employ in obtaining our empirical results, and defer to Appendix C.2 a more general discussion of the estimation of models with interacted effects.

A simple instrumental-variable estimator. Let us consider the model

$$\ell_{jit} = \beta_0 r_{jit} + \beta_1 y_{it} + \left(1 + \beta_2 r_{jit}\right) \eta_i + \varepsilon_{jit}, \tag{4.17}$$

which boils down to the standard linear income process when $\beta_2 = 0$. However, solving for the conditional quantile function, we can see that this model corresponds to a very different income process with heterogeneous risk and persistence that generalizes equation (4.6) to

$$y_{i,t+1} = -\frac{\eta_i}{\beta_0 + \beta_2 \eta_i} - \frac{\beta_1}{\beta_0 + \beta_2 \eta_i} y_{it} + \frac{1}{\beta_0 + \beta_2 \eta_i} v_{i,t+1}.$$

To get an estimating equation, first note that taking deviations from individual means does not remove unobserved heterogeneity from model (4.17):

$$\tilde{\ell}_{jit} = \beta_0 \tilde{r}_{jit} + \beta_1 \tilde{y}_{it} + \beta_2 \tilde{r}_{jit} \eta_i + \tilde{\varepsilon}_{jit}, \tag{4.18}$$

where $\tilde{\ell}_{jit} = \ell_{jit} - \bar{\ell}_i$, $\tilde{r}_{jit} = r_{jit} - \bar{r}_i$, and so on. However, we can use the transformation

$$r_{jit}\bar{\ell}_i - \bar{r}_i \ell_{jit} = \beta_1 \left(r_{jit}\bar{y}_i - \bar{r}_i y_{it}\right) + \tilde{r}_{jit}\eta_i + \left(r_{jit}\bar{\varepsilon}_i - \bar{r}_i \varepsilon_{jit}\right)$$

to substitute out $\tilde{r}_{jit}\eta_i$ in (4.18) and obtain

$$\tilde{\ell}_{jit} = \beta_0 \tilde{r}_{jit} + \beta_1 \tilde{y}_{it} + \gamma \left(r_{jit}\bar{y}_i - \bar{r}_i y_{it}\right) + \beta_2 \left(r_{jit}\bar{\ell}_i - \bar{r}_i \ell_{jit}\right) + \xi_{jit}, \tag{4.19}$$

where $\xi_{jit} = \left(1 + \bar{r}_i\right) \varepsilon_{jit} - (1 + r_{jit})\bar{\varepsilon}_i$ and $\gamma = -\beta_1\beta_2$. Whereas the error term ξ_{jit} is mean independent of r_{jit} and y_{it} for all (t, j) (which we collect into w_i), $r_{jit}\bar{\ell}_i - \bar{r}_i\ell_{jit}$ is an endogenous variable in equation (4.19). To motivate an IV estimator, note that

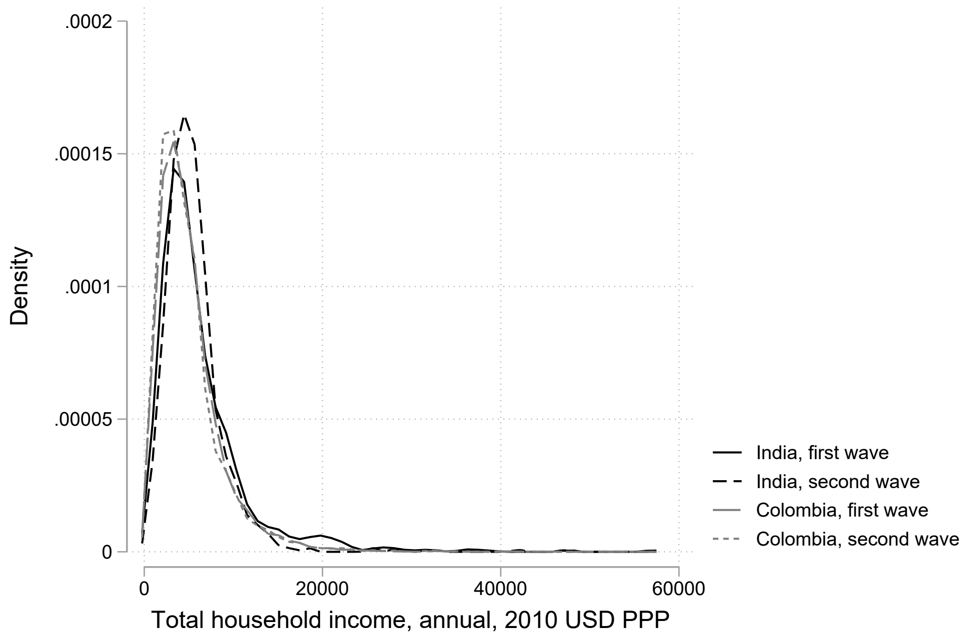
$$E\left[r_{jit}\bar{\ell}_i - \bar{r}_i \ell_{jit} \middle| w_i\right] = \beta_1 \left(r_{jit}\bar{y}_i - \bar{r} y_{it}\right) + \tilde{r}_{jit} E\left[\eta_i \middle| w_i\right].$$

Approximating $E\left[\eta_i \middle| w_i\right]$ by the projection of η_i on $\bar{w}_i$ suggests using $\tilde{r}_{jit}\bar{r}_i$ and $\tilde{r}_{jit}\bar{y}_i$ as external instruments for $r_{jit}\bar{\ell}_i - \bar{r}_i\ell_{jit}$ and estimating equation (4.19) by two-stage least squares (TSLS). The restriction $\gamma = -\beta_1\beta_2$ is not required for identification and might be ignored to avoid nonlinear estimation.

4.4 Data

We use data on subjective income expectations in combination with data on realized income from two developing country contexts — rural India and Colombia. In both cases, the subjective income expectations were collected as part of broad surveys aimed at evaluating development interventions. Both interventions were targeted at a poor and rural population.

In Figure 4.2, we plot the distribution of reported total household income across countries and survey waves. Income is measured in 2010 PPP USD, which we use for both countries throughout the analysis. While the contexts we are studying are very distinct, the two distributions are remarkably similar, indicating that we are concerned with comparably poor populations. Average annual household income in the sample is $5,924 for India and $5,013 for Colombia, with a standard deviation of $4,632 and $3,759, respectively.



Note. The Figure shows the distribution of total household income in the two study populations, in 2010 PPP USD. Monthly income in Colombia is annualized for comparability.

FIGURE 4.2. Household income across study populations.

In what follows, we briefly describe the survey contexts and the characteristics of the respondents and their households. We then provide some evidence about the validity of the expectations data. In both surveys, subjective expectations data were elicited using the approach described in Section 4.2. The main difference between the two surveys is in the horizon of future income: in India future income refers to the following year, while in Colombia is the following month.

4.4.1 India

The data in India were collected in 64 villages in Anantapur, a district located in the southern state of Andhra Pradesh, for the evaluation of a microfinance intervention (loans for cow or buffalo); see Augsburg (2009) for additional details. A typical household we consider has five members and a male, 45 years-old household head. Most households belong to the "Other Backward caste", a collective term used by the Government of India to classify castes which are educationally or socially disadvantaged. A further 13% belong to the Scheduled Castes, 5% to the Scheduled Tribes, and the remaining 28% to the General Caste. More than 60% of household heads had not undergone any formal education, and only 10% had some primary education. The average household depends on three income sources, with agriculture being the primary activity — as farmers (25%) or as agricultural labourers (64%). Additional details can be found in Attanasio and Augsburg (2016, Table 1).

In Table 4.1, we present descriptive information on income sources and shocks, which we later integrate in our models and which provide contextual information on the importance of different sources of risk that

TABLE 4.1. India – income shocks and sources

	≤2 sources	3 sources	4+ sources
Proportion	0.423	0.372	0.205
0% farm	0.347	0.155	0.054
Up to 50% farm	0.235	0.431	0.502
More than 50% farm	0.419	0.414	0.444
	1.00	1.00	1.00
No shocks	0.113	0.087	0.092
Health shock	0.233	0.194	0.200
Agriculture shock	0.531	0.606	0.603
Other shocks	0.123	0.113	0.105
	1.00	1.00	1.00

Note. The table shows relative frequencies (proportions) of different components of total household income for the Indian data, pooling across the two waves. The first row displays the proportion of households reporting up to two different income sources, three income sources, and more, respectively. The next three rows report the proportion of current income stemming from farm-related activities for each subgroup by income source. The final four rows show the relative frequency of the most important types of (negative) shocks faced by households during the previous year, again for each income source subgroup.

households face. Households with less than three, three, and four or more income sources account for 42.3%, 37.2%, and 20.5% of the sample, respectively. For each such category, we report the percentage of income from farm-related activities (which includes agriculture) and the types of shocks experienced. Households with at most two income sources are relatively more likely to report no income from farm-related activities. Moreover, the likelihood of reporting no shocks is about 10%, health shocks about 20%, and agricultural shocks about 50-60%, quite uniformly across household categories.

After the household baseline sample was interviewed in January/February 2008, a follow-up survey was conducted in April/June 2009. Respondents were asked to provide information on income and subjective expectations in both survey rounds. Of the 1,036 households that made the original sample, 947 were re-interviewed in the second wave. We drop observations with missing income or at least one reported probability and those with elicited expectations that violate basic probability laws, following the analysis in Attanasio and Augsburg (2016). This yields a balanced panel with $N = 770$ households. Details are reported in Table C.1.1 in Appendix C.1.1.[11]

About a quarter of these households were clients of a microfinance institution (MFI) and had in 2008 loans provided livestock investment. The remaining households were either residing in the same villages or in villages the MFI considered targeting in the future. The data was collected to evaluate the provision of these livestock loans, which aimed at enabling households to engage in milk-selling as an additional income-generating activity, thereby reducing their dependence on outcomes of the main cropping seasons.

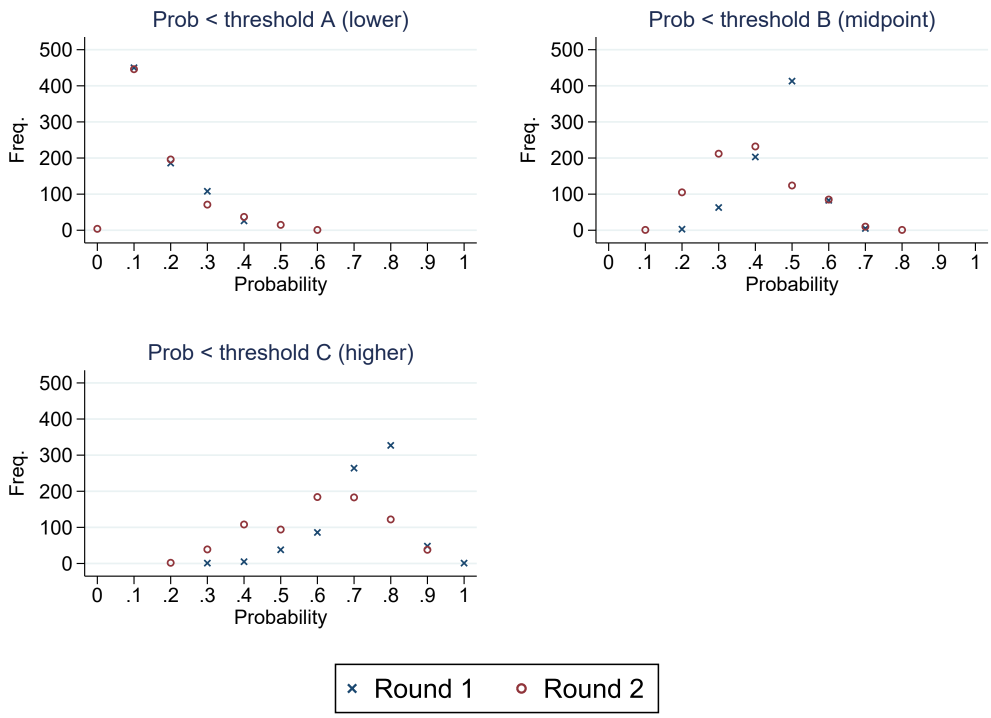
In both survey waves, the interviewers — who visited the respondents in their homes — elicited information on points on the respondents' subjective household income distribution. The technique discussed in Section 4.2 was used after explaining the approach in detail, practicing with rainfall questions, and using a ruler as a visual aid. Respondents were asked about their expected household income for the year following the interview. This interval was chosen considering the irregularity of income and to ensure key income

[11] This attrition rate is slightly higher than that reported in Attanasio and Augsburg (2016). As documented in the Appendix, these differences are mostly due to removing outliers in reported and expected income.

periods were covered.

As discussed in detail in Attanasio and Augsburg (2016), respondents were not only willing to provide (expected) income information, but also provided sensible answers that reflected their beliefs. In particular, (i) over 97% of respondents provided responses to all three thresholds, (ii) violations of basic probability laws (monotonicity and wrong "direction") make up less than 1% of the sample, (iii) very few households bunch at 100% for the highest threshold or 0% for the lowest, indicating that the minimum and maximum expected income are well elicited, (iv) respondents made otherwise use of the entire range, although some bunching at multiple of 5s was observed (possibly because these were indicated on the ruler), and (v) expectations correlate sensibly with household characteristics. We report further details in Appendix C.1.1.

FIGURE 4.3. India: frequencies by threshold

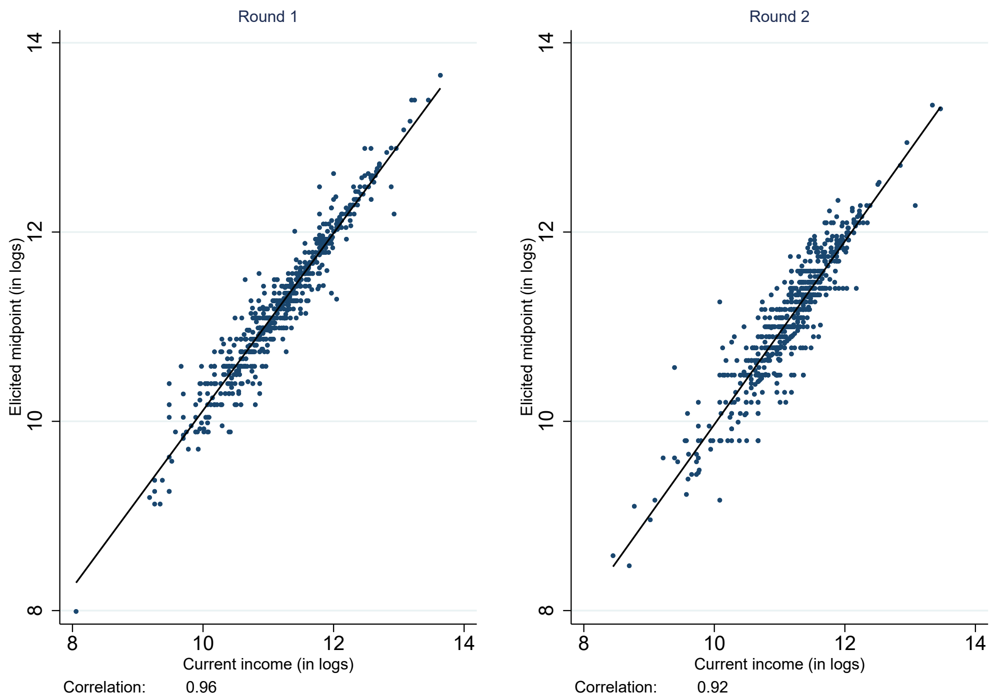


Note. The figure displays subjective probability frequency plots for each threshold and survey round, as shown in the legend. Probabilities are rounded to the nearest tenth.

Figure 4.3 provides a summary of the distribution of elicited probabilities by threshold and survey round. The upper left panel displays the absolute frequencies for reported cumulative probabilities below the first threshold, which as expected are skewed to the right — the mode being at 0.1 in both survey rounds. On a similar vein, the distribution is mostly concentrated within the 0.3-0.6 range for the midpoint threshold, and skewed to the left for the highest threshold.

Figure 4.4 plots together the elicited midpoint of the predictive distribution and realized current income, and shows an extremely high cross-sectional correlation between household's current income and their median subjective assessment for next year's income. This evidence suggests that the subjective income expectations data provide useful and meaningful information.

FIGURE 4.4. India: current income and reported midpoint



Note. The solid black corresponds to the linear regression fit.

4.4.2 Colombia

The data in Colombia were collected in 122 of the country's poorest municipalities located in 26 of 34 departments to evaluate the introduction of a Conditional Cash Transfer (CCT) program, called *Familias en Acción* (FEA), a welfare program run by the Colombian government to foster the accumulation of human capital through improved nutrition, health, and education in rural Colombia. As many CCTs around the world, FEA pursued its objective through a cash transfer conditional on child vaccinations, development checks, school attendance, and courses for the mother. The program was targeted to the poorest sectors of society; recipients typically fall into the bottom 20% of Colombian households living in rural areas.[12]

The evaluation first conducted a baseline survey in 2002, approaching 11,500 and interviewing 11,462 households. We use data from the two follow-up survey rounds, conducted from July to November 2003 and again from November 2005 to March 2006, completing interviews to 10,743 and 9,463 households, respectively.[13] At the time of the second survey, about half of respondent households were target beneficiaries of *Familias en Acción*.

Survey respondents are predominantly female (65%). Just over half (54%) are household heads, living in households that, on average, included another five members, with an average age of 43 years and with low levels of education: 24% of household heads have less than primary education, with an average of 3.5 years

[12]In particular, recipients (or potential recipients, in the case of the evaluation sample) were in the lowest category of the SISBEN indicator, which is used to target most social programs and to set utility prices. See Attanasio, Battistin, and Mesnard (2012, Section 2).

[13]These figures correspond to the first row in Table C.1.3 in Appendix C.1.2, which also provides additional information on response rates and sample sizes.

TABLE 4.2. Colombia – income shocks and providers

	1 earner	2 earners	3+ earners
Proportion	0.380	0.413	0.207
Up to 75% regular	0.145	0.269	0.223
More than 75% regular	0.080	0.359	0.434
100% regular	0.775	0.372	0.344
	1.00	1.00	1.00
No shocks	0.773	0.749	0.712
Health shock	0.089	0.093	0.124
Other shocks	0.138	0.158	0.163
	1.00	1.00	1.00

Note. The table shows relative frequencies (proportions) of different components of total household income for the Colombian data, pooling across the two waves. The first row displays the proportion of households with one earner, two earners or three or more earners (during the previous month). The next three rows report the proportion of current income stemming from regular sources (as opposed to occasional) for each subgroup by number of earners. Note that "more than 75% regular" excludes 75% and 100%. The final three rows show the relative frequency of the most important types of (negative) shocks faced by households during the previous year, again for each category of number of earners. Note that around 2.5% of households report having suffered both health- and non health-related shocks, which implies that the absolute frequencies for these two categories are slightly larger than reported above, for each category of number of earners.

of schooling. Income is predominantly earned in the form of labour income, where most individuals tend to be informally employed (93%). About half of the working individuals in the sample work in agriculture, while others, for example, work as domestic servants. The survey design and context are described in detail in Attanasio et al. (2012).[14]

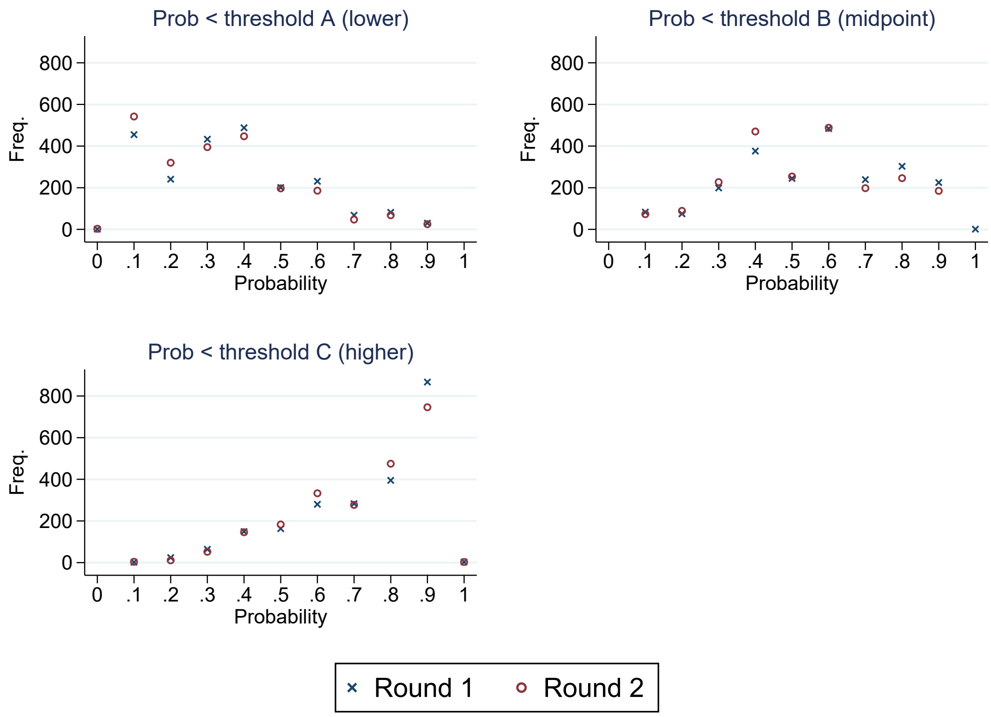
In Table 4.2, as for the Indian data, we provide descriptive statistics of time-varying characteristics related to income sources and shocks. We divide households into three categories according to the number of household members who report a source of income during the previous month: one, two or three or more members, which account for 38%, 41.3% and 20.7% of the sample, respectively. We also report the proportion of income that comes from regular sources, defined as the share of labor and non-labor income in total household income, excluding occasional labour, monthly CCT subsidies (if any) and transfers. Households with only one working member tend to receive most of their income from regular sources (77.5%), while those with three or more providers are the least likely to do so (34.4%). Income shocks are evenly distributed across these earner categories, similar to the pattern observed in India.

Elicitation of expectations was conducted in a similar fashion to the survey in India; see again Section 4.2 for a description of the elicitation approach. Figure 4.5 summarizes the distribution of elicited probabilities by threshold and survey round. Figure 4.6 displays a high positive correlation between the midpoint of the reported probability distribution and current income in both survey rounds. This relationship is somewhat weaker than in the Indian context, in line with our results on risk and persistence and the fact that expectations here refer to a much shorter time span.

Since the subjective expectations data have not been used before, we provide a detailed analysis and validation in Appendix C.1.2. Overall, the elicitation of subjective expectations was less precise in the Colombian data. We find a substantially larger degree of logical response errors, although still within reasonable

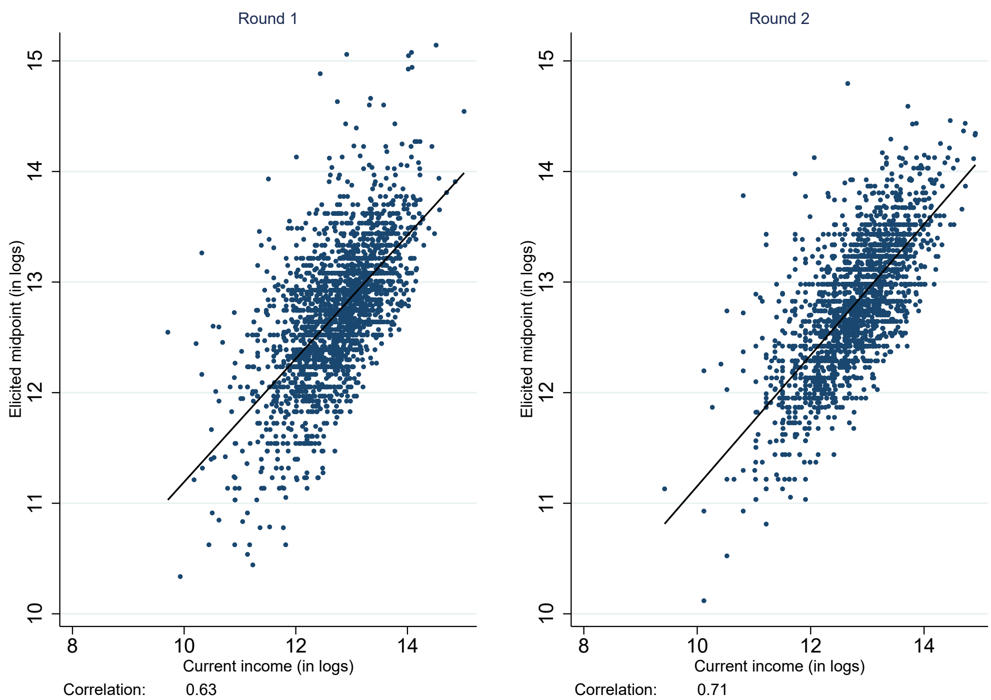
[14]The baseline evaluation report can be accessed at https://ifs.org.uk/publications/baseline-report-evaluation-familias-en-accion.

FIGURE 4.5. Colombia: frequencies by threshold



Note. The figure displays subjective probability frequency plots for each threshold and survey round, as shown in the legend. Probabilities are rounded to the nearest tenth.

FIGURE 4.6. Colombia: current income and reported midpoint



Note. The solid black corresponds to the linear regression fit.

ranges (for instance, around 4% of households report distributions that violate monotonicity). Validation and sample selection leave us with a significantly reduced balanced panel sample of $N = 2,230$ households.

A detailed step-by-step analysis is reported in Appendix C.1.2. Tables C.1.5 and C.1.6 show that these decisions do not imply strong sample selection (at least, based on observable characteristics). Households in the final sample tend to have fewer adults, and household heads are slightly younger (by around a year, on average). They are also slightly less likely to have experienced health or other types of shocks, but are generally very comparable in terms of household composition, income, income sources and education level.

4.5 Results

In this section, we report the results we obtain for both countries when estimating various specifications of the income process. We start with a linear AR(1) process with time effects and consider versions with and without fixed effects, and then augment it with a number of state variables. When these are interacted with current income, persistence is allowed to be heterogeneous in the cross-section. Finally, we consider nonlinear processes of the type introduced in section 4.3.4. In such models, all the features of the distribution vary across units and over time.

4.5.1 Linear models

In what follows, we present the estimates of the parameters of the linear process (4.6), which are obtained from least-squares estimates of the reduced-form parameters in equation (4.7), using the one-to-one mapping between the two sets of parameters.

India

We begin by reporting estimates using the Indian data in Table 4.3. The first column contains the estimates of the model without fixed effects, while the second contains the estimates of the model with fixed effects. In addition to the parameters of the model (the persistence parameter ρ, the standard deviation of innovations σ, the residual variance σ_ε^2 and, in the case of the fixed-effect model, the variance of the individual and village-level fixed effects σ_η^2 and $\sigma_{\eta,\text{village}}^2$), for comparability with some of the results we report below, we also include the differences between the 75th and 25th quantiles and between the 90th and 10th quantiles implied by these estimates.

In the model without fixed effects, ρ is close to one and estimated very precisely. The standard deviation of the innovation to the (log) income process is substantive at 0.56, reflected in large values of the interquantile ranges reported. This is a measure of risk, which is identified from the association between the self-reported range of variation of future income and elicited probabilities. The residual variance is estimated at 1.24, which is sizable.

The introduction of fixed effects reduces the degree of persistence from 0.97 to 0.93, which is now significantly different from unity. Fixed effects also play an important role in assessing risk, as the standard error of the income process innovations is reduced from 0.56 to 0.31. The variance of the individual fixed effect at 0.22 (measured as η_i) is one and a half times the variance of the village level fixed effect. These results are surprisingly comparable to those used in standard macro calibrations of the income process based on realized earnings (see, for example, Kaplan and Violante (2010) or Alvarez and Arellano (2022)).

The residual variance is somewhat smaller after fixed effects are introduced, but remains substantial. In particular, it is too large for the residual to be interpreted solely as measurement error in elicited probabilities. This impression is reinforced by the fact that the residual variance in a regression of ℓ_{jit} on r_{jit} with period and unit specific effects is 0.37. Such calculation can be regarded as a lower bound for the measurement error component of the residual in a separable model and implies that a subjective probability of 0.5 would be elicited in the survey with a standard error of one percentage point. The likely presence of additional sources of residual variation provides further motivation for examining the roles of other state variables, nonlinearities, and neglected heterogeneity. However, a variance decomposition with period and unit effects is only suggestive because it preserves the separability between r_{jit} and state variables, while our flexible models emphasize the interactions between the two.

	No FE	FE
ρ	0.97	0.93
	(0.94, 1.00)	(0.90, 0.96)
σ	0.56	0.31
	(0.51, 0.60)	(0.29, 0.33)
$IQR_{0.75}$	1.22	0.69
	(1.13, 1.33)	(0.64, 0.74)
$IQR_{0.90}$	2.44	1.38
	(2.25, 2.65)	(1.29, 1.47)
σ_η^2		0.22
		(0.18, 0.27)
σ_η^2 village		0.14
		(0.14, 0.19)
σ_ε^2	1.24	1.14
	(1.21, 1.27)	(1.10, 1.18)

Note. The table reports results for the linear model in (4.7) using the data for India, without fixed effects (and a common intercept) and with fixed effects. We also include year (survey round) dummies in both cases. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.3. India — linear model

Colombia

In Table 4.4, we report the results obtained by estimating the same two versions of the linear model (with and without fixed effects) reported in Table 4.3 for India. In the model without fixed effects, the parameter ρ is estimated to be 0.71. Remarkably, the standard deviation of income innovations is very large at 0.98. It should be remembered that in the Colombian data, future income refers to *next month* rather than *next year*. Annual income is likely to be less volatile than monthly income, although the two economies might be very different. The residual variance is somewhat larger in the Colombian data than in the Indian data, although of comparable magnitude, so the previous comments about the size of the residuals apply here as well.

When adding fixed effects, the estimated ρ is much smaller at 0.5 and the standard deviation of innovations is reduced from 0.98 to 0.65. Moreover, the variance of the individual fixed effect is much larger than in the Indian sample and, similar to India, much larger than the variance of the village component of the

	No FE	FE
ρ	0.71	0.50
	(0.67, 0.74)	(0.46, 0.55)
σ	0.98	0.65
	(0.93, 1.03)	(0.63, 0.67)
$IQR_{0.75}$	2.16	1.43
	(2.05, 2.26)	(1.38, 1.48)
$IQR_{0.90}$	4.31	2.86
	(4.10, 4.52)	(2.75, 2.96)
σ_η^2		0.48
		(0.44, 0.52)
σ_η^2 village		0.12
		(0.12, 0.17)
σ_ε^2	1.46	1.09
	(1.42, 1.49)	(1.05, 1.12)

Note. The table reports results for the linear model in (4.7) using the data for Colombia, without fixed effects (and a common intercept) and with fixed effects. We also include year (survey round) and month (interview) dummies in both cases. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.4. Colombia — linear model

fixed effect (four times). Taken together, these results suggest higher risks and lower persistence in Colombian monthly earnings compared to Indian annual earnings, and greater unobserved heterogeneity in the Colombian data.

4.5.2 Linear models with additional state variables

We now augment the linear model with time-varying characteristics x_{it}, along the lines of Subsection 4.3.3 and equation (4.8). When interacted with current income, we allow for differential subjective persistence along these characteristics. We include sources of income and shocks experienced in the current year as dimensions of x_{it}, as described in Tables 4.1 and 4.2 for India and Colombia, respectively. We only report specifications that include fixed effects.

India

The results obtained estimating equation (4.8) on the India data are reported in Table 4.5 when introducing indicators of type and number of income sources and in Table 4.6 when interacting the number of sources with types of shocks.

These results show that persistence is not greatly affected by the presence of additional variables, even when interacted with current income. Households with different sources of income, and households who have in the past year experienced either a health, agricultural or other shock[15] all have subjective persistence around the 0.90 mark. Having said that, however, households with no farm activities have lower levels of

[15] "Sources" in the India data refer to the number of activities that the household generates income from (farming, agricultural labour, relief work, crafts, trading etc.), health shocks refer to illness or death of a household member, agricultural shocks include crop failure due to disease or floods, and other shocks include events such as job loss or being the victim of crime.

persistence.

Remarkably, the introduction of this additional variables does not affect much the variability of the income innovations or of the fixed effects. Similar considerations apply to the residual variance. The conclusion we draw from these tables is that, although marginally significant, the introduction of the interactions with the observable considered, it does not have a large effect on the estimated risk and persistence of the income process.

ρ	≤2 sources	3 sources	4+ sources
0% farm	0.87	0.90	0.83
	(0.83, 0.92)	(0.85, 0.95)	(0.76, 0.90)
50% farm	0.91	0.94	0.87
	(0.87, 0.95)	(0.90, 0.98)	(0.80, 0.93)
75% farm	0.93	0.96	0.89
	(0.88, 0.98)	(0.92, 1.01)	(0.82, 0.96)
σ		0.30	
		(0.28, 0.32)	
$IQR_{0.90}$		1.33	
		(1.24, 1.41)	
σ^2_η		0.23	
		(0.19, 0.28)	
σ^2_η village		0.13	
		(0.13, 0.18)	
σ^2_ε		1.12	
		(1.07, 1.16)	

Note. The table reports results for India for the linear model (4.7) in augmented with household-level characteristics, along the lines of (4.8). "Farm" refers to the proportion of current income obtained from farming-related activities; see Table 4.1 for a full description of these variables. We also include year (survey round) dummies. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.5. India — linear model augmented with household characteristics (% of income from farming)

ρ	≤2 sources	3 sources	4+ sources
No shock	0.87	0.91	0.83
	(0.79, 0.95)	(0.84, 0.99)	(0.74, 0.93)
Health	0.92	0.97	0.89
	(0.86, 0.98)	(0.91, 1.04)	(0.81, 0.97)
Agricultural	0.90	0.97	0.87
	(0.86, 0.95)	(0.92, 1.01)	(0.81, 0.94)
Other	0.99	1.04	0.97
	(0.88, 1.09)	(0.93, 1.14)	(0.84, 1.08)
σ		0.30	
		(0.28, 0.32)	
$IQR_{0.90}$		1.34	
		(1.23, 1.42)	
σ_η^2		0.25	
		(0.22, 0.32)	
σ_η^2 village		0.15	
		(0.15, 0.21)	
σ_ε^2		1.13	
		(1.08, 1.16)	

Note. The table reports results for India for the linear model in (4.7) augmented with household-level characteristics, along the lines of (4.8). See Table 4.1 for a full description of these variables. We also include year (survey round) dummies. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.6. India — linear model augmented with household characteristics (shocks and income sources)

Colombia

In the Colombian data, we note more substantial differences in persistence according to the number of sources of income than in India. We observe that for three or more working members there is higher persistence. This seems to be a case of income diversification, which makes a lot of sense for Colombia given that predictions are one-month ahead. This is particularly true for households with a regular (non-occasional) stream of income, as can be seen in Table 4.7.

It is noteworthy that the introduction of the additional controls interacted with income does not make much difference to the size of the uncertainty, which remains more or less at the same level (0.64) and to the variability of both individual and village level fixed effects. The stability of these coefficients and the limited variability of the persistence estimates are an indication of the fact that these observables play a limited role in this model.

The most noticeable difference between these results and those obtained for India is the size of the persistence coefficient. In the case of Colombia, although some variability is observed by income sources, persistence is never larger than 0.65, while in India is never below 0.83. We also notice that the idiosyncratic variability of fixed effects is considerably larger in Colombia, being almost twice as large as in India. Instead, the variability of village level fixed effects is roughly similar (around 0.12). An implication of this finding is that the variability across villages accounts for a larger fraction of the fixed effects variability in India.

ρ	1 earner	2 earners	3+ earners
0% regular	0.34	0.36	0.48
	(0.21, 0.48)	(0.24, 0.47)	(0.34, 0.63)
75% regular	0.51	0.52	0.61
	(0.43, 0.58)	(0.43, 0.59)	(0.51, 0.71)
100% regular	0.56	0.57	0.65
	(0.49, 0.63)	(0.48, 0.66)	(0.55, 0.76)
σ		0.64	
		(0.62, 0.67)	
$IQR_{0.90}$		2.83	
		(2.72, 2.92)	
σ_η^2		0.48	
		(0.45, 0.52)	
σ_η^2 village		0.11	
		(0.12, 0.16)	
σ_ε^2		1.08	
		(1.04, 1.11)	

Note. The table reports results for Colombia for the linear model in (4.7) augmented with household-level characteristics, along the lines of (4.8). See Table 4.2 for a full description of these variables. We also include year (survey round) and month (interview) dummies. The R^2 are adjusted for the presence of fixed effects. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.7. Colombia — linear model augmented with household characteristics (proportion of regular income and income sources)

ρ	1 earner	2 earners	3+ earners
No shock	0.54	0.55	0.58
	(0.47, 0.62)	(0.47, 0.63)	(0.48, 0.68)
Health	0.64	0.65	0.67
	(0.50, 0.77)	(0.51, 0.79)	(0.53, 0.81)
Other	0.44	0.46	0.48
	(0.32, 0.56)	(0.33, 0.58)	(0.34, 0.61)
σ		0.65	
		(0.62, 0.67)	
$IQR_{0.90}$		2.84	
		(2.73, 2.93)	
σ_η^2		0.48	
		(0.45, 0.53)	
σ_η^2 village		0.11	
		(0.12, 0.16)	
σ_ε^2		1.09	
		(1.04, 1.11)	

Note. The table reports results for Colombia for the linear model (4.7) in augmented with household-level characteristics, along the lines of (4.8). See Table 4.2 for a full description of these variables. We also include year (survey round) and month (interview) dummies. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.8. Colombia — linear model augmented with household characteristics (income shocks and income sources)

4.5.3 Nonlinear models

We now turn to the central empirical results of the chapter. In this section, we discuss the estimation of the nonlinear model with additive fixed effects:

$$\ell_{jit} = \beta_0^{\dagger}(s_{jit}) + \beta_1^{\dagger}(s_{jit})\psi\left(y_{it}\right) + \eta_i + \varepsilon_{jit}, \qquad (4.20)$$

which corresponds to equation (4.16) with $\beta_2^{\dagger}(\cdot) = 1$. We report results for the most parsimonious specification that would still allow for nonlinear persistence and skewness — that is, we choose $L = 3$ for $\beta_0^{\dagger}(\cdot)$ and $\beta_1^{\dagger}(\cdot)$ (cubic splines with two boundary knots and one intermediate knot) and $\psi(\cdot)$ of order 1 (current income enters in log levels, but since $s_{jit} = r_{jit} - y_{i,t}$ quadratic income terms are also involved). This configuration is in what follows referred to as the *baseline* specification for the nonlinear model.

We experimented with higher values of L and higher order polynomials. While for $L > 3$ we obtained qualitatively similar results, as long as the position of the knots was judiciously chosen, the use of higher-order $\psi(\cdot)$ terms required substantial penalization to avoid unstable results. We also experimented with nonlinear models involving additional state variables, but since the interaction terms did not contribute much, in line with what we saw for linear models, we do not present them here.

In our data, identification of nonlinear persistence comes directly from the association between current income and the shape of the distribution of subjective probabilities of future income, net of fixed effects. In particular, to obtain the results we present, it is key to consider distributional models that are flexible enough to allow for conditional skewness that may change with current income and unobserved heterogeneity.

Overall, the linear autoregressive model is soundly rejected on both the Indian and Colombian data. Moreover, similar patterns of heterogeneous risks and nonlinear persistence emerge in the two data sets. This is particularly remarkable in view of the differences between the two surveys (annual vs monthly) and the characteristics of their underlying populations.

India

Table 4.9 presents the results obtained in estimating the nonlinear model with additive effects on the Indian data. Firstly, with regard to risk, it is noticeable that *dispersion risk* decreases with current income (interquantile range measures for the 90th income percentile are two-thirds of those for the 10th percentile), while *skewness risk* increases moderately. That is, the rich have less dispersion risk but more skewness risk than the poor.

Turning to persistence, we observe the presence of nonlinear persistence, which depends on both the percentile of current income and the rank of the quantile shock to next-period's income. Persistence is close to one for high-income households throughout, but only when hit by a bad shock for low-income households. When a good shock hits a low-income household, persistence is much lower. This pattern, which is depicted in Figure 4.7, features prominently in our results and is only partially consistent with the nonlinear persistence reported in Arellano et al. (2017), who found reduced persistence not only at the bottom of the income distribution (with good shocks) but also at the top of the income distribution (with bad shocks). Those differences do not necessarily imply a contradiction between results based on subjective expectations and those based on realized incomes because the populations of reference in the two studies are very different.

	y_{p10}	y_{p50}	y_{p90}
$IQR_{0.75}$	0.79	0.60	0.52
	(0.72, 0.90)	(0.54, 0.63)	(0.45, 0.55)
$IQR_{0.90}$	1.61	1.23	1.05
	(1.48, 1.85)	(1.15, 1.31)	(0.93, 1.14)
$SK_{0.90}$	−0.02	−0.10	−0.14
	(−0.15, 0.06)	(−0.20, −0.04)	(−0.28, −0.04)
$\rho_{\tau 0.25}$	0.96	1.02	1.05
	(0.92, 1.05)	(0.99, 1.06)	(1.00, 1.08)
$\rho_{\tau 0.50}$	0.79	0.96	1.01
	(0.72, 0.86)	(0.92, 0.98)	(0.97, 1.03)
$\rho_{\tau 0.75}$	0.58	0.86	0.97
	(0.38, 0.71)	(0.82, 0.89)	(0.91, 0.99)
σ^2_η		0.23	
		(0.19, 0.28)	
σ^2_η village		0.14	
		(0.13, 0.19)	
σ^2_ε		1.11	
		(1.06, 1.14)	

Note. The table reports results for India for the flexible model with additive fixed effects in (4.20). We also include year (survey round) dummies. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.9. India — flexible model (additive fixed effects)

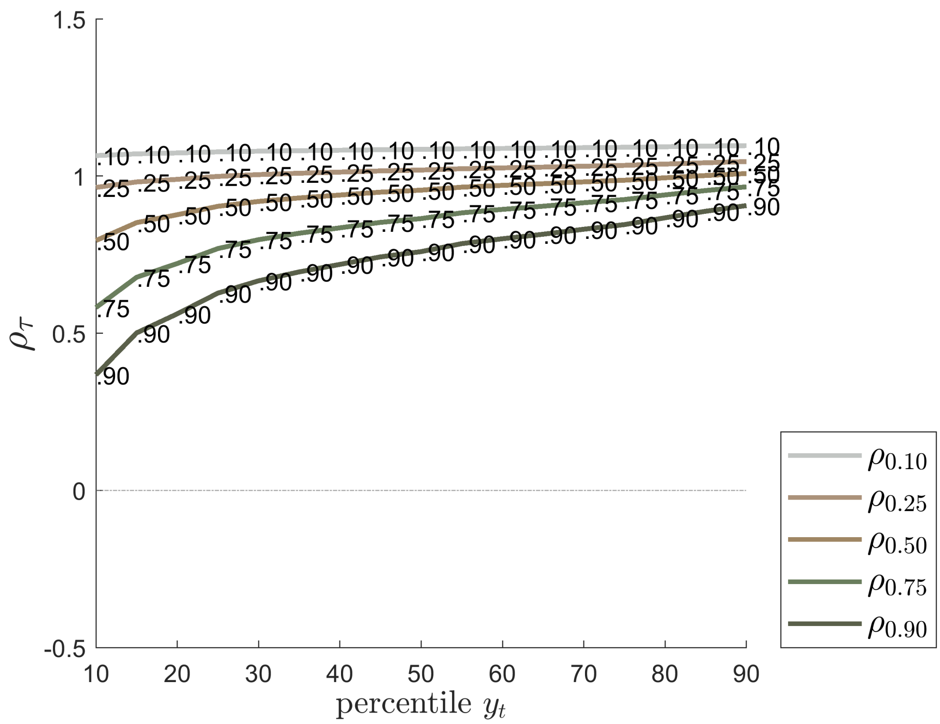
In our developing country databases all households are poor by comparison to PSID households.

An economic implication of the nonlinear income process comes from the fact that a positive shock for lower-income households reduces the persistence of the past and is therefore beneficial for those households in terms of expected future income. Relative to the predictions of a linear income process, this asymmetry will induce lower saving and higher consumption at younger ages for self-insured low-income households. However, for higher-income households, given the estimated process, the opposite effect (associated to negative shocks) would not be expected to happen.

The nonlinear persistence that we find among the poorest households is consistent with a poverty trap interpretation. When income is too low it is difficult to escape poverty, but a large positive shock can weaken the weight of the past history and get the household (persistently) off the hook at a higher income level.[16] We find it quite interesting that this kind of poverty trap dynamics seems to be reflected in the subjective income expectations of poor households.

All these summary measures are computed for the model's probability distributions evaluated at the median value of the fixed effects. We extend the analysis to other percentiles (corresponding to a normal distribution with the estimated variance of the fixed effects) in Figure C.3.1 (see Appendix C.3) and obtain a similar pattern of nonlinear persistence together with an additional pattern of unobserved heterogeneity. Specifically, as the selected percentile of the effects increases, overall persistence increases and the amount of

[16] See Banerjee, Duflo, Goldberg, Karlan, Osei, Parienté, Shapiro, Thuysbaert, and Udry (2015) for evidence on how a multi-faceted program can help the extreme poor to persistently increase their income; and Genicot and Ray (2017) for an aspirations-based theory of poverty traps and references to the earlier theoretical literature.



Note. The figure reports estimates of nonlinear persistence for India for the flexible model with additive fixed effects in (4.20). Specifications also include year (survey round) dummies. See Figure C.3.3 for pointwise confidence bands.

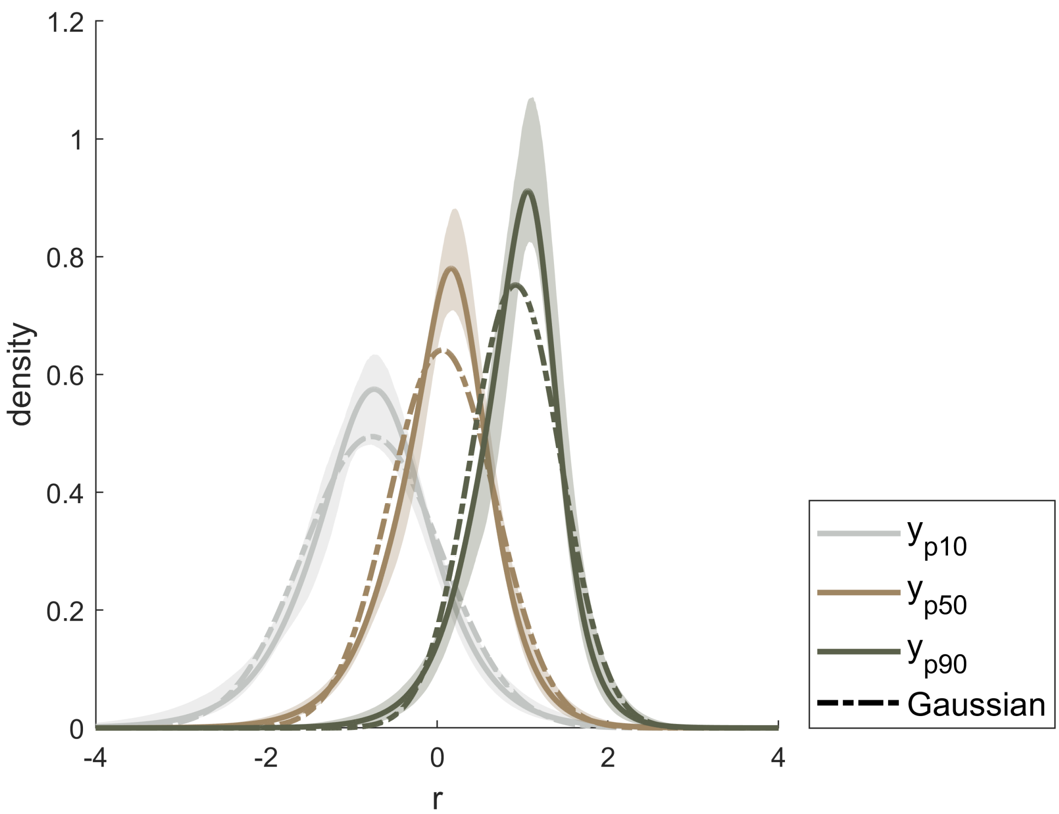
FIGURE 4.7. India — flexible model, nonlinear persistence (additive fixed effects)

persistence for different shocks becomes more compressed, with a flatter gradient along current income for large shocks.

Probability density functions for India. Figure 4.8 shows estimated conditional probability density functions (*pdf* s) at the 10th, 50th and 90th percentiles of current income for the baseline model. They are calculated by numerical differentiation of the corresponding estimated cumulative probabilities. Estimated *cdf* s with or without rearrangement coincide since there are no instances of non-monotonicities in the baseline specification. Figure 4.8 also shows block-bootstrap point-wise confidence bands and Normal *pdf* s with the same empirical mean and variance for comparisons.

As expected, the predictive subjective density for poorer households is shifted to the left relative to that of richer households, indicating that at any given reference level for future income, poorer households tend to assign a higher probability to their future income falling below that level. Moreover, consistent with the results in Table 4.9, subjective predictive densities tend to be more symmetric and are noticeably more dispersed for the poor. Beyond non-normality, the figure also portrays more pronounced differences between the current rich and the current poor in terms of relatively bad and relatively good outcomes. These differences are suggestive of nonlinear persistence,[17] which is more relevant for the currently poor, consistently with the patterns we find in Table 4.9.

[17] Recall that, according to the chain rule in equation (4.15), nonlinear persistence can be obtained as the derivative of the *cdf* with respect to y relative to the derivative of the *cdf* with respect to r.



Note. The figure shows estimated *pdf* s at the 10th, 50th and 90th percentiles of current income for India, calculated by numerical differentiation on the estimated (conditional) cumulative probabilities. Shared areas are 90% pointwise confidence bands using block bootstrap (1000 repetitions). Dotted lines correspond to Normal *pdf* s with the same mean and variance as the empirical *pdfs* of the same color. The range of variation is standardized (future) log income, see Appendix C.2.4.

FIGURE 4.8. India — flexible model, probability density function (additive fixed effects)

Colombia

Table 4.10 presents the results for the nonlinear model with additive effects on the Colombian data. Similar to India, we observe dispersion and skewness decreasing with current income (decreasing dispersion risk and increasing skewness risk). However, while in India we found no skewness at the bottom of the income distribution and negative skewness at the top, in Colombia we find positive skewness at the bottom and no skewness at the top.

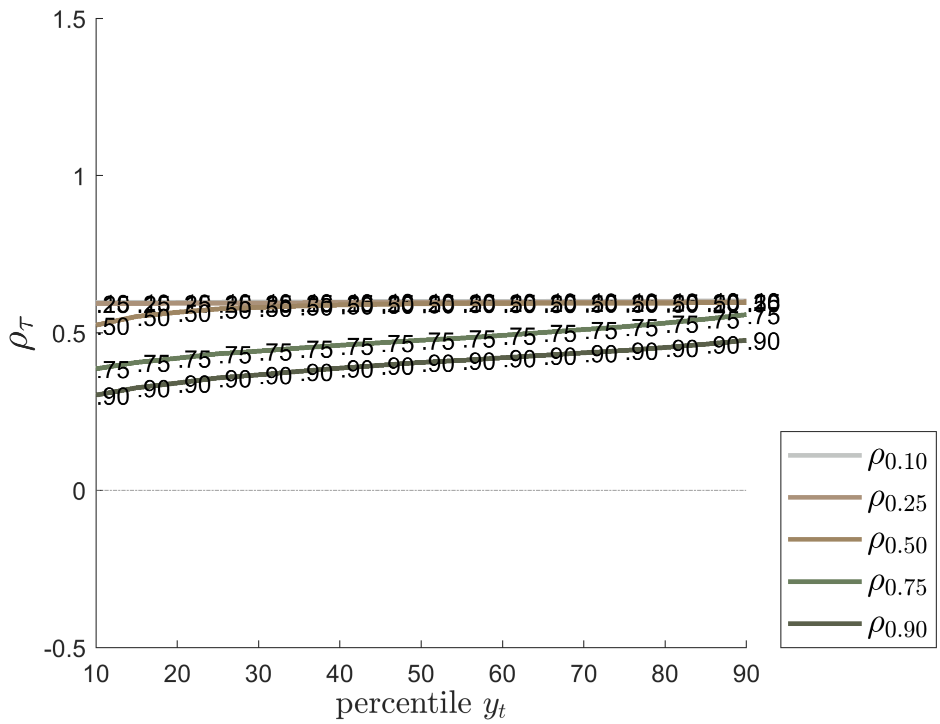
Regarding persistence, although at lower levels than in India (similar to the linear model), we find the same pattern of nonlinearities, with persistence decreasing with relatively good shocks for low income households but not for high income households (Table 4.10 and Figure 4.9). The impact of unobserved heterogeneity on persistence is also similar to the one for India; see Figure C.3.2 in Appendix C.3.

Probability density functions for Colombia. We also observe marked departures from normality in Colombia according to Figure 4.10, which depicts estimated *pdf* s together with Normal distribution fits. The estimated densities are consistent with the pattern in Table 4.10 of decreasing dispersion risk as we move along the income gradient and feature prominent deviations from normality, more so for poorer households.

	y_{p10}	y_{p50}	y_{p90}
$IQR_{0.75}$	1.48	1.34	1.28
	(1.39, 1.58)	(1.26, 1.40)	(1.21, 1.36)
$IQR_{0.90}$	2.97	2.75	2.63
	(2.80, 3.13)	(2.63, 2.86)	(2.50, 2.77)
$SK_{0.90}$	0.13	0.08	0.04
	(0.05, 0.19)	(0.03, 0.12)	(−0.01, 0.08)
$\rho_{\tau 0.25}$	0.59	0.60	0.60
	(0.54, 0.65)	(0.53, 0.67)	(0.48, 0.69)
$\rho_{\tau 0.50}$	0.52	0.59	0.60
	(0.42, 0.61)	(0.54, 0.64)	(0.52, 0.66)
$\rho_{\tau 0.75}$	0.39	0.48	0.56
	(0.30, 0.47)	(0.41, 0.54)	(0.50, 0.62)
σ^2_η		0.47	
		(0.44, 0.52)	
σ^2_η village		0.12	
		(0.12, 0.17)	
σ^2_ε		1.09	
		(1.05, 1.12)	

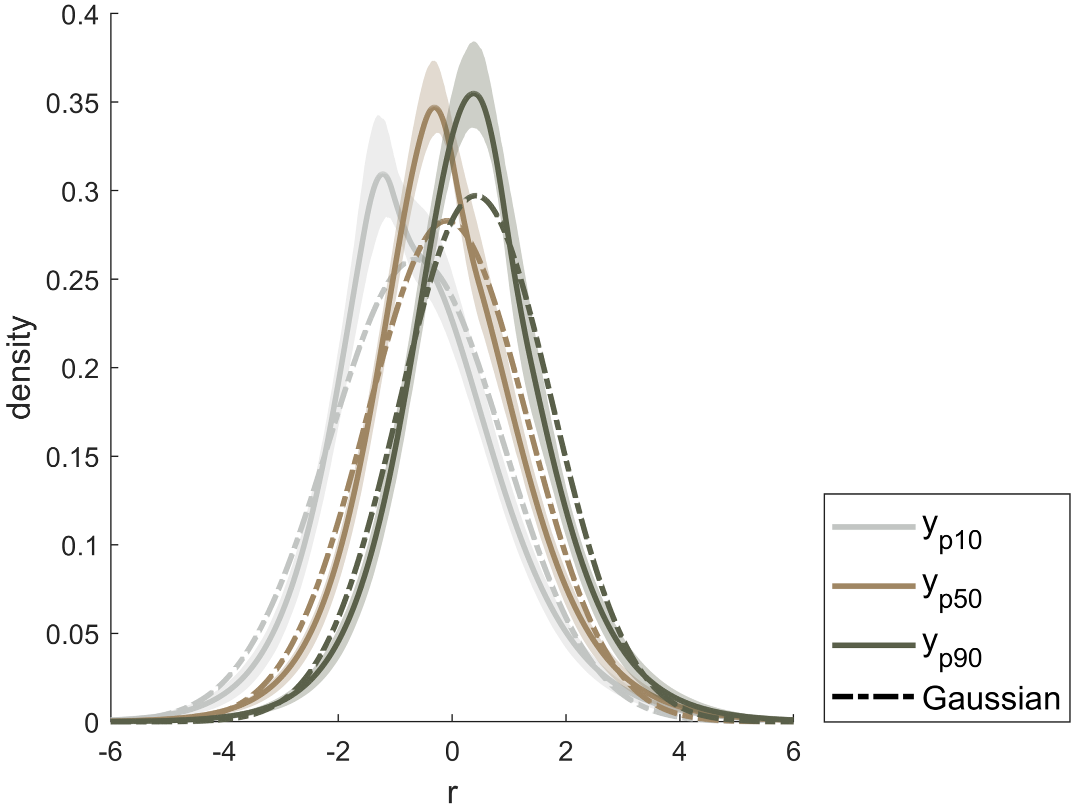
Note. The table reports results for Colombia for the flexible model with additive fixed effects in (4.20). We also include year (survey round) and month (interview) dummies. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.10. Colombia — flexible model (additive fixed effects)



Note. The figure reports estimates of nonlinear persistence for Colombia for the flexible model with additive fixed effects in (4.20). Specifications also include year (survey round) dummies. See Figure C.3.4 for pointwise confidence bands.

FIGURE 4.9. Colombia — flexible model, nonlinear persistence (additive fixed effects)



Note. The figure shows estimated *pdf* s at the 10th, 50th and 90th percentiles of current income for Colombia, calculated by numerical differentiation on the estimated (conditional) cumulative probabilities. Shared areas are 90% pointwise confidence bands using block bootstrap (1000 repetitions). Dotted lines correspond to Normal *pdf* s with the same mean and variance as the empirical *pdfs* of the same color. The range of variation is standardized (future) log income, see Appendix C.2.4.

FIGURE 4.10. Colombia — flexible model, probability density function (additive fixed effects)

4.5.4 Generalizing heterogeneity patterns: interacted fixed effects

In this section, we discuss the estimation of the nonlinear model (4.16) with interacted fixed effects, in which $\beta_2^{\dagger}(s_{jit})$ is allowed to depend on s_{jit}. In these models log odd ratios can vary differentially with fixed effects and therefore allow for a greater distributional role of unobserved heterogeneity in accounting for nonlinearities.

In line with the nonlinear estimates with additive effects, we report results for a parsimonious specification where we choose $L = 3$ for $\beta_0^{\dagger}(\cdot)$ and $\beta_1^{\dagger}(\cdot)$, $L = 2$ for $\beta_2^{\dagger}(\cdot)$ and $\psi(\cdot)$ of order 1. Thus, the model contains a total of 6 parameters — two in the intercept function, three in the interactions with current income, and one in the multiplicative term interacted with the fixed effect. We resort to the linear TSLS estimator introduced in Section 4.3.5 (which does not impose the restrictions in equation (4.19)) using a full set of first-stage interaction terms.[18]

Tables 4.11 and 4.12 report the results for India and Colombia, respectively. Starting with the results for India, we observe some noticeable differences relative to the nonlinear additive model estimates in Table 4.9. First, dispersion risk is now smaller overall, although it is still decreasing with current income. Thus, it appears that a larger fraction of the spread in the subjective probability distributions is now accounted for by unobserved heterogeneity as opposed to risk. Secondly, negative skewness is now more prominent overall, while the increase in skewness risk with current income is much reduced. Finally, although the pattern of nonlinear persistence remains the same, there is a smaller reduction in persistence at the bottom of the income distribution in the presence of a positive shock. The results for Colombia in Table 4.12 tell a similar story relative to those in Table 4.10 for the nonlinear additive model.

State dependence versus unobserved heterogeneity. We have found that state dependence and unobserved heterogeneity compete as sources to explain persistence, not only in linear models (the comparison between columns 1 and 2 in tables 4.3 and 4.4), but also in the case of nonlinear persistence when fixed effects are allowed a flexible distributional role.[19] Our results show that both state dependence and unobserved heterogeneity matter, linearly and nonlinearly, and illustrate how to quantify the relative contributions of each one to different features of a distributional income process estimated from subjective expectations data.

[18] Remember that we also include year indicators in all specifications, and that the TSLS estimator on the transformed equation (4.19) requires us to account for additional "included" regressors (even if the nonlinear restrictions are not imposed). In the case of Colombia, relative to the nonlinear estimates with additive fixed effects, we excluded survey (month) indicators, which would increase the regressor set by 16 additional coefficients and tend to introduce instability in the estimates.

[19] Using a different model for realized outcomes, Almuzara (2020) considers a related problem of distinguishing between nonlinear state dependence and (variance) unobserved heterogeneity. He shows that a fixed effect in the variance of transitory shocks may give rise to spurious nonlinear dynamics.

	y_{p10}	y_{p50}	y_{p90}
$IQR_{0.75}$	0.56	0.46	0.42
	(0.49, 0.79)	(0.39, 0.56)	(0.33, 0.48)
$IQR_{0.90}$	1.31	1.04	0.90
	(1.04, 3.32)	(0.83, 1.50)	(0.70, 1.12)
$SK_{0.90}$	−0.25	−0.29	−0.29
	(−0.70, −0.04)	(−0.50, −0.11)	(−0.45, −0.12)
$\rho_{\tau 0.25}$	1.00	1.05	1.07
	(0.93, 1.11)	(1.01, 1.10)	(1.03, 1.10)
$\rho_{\tau 0.50}$	0.93	1.01	1.04
	(0.83, 0.97)	(0.95, 1.03)	(0.99, 1.06)
$\rho_{\tau 0.75}$	0.82	0.97	1.02
	(0.63, 0.88)	(0.89, 0.99)	(0.95, 1.04)
σ^2_η		0.49	
		(0.38, 0.63)	
σ^2_η village		0.19	
		(0.18, 0.29)	
σ^2_ε		1.10	
		(1.01, 1.26)	

Note. The table reports results for India for the flexible model in (4.16). We also include year (survey round) dummies. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.11. India — flexible model (multiplicative fixed effects)

	y_{p10}	y_{p50}	y_{p90}
$IQR_{0.75}$	1.91	1.62	1.52
	(1.22, 4.19)	(1.11, 3.00)	(1.10, 2.41)
$IQR_{0.90}$	3.85	3.57	3.48
	(2.49, 8.13)	(2.39, 6.74)	(2.37, 6.18)
$SK_{0.90}$	0.37	0.27	0.16
	(0.21, 0.56)	(0.13, 0.50)	(0.05, 0.37)
$\rho_{\tau 0.25}$	0.59	0.49	0.38
	(0.46, 0.69)	(0.26, 0.65)	(−0.07, 0.62)
$\rho_{\tau 0.50}$	0.50	0.58	0.49
	(−0.31, 0.68)	(0.41, 0.68)	(0.20, 0.65)
$\rho_{\tau 0.75}$	0.19	0.26	0.39
	(−1.24, 0.53)	(−0.72, 0.57)	(−0.39, 0.63)
σ^2_η		0.47	
		(0.41, 0.58)	
σ^2_η village		0.11	
		(0.11, 0.17)	
σ^2_ε		1.10	
		(1.05, 1.23)	

Note. The table reports results for Colombia for the flexible model in (4.16). We also include year (survey round) dummies. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

TABLE 4.12. Colombia — flexible model (multiplicative fixed effects)

CHAPTER 5

CONCLUSIONS

In this final chapter, I revisit the contributions of the thesis and reflect on the challenges and opportunities afforded by modern, rich microdata environments.

Possibilities are wide-ranging, and include a better understanding of the transmission of shocks and the nature of heterogeneity, sharpened inference in contexts where a substantial fraction of the population of interest is observed, and incorporating into the econometric toolbox data on hypothetical scenarios and subjective expectations that directly inform us of economic agents' beliefs and perceptions.

Challenges are ubiquitous as well, and often require the development of new econometric approaches. Non-standard asymptotics are one recurrent theme throughout this thesis, incorporating key ingredients of the setup — such as additional information about the sampling process or embeddings regulating the signal-to-noise value of different shocks — in order to provide better approximations. Another is methods to handle new types of dependent variables, such as subjective probabilistic assessments of future outcomes.

Another central message is the imperative to explicitly incorporate and account for microeconomic heterogeneity in making progress towards tackling these challenges. It modulates the extent to which finite-populations inference makes a practical difference in applications, motivates the use of microdata to answer macro questions, plays a crucial role in devising robust inference procedures in the presence of aggregate shocks, and permeates every aspect of modelling income processes with subjective expectations data. This thesis thus emphasizes the importance of keeping heterogeneity at the forefront of econometric research.

At a higher level of generality, a further lesson from this thesis is the need for econometrics to keep developing in continuous dialogue with empirical research. All methodological questions explored here emerge in response to pressing empirical needs. At the same time, the thesis illustrates the enduring versatility of panel data econometrics, which remains more relevant than ever. Many of the tools employed here draw from well-established techniques for random coefficient models or time series methods for panels with a short or moderate time dimension.

Finite populations

Finite population problems, where the sample at hand is a relevant fraction of the population of interest, are ubiquitous in empirical work. Despite its salience, the standard treatments of inference in finite populations assume that the features of interest are observable upon sampling, which limits their adoption in applications.

In Chapter 2, I proposed new methods to assess estimation uncertainty in problems where a finite population coexists with a measurement problem. I introduced finite-population variance estimators that guarantee non-conservative inference and applied these methods to two different empirical applications: on predicting police violence and on studying firm misallocation with census data. Finite-population inference allows for a systematic approach to uncertainty quantification in setups where uncertainty has been previously understood in different ways and leads to large gains in precision in setups where routine practice has been to report standard errors as if the sample were negligible relative to the population.

I also leave some interesting dimensions for future work. Extending the finite-populations framework to a "many measurements" context presents no conceptual difficulty. If anything, some tasks are simplified: whereas weak dependence remains a key assumption, more agnostic approaches to dependence are possible, as those in the time series tradition. Having access to many measurements also allows extensions to more general nonlinear models and facilitates constructing Finite Population Corrections. Further formalizing these ideas also seems a promising direction for future work. Similarly, the framework in this chapter can be extended to more general persistent–transitory measurement models and dynamic panel data problems in short panels.

Macro shocks

The use of micro data to answer macro questions offers an exciting avenue to study how agents respond to economy-wide policies. Yet, this is not without difficulties. In Chapter 3, we proposed a disciplined approach to uncertainty quantification when both aggregate and idiosyncratic shocks coexist and interest is in parameters identified solely by macro shocks. One such scenario is the estimation of impulse responses to macro shocks when rich micro data and a measurement of the shock of interest are available. Despite the complex environment, inference is simple and robust: it involves lag augmentation and clustering at the time level, and is valid regardless of the relative signal of macro shocks in the microdata.

Our basic 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 an object of interest too — see, for instance, the literature on cross-sectional fiscal multipliers (Chodorow-Reich, 2019).

Finally, we leave some interesting dimensions for future research. Quantifying signal-to-noise (perhaps a lower bound) seems relevant in settings where uniform inference is not possible; we expect that these issues become 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 signal-to-noise 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 impulse response horizons could be made building on the techniques in Jordà (2009) and Montiel Olea and Plagborg-Møller (2019).

Subjective expectations

In Chapter 4, we developed an econometric framework for modeling income risk and heterogeneity from the responses to subjective expectation questions of Indian and Colombian households. One main conclusion is that linear income processes are soundly rejected in both datasets. Subjective income distributions feature heteroskedasticity, conditional skewness, and nonlinear persistence. We found a negative association between conditional dispersion and current income, and between conditional skewness and current income. We also documented that persistence diminishes for poor households experiencing large positive shocks, but not for richer households experiencing large negative shocks.

Unobserved heterogeneity matters and is composed of both household-specific and aggregate-level factors. We found that state dependence and unobserved heterogeneity compete as explanations of risk and persistence, both linearly and nonlinearly, which emphasizes the importance of allowing for flexible distributional unobserved heterogeneity to capture their relative contributions. Finally, we also explored whether not only current income but also its sources matter for risk, thereby calling for a larger state space than is common in the literature, and found only moderate evidence for the role of those additional state variables.

Taken together, our results suggest complex and heterogeneous patterns of transmission of income shocks to consumption, involving precautionary dispersion and skewness motives, which depend on the household position in the income distribution.

APPENDIX A

APPENDIX TO CHAPTER 2

A.1 Proofs

The derivations here are a proof sketch of Propositions 2.1 and 2.2; the former is here integrated in the proof of the latter.

Let λ denote a fixed column vector $\lambda \neq 0_{(k+p)\times 1}$ and let $\hat{f} = N/n$. The finite-population variance is $\hat{\mathrm{V}}(\hat{f})$ in (2.14). Throughout, we condition on $\hat{\mathrm{V}}(\hat{f}) \geq 0$ (in the matrix sense), which is a measure-one event in the limit. At all times, we maintain Assumption 2.2 and the regularity conditions in Assumption A.1. For (B) below we also invoke Assumption 2.1 and assume rank $Q_i^*(\delta)\, S_{(m)} = m$. Then, as $n \to \infty$:

(A) $\left(\lambda' \mathrm{V}(f) \lambda\right)^{-1/2} \sqrt{N} \lambda' \left(\hat{\gamma} - \gamma_n\right) \xrightarrow{\text{d}} \mathrm{N}(0, 1)$,

(B) $\left(\lambda' \hat{\mathrm{V}}(\hat{f}) \lambda\right) / \left(\lambda' \mathrm{V}(f) \lambda\right) \xrightarrow{\text{p}} 1.$

where $\mathrm{V}(f)$ is defined in (2.12). (A) and (B) are established in Lemmas A.1 and A.2, respectively. Since λ can be chosen arbitrarily, (A) and the Cramér-Wold device imply Proposition 2.1. (A) and (B) imply that

$$\left(\lambda' \hat{\mathrm{V}}(\hat{f}) \lambda / N\right)^{-1/2} \lambda' \left(\hat{\gamma} - \gamma_n\right) \xrightarrow{\text{d}} \mathrm{N}(0, 1),$$

and thus Proposition 2.2 follows.

Assumption A.1 (Regularity conditions for limit theorems). *(Sketch) (i) (Identification) For large enough n and for every $\tilde{\gamma} \in \Gamma$, $\inf_{\tilde{\gamma}:\, ||\tilde{\gamma} - \gamma_n|| \geq \varepsilon} ||E_n \left[E \left[\psi(Y_i, W_i, \tilde{\gamma})\right]\right]|| > 0$, (ii) compact parameter space $\Gamma \subseteq \mathbb{R}^{p+k}$, (iii) square integrability and local Lipschitz conditions on $\psi(Y_i, W_i, \tilde{\gamma})$ (regularity for Z-estimators; van der Vaart (1998, Chapter 5)), (iv) existence of the appropriate limits of sequences, (v) (Moments) $E\left[||\varepsilon_i||^4\right] \leq C < \infty$.*

Lemma A.1 (Asymptotic normality of the rescaled estimation error). *For an arbitrary column vector $\lambda \neq 0_{(k+p)\times 1}$,*

$$\left(\lambda' V(f) \lambda\right)^{-1/2} \sqrt{N} \lambda' \left(\hat{\gamma} - \gamma_n\right) \xrightarrow{\text{d}} \mathrm{N}(0, 1).$$

Proof. Under regularity conditions in Assumption A.1, the sample moment condition (2.9) admits an expansion

$$\sqrt{N}\left(\hat{\gamma}-\gamma_n\right) = H_n^{-1} n^{-1/2} \sum_{i=1}^{n} \frac{R_i}{\sqrt{f_n}} \psi(Y_i, W_i, \gamma_n) + o_p(1),$$

where we have also used that $\hat{f}/f_n \xrightarrow{\text{p}} 1$ by Assumption (2.2). Now, note that from independent random sampling and repeatedly using $E\left[R_i\right] = f_n$,

$$\operatorname{Var}\left(\frac{R_i}{\sqrt{f_n}}\psi(Y_i, W_i, \gamma_n)\right) = E\left[\psi(Y_i, W_i, \gamma_n)\psi(Y_i, W_i, \gamma_n)'\right] \\ - f_n E\left[\psi(Y_i, W_i, \gamma_n)\right] E\left[\psi(Y_i, W_i, \gamma_n)\right]', \tag{A.1}$$

and that averaging over the population yields $\mathrm{V}_{\psi,n}(f_n)$ in (2.11). For an arbitrary vector $\lambda \neq 0_{(k+p)\times 1}$, $\{\lambda'\psi(Y_i, W_i, \gamma_n)\}_n$ is a row-wise independent triangular array and note that $a'E_n\left[\psi(Y_i, W_i, \gamma_n)\right] = 0$ from (2.10). Asymptotic normality of these averages follows by a Lyapunov-type condition; here we invoke Lemma A.1 in Abadie et al. (2020). Letting $\mathrm{V}_\psi(f) = \lim_{n\to\infty} \mathrm{V}_{\psi,n}(f_n)$, the asymptotic variance is given by $\lambda'\mathrm{V}_\psi(f)\lambda$, and the result follows via the the Cramér-Wold device. □

Lemma A.2 (Consistency of the finite-population standard error). *For an arbitrary column vector* $\lambda \neq 0_{(k+p)\times 1}$,

$$\frac{\lambda'\hat{V}(\hat{f})\lambda}{\lambda' V(f)\lambda} \xrightarrow{\text{p}} 1.$$

Proof. I focus on $\hat{\mathrm{V}}_\psi(\hat{f})$; the regularity conditions in Assumption A.1 immediately imply convergence of $\hat{H}$ to its limits. We first characterize $\mathrm{V}_{\psi,n}(f_n)$. It is immediate from the expressions in (2.10), $E\left[\varepsilon_i\varepsilon_i'\right] = \mathbf{\Omega}_i$ and the result in (A.1) that

$$\mathrm{V}_{\psi,n}(f_n) = \begin{pmatrix} E_n\left[\mathbf{\Psi}_{\delta\delta,i}\right] & E_n\left[\mathbf{\Psi}_{\delta\beta,i}\right] \\ E_n\left[\mathbf{\Psi}'_{\delta\beta,i}\right] & (1-f_n)E_n\left[\mathbf{\Psi}_{\beta\beta,i}\right] + f_n E_n\left[\tilde{\mathbf{\Psi}}_{\beta\beta,i}\right] \end{pmatrix}.$$

where $\psi_{\beta,i} = h_1(W_i, \beta_n)\left(\theta_i - h_0\left(W_i;\beta_n\right)\right)$ and

$$\begin{aligned} \mathbf{\Psi}_{\delta\delta,i} &= \mathrm{A}(W_i, \delta)Q_i(\delta)\mathbf{\Omega}_i Q_i(\delta)\mathrm{A}(W_i, \delta)', \\ \mathbf{\Psi}_{\delta\beta,i} &= \mathrm{A}(W_i, \delta)Q_i(\delta)\mathbf{\Omega}_i\left(g_1\left(X_i;\delta\right)^\dagger h_1(W_i, \beta_n)\right)', \\ \mathbf{\Psi}_{\beta\beta,i} &= \psi_{\beta,i}\psi'_{\beta,i} + \underbrace{h_1(W_i, \beta_n)g_1\left(X_i;\delta\right)^\dagger \mathbf{\Omega}_i\left(g_1\left(X_i;\delta\right)^\dagger h_1(W_i, \beta_n)\right)'}_{\equiv\tilde{\mathbf{\Psi}}_{\beta\beta,i}}. \end{aligned}$$

Next, define

$$\Lambda_i\left(f_n\right) = \operatorname{vec}^{-1}\left[\left(1-f_n\right)I_{T^2} + f_n S_{(m)}\left(Q_i^*\left(\delta\right)S_{(m)}\right)^\dagger Q_i^*\left(\delta\right)\right]\left(u\left(Y_i, W_i, \gamma_n\right)\otimes u\left(Y_i, W_i, \gamma_n\right)\right),$$

and note that

$$u\left(Y_i, W_i, \gamma_n\right) = g_1\left(X_i;\delta\right)\left(\theta_i - h_0\left(W_i;\beta_n\right)\right) + \varepsilon_i.$$

Furthermore, let $g_1^*\left(X_i;\delta\right) = g_1\left(X_i;\delta\right) \otimes g_1\left(X_i;\delta\right)$ and note that using Assumption 2.1,

$$E\left[u\left(Y_i, W_i, \gamma_n\right) \otimes u\left(Y_i, W_i, \gamma_n\right)\right] = g_1^*\left(X_i;\delta\right)\left(\theta_i - h_0\left(W_i;\beta_n\right)\right)^2 + S_{(m)}\omega_i$$

and $Q_i^*\left(\delta\right) E\left[u\left(Y_i, W_i, \gamma_n\right) \otimes u\left(Y_i, W_i, \gamma_n\right)\right] = Q_i^*\left(\delta\right) S_{(m)}\omega_i$. Further, using that $Q_i^*\left(\delta\right) S_{(m)}$ has column rank, it follows that

$$E\left[\Lambda_i\left(f_n\right)\right] = \left(1 - f_n\right)\left[g_1\left(X_i;\delta\right)\left(\theta_i - h_0\left(W_i;\beta_n\right)\right)^2 g_1\left(X_i;\delta\right)' + \Omega_i\right] + f_n\Omega_i.$$

The above shows unbiasedness of $\Lambda_i(f_n)$ precisely for the term in the (finite-population) score. We can then bound the variance of this term . One can proceed similarly for the residual term (using $\hat{\gamma}$ instead of γ_n). Note that $Q_i(\delta)g_1(X_i;\delta) = 0_T$ by construction, which shows why the finite-population variance is independent of f_n for common parameters despite the way it is contructed.

□

A.2 Additional derivations

A.2.1 Section 3.2: conservativeness of the cluster-robust variance

Here we show that $E\left[\hat{\mathrm{V}}^{\text{cluster}}\right] = \mathrm{V}(0)$, where

$$\hat{\mathrm{V}}^{\text{cluster}} = \frac{1}{N(N-1)}\sum_{i=1}^{n} R_i\left(\bar{Y}_i - \hat{\beta}\right)^2.$$

In order to see this, it is helpful to rewrite the expression as

$$\hat{\mathrm{V}}^{\text{cluster}} = \frac{1}{N^2}\sum_{i=1}^{n} R_i\bar{Y}_i^2 - \frac{1}{N^2(N-1)}\sum_{i=1}^{n}\sum_{j\neq i} R_iR_j\bar{Y}_i\bar{Y}_j.$$

Taking expectations, we have

$$\begin{aligned} E\left[\hat{\mathrm{V}}^{\text{cluster}}\right] &= \frac{E_n\left[E\left[\bar{Y}_i^2\right]\right]}{N} - \frac{\frac{1}{n-1}E_n\left[\sum_{j\neq i} E\left[\bar{Y}_i\bar{Y}_j\right]\right]}{N} = \frac{E_n\left[\theta_i^2\right] + \sigma^2/T}{N} - \frac{\frac{1}{n-1}E_n\left[\sum_{j\neq i}\theta_i\theta_j\right]}{N} \\ &= \frac{\frac{n}{n-1}E_n\left[\left(\theta_i - \beta_n\right)^2\right]}{N} + \frac{\sigma^2/T}{N} = \mathrm{V}(0), \end{aligned}$$

where we have used Assumption 2.S1 in the first line[1] and Assumption 2.S2 in the second one.

[1]In particular, note that $E\left[R_i\right] = N/n$ and $E\left[R_iR_j\right] = N(N-1)/n(n-1)$ for $j \neq i$ for simple random sampling without replacement.

A.2.2 Finite-population inference for variances

Here I extend the results in Section 3.3 to cover $\beta_n = \text{Var}_n(\theta_i) = (n-1)^{-1}\sum_{i=1}^n(\theta_i - \bar{\theta}_n)^2$ (where $\bar{\theta}_n$ is the average θ_i in the population) in the context of the misallocation empirical application in Section 2.5.2 and the measurement model in equation (2.20).

In particular, consider the variance estimator in equation (2.22), which we can extend to allow for weak dependence as in Assumption 2.1:

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{n} R_i T^{-2} 1'_{T^2}\left[I_{T^2} - S_{(m)}\left(Q_i^* S_{(m)}\right)^{\dagger} Q_i^*\right]\hat{Y}_i^* = \frac{1}{N}\sum_{i=1}^{n} R_i T^{-2} 1'_{T^2}\tilde{Y}_i^*, \tag{A.2}$$

with an obvious definition of $\tilde{Y}_i^*$. The motivation for this estimator can be traced back to equation (2.16) in Remark 2.8, which recasts the measurement system for θ_i as a measurement system for θ_i^2. A valid, conservative (large-sample) variance estimator for $\hat{\beta}$ is given by

$$\hat{\text{V}}(0) = \frac{1}{N}\sum_{i=1}^{n} R_i\left(T^{-2}1'_{T^2}\tilde{Y}_i^* - \hat{\beta}\right)^2.$$

Again through the lens of equation (2.16), it can be seen that the corresponding FPC is given by

$$\text{FPC} = \lim_{n\to\infty} E_n\left[(\theta_i - \bar{\theta}_n)^4\right] - \beta_n^2.$$

We can then leverage Remark 2.6, which shows that we can construct finite-population variance estimators if we have access to a conservative estimator that is consistent for $\text{V}(0)$ and a valid estimator of the FPC. A candidate for the latter follows by noting that the FPC is also equal to (the limit of) $\kappa_{4n} + 2\beta_n^2$, where

$$\kappa_{4n}(\theta_i) = E_n\left[(\theta_i - \bar{\theta}_n)^4\right] - 3E_n\left[(\theta_i - \bar{\theta}_n)^2\right]^2.$$

Arellano and Bonhomme (2012, Appendix A) propose estimators of fourth-order cumulants. This approach allows us to obtain valid finite-population estimators under the same notion of dependence over measurements used in estimation, but we do need to restrict the higher-order dependence between measurement errors and unobserved attributes; statistical independence would be a sufficient condition.[2]

A.3 Empirics: additional results

A.3.1 Additional results from Section 2.5.1

Tables A.1, A.2 and A.3 report finite-population confidence intervals for all results reported in the main empirical section in Montiel-Olea et al. (2021) (section 5.2 of the paper). The entries correspond to the ten

[2]One way to operationalize independence is to define a probability distribution over θ_i and measurement errors in the limit over sequences of growing finite populations and then impose these restrictions. Weaker conditions are possible imposing restrictions on limits of certain sums over the finite population. Similarly, independence can be relaxed to zero cross-cumulants up to fourth order.

largest police departments by population served. Table A.1 considers counterfactuals based on both observed and unobserved determinants, Table A.2 considers only counterfactual unobserved determinants and Table A.3 only observed ones.

TABLE A.1. Counterfactual homicides for 2013-2018: observed and unobserved determinants

(a) Conventional inference

	Phoenix	Las Vegas	Dallas	San Antonio	Los Angeles	Houston	San Diego	Chicago	Philadelphia	New York
Phoenix	**93**	[52,53]	[33,47]	[31,45]	[34,40]	[31,34]	[31,32]	[28,32]	[21,30]	[5,9]
Las Vegas	[93,93]	**51**	[33,46]	[31,44]	[34,40]	[31,33]	[30,32]	[28,32]	[21,30]	[5,9]
Dallas	[68,97]	[38,54]	**33**	[32,33]	[27,37]	[24,33]	[23,33]	[22,30]	[21,23]	[5,7]
San Antonio	[76,109]	[42,61]	[37,39]	**35**	[30,41]	[26,37]	[25,37]	[24,34]	[23,25]	[6,8]
Los Angeles	[269,314]	[150,176]	[105,141]	[99,135]	**113**	[95,106]	[89,106]	[86,98]	[66,91]	[17,27]
Houston	[145,156]	[81,87]	[54,74]	[50,71]	[56,62]	**51**	[48,53]	[46,50]	[34,48]	[9,14]
San Diego	[79,83]	[44,46]	[28,40]	[27,39]	[29,35]	[26,29]	**26**	[23,28]	[18,26]	[4,8]
Chicago	[189,216]	[105,121]	[72,100]	[68,96]	[75,85]	[67,73]	[62,74]	**63**	[46,64]	[12,19]
Philadelphia	[89,130]	[50,73]	[44,47]	[41,45]	[36,50]	[31,44]	[30,44]	[29,40]	**28**	[7,9]
New York	[568,1013]	[317,566]	[270,371]	[259,348]	[237,375]	[201,342]	[189,342]	[184,310]	[175,236]	**55**
Totals	[1689,2279]	[942,1273]	[753,882]	[717,836]	[699,848]	[596,769]	[562,776]	[545,700]	[481,567]	[125,166]

(b) Finite-population inference

	Phoenix	Las Vegas	Dallas	San Antonio	Los Angeles	Houston	San Diego	Chicago	Philadelphia	New York
Phoenix	**93**	[52,53]	[35,44]	[34,42]	[35,39]	[32,33]	[31,32]	[28,32]	[23,28]	[6,8]
Las Vegas	[93,93]	**51**	[35,43]	[33,41]	[35,39]	[31,33]	[30,32]	[28,31]	[23,28]	[6,8]
Dallas	[73,90]	[41,50]	**33**	[32,33]	[28,36]	[25,31]	[24,30]	[23,29]	[21,23]	[5,7]
San Antonio	[81,101]	[45,56]	[37,39]	**35**	[32,40]	[28,35]	[27,34]	[25,33]	[23,25]	[6,7]
Los Angeles	[274,307]	[153,172]	[107,137]	[102,130]	**113**	[96,104]	[91,104]	[87,97]	[69,88]	[19,24]
Houston	[146,155]	[82,86]	[56,71]	[53,67]	[57,62]	**51**	[48,53]	[46,50]	[36,45]	[10,13]
San Diego	[79,83]	[44,46]	[30,38]	[29,36]	[30,34]	[27,29]	**26**	[24,28]	[19,24]	[5,7]
Chicago	[190,214]	[106,120]	[74,96]	[71,92]	[75,84]	[67,72]	[62,73]	**63**	[49,61]	[13,18]
Philadelphia	[96,120]	[54,67]	[44,47]	[41,45]	[38,48]	[33,42]	[32,41]	[30,38]	**28**	[7,9]
New York	[643,870]	[359,486]	[281,355]	[269,333]	[261,332]	[225,298]	[214,292]	[204,275]	[180,228]	**55**
Totals	[1791,2094]	[1000,1170]	[751,882]	[716,833]	[723,809]	[626,717]	[597,704]	[565,667]	[481,564]	[134,155]

Note: Diagonal entries are observed lethal encounters (totalling 548 encounters). Off-diagonal entries are 90% confidence intervals for counterfactual values of lethal encounters $\hat{Y}_t^*(i, j, z_j)$ in equation (2.19), which replace characteristics of agency i in the rows with that of agency j in the columns; see the text in Section 2.5.1 for additional details. In this case, we replace both observed and unobserved determinants of police use of deadly force.

TABLE A.2. Counterfactual homicides for 2013-2018: unobserved determinants

(a) Conventional inference

	Phoenix	Las Vegas	Dallas	Philadelphia	San Diego	Chicago	Los Angeles	Houston	San Antonio	New York
Phoenix	**93**	[64,70]	[52,74]	[46,78]	[51,66]	[48,64]	[44,54]	[44,53]	[40,57]	[13,25]
Las Vegas	[70,77]	**51**	[40,58]	[36,60]	[40,51]	[38,49]	[34,43]	[34,41]	[30,45]	[10,20]
Dallas	[43,61]	[31,44]	**33**	[28,37]	[26,38]	[25,36]	[23,31]	[22,31]	[24,28]	[8,12]
Philadelphia	[35,60]	[25,42]	[26,35]	**28**	[22,35]	[23,32]	[19,30]	[19,29]	[19,29]	[7,11]
San Diego	[39,50]	[28,35]	[24,36]	[22,35]	**26**	[23,29]	[21,26]	[21,25]	[18,28]	[6,11]
Chicago	[93,125]	[67,87]	[60,86]	[57,81]	[60,74]	**63**	[50,64]	[53,59]	[44,68]	[16,28]
Los Angeles	[198,244]	[139,175]	[125,170]	[111,172]	[121,150]	[115,146]	**113**	[103,125]	[95,130]	[34,54]
Houston	[92,112]	[66,79]	[57,79]	[53,78]	[56,68]	[57,63]	[47,58]	**51**	[42,62]	[15,26]
San Antonio	[60,85]	[42,62]	[43,52]	[36,56]	[35,54]	[34,53]	[32,43]	[30,45]	**35**	[11,17]
New York	[207,402]	[147,284]	[152,234]	[148,222]	[133,233]	[130,224]	[117,189]	[111,197]	[116,178]	**55**
Totals	[952,1279]	[677,909]	[656,802]	[591,809]	[595,762]	[577,731]	[531,614]	[510,630]	[489,630]	[177,257]

(b) Finite-population inference

	Phoenix	Las Vegas	Dallas	Philadelphia	San Diego	Chicago	Los Angeles	Houston	San Antonio	New York
Phoenix	**93**	[65,69]	[55,70]	[51,70]	[54,61]	[51,61]	[46,51]	[46,51]	[43,53]	[16,21]
Las Vegas	[71,75]	**51**	[43,54]	[40,53]	[43,47]	[40,47]	[35,41]	[36,39]	[34,41]	[13,16]
Dallas	[45,58]	[33,41]	**33**	[30,35]	[28,35]	[27,34]	[23,31]	[24,30]	[25,27]	[9,11]
Philadelphia	[39,54]	[28,38]	[28,33]	**28**	[24,32]	[24,31]	[20,29]	[21,27]	[20,26]	[8,10]
San Diego	[41,47]	[30,33]	[26,33]	[24,32]	**26**	[24,28]	[21,25]	[22,24]	[20,25]	[8,10]
Chicago	[99,118]	[71,83]	[63,82]	[60,77]	[62,71]	**63**	[51,62]	[53,58]	[48,63]	[19,24]
Los Angeles	[208,232]	[146,168]	[126,168]	[116,165]	[124,146]	[119,142]	**113**	[106,121]	[98,125]	[39,49]
Houston	[96,107]	[69,76]	[60,75]	[56,72]	[59,65]	[58,62]	[49,56]	**51**	[46,58]	[18,22]
San Antonio	[64,78]	[46,56]	[45,50]	[39,51]	[38,49]	[36,48]	[33,42]	[32,41]	**35**	[12,16]
New York	[248,326]	[177,232]	[169,212]	[159,204]	[152,199]	[147,190]	[131,165]	[130,164]	[129,162]	**55**
Totals	[1026,1165]	[732,828]	[662,793]	[611,777]	[626,717]	[600,691]	[535,603]	[536,585]	[508,602]	[198,232]

Note: Diagonal entries are observed lethal encounters (totalling 548 encounters). Off-diagonal entries are 90% confidence intervals for counterfactual values of lethal encounters $\hat{Y}_t^*(i, j, z_i)$ in equation (2.19), which replace characteristics of agency i in the rows with that of agency j in the columns; see the text in Section 2.5.1 for additional details. In this case, we replace only unobserved determinants of police use of deadly force.

TABLE A.3. Counterfactual homicides for 2013-2018: observed determinants

(a) Conventional inference

	Phoenix	San Antonio	Las Vegas	Los Angeles	Houston	Dallas	San Diego	Chicago	Philadelphia	New York
Phoenix	**93**	[69,79]	[70,78]	[65,77]	[57,70]	[53,67]	[45,59]	[43,59]	[31,52]	[27,46]
San Antonio	[43,49]	**35**	[33,39]	[34,35]	[29,33]	[27,32]	[23,27]	[22,27]	[16,24]	[14,21]
Las Vegas	[63,69]	[48,56]	**51**	[46,55]	[40,49]	[37,47]	[32,41]	[31,41]	[22,37]	[19,32]
Los Angeles	[139,164]	[116,121]	[109,130]	**113**	[95,108]	[89,103]	[77,88]	[73,89]	[52,79]	[47,68]
Houston	[70,86]	[56,65]	[55,67]	[55,63]	**51**	[48,50]	[38,47]	[39,44]	[28,39]	[24,35]
Dallas	[48,61]	[39,46]	[38,47]	[38,44]	[36,37]	**33**	[26,33]	[28,30]	[20,27]	[17,24]
San Diego	[43,56]	[36,42]	[35,44]	[35,40]	[30,37]	[28,35]	**26**	[23,30]	[17,26]	[16,22]
Chicago	[103,140]	[84,105]	[81,108]	[82,100]	[76,84]	[73,78]	[58,74]	**63**	[45,58]	[39,52]
Philadelphia	[52,89]	[43,66]	[41,69]	[42,63]	[39,54]	[37,50]	[30,46]	[32,41]	**28**	[23,29]
New York	[114,195]	[96,143]	[91,151]	[94,135]	[84,119]	[80,110]	[69,94]	[69,91]	[56,71]	**55**
Totals	[774,996]	[642,738]	[610,773]	[626,699]	[564,611]	[532,572]	[442,516]	[445,493]	[320,435]	[286,378]

(b) Finite-population inference

	Phoenix	San Antonio	Las Vegas	Los Angeles	Houston	Dallas	San Diego	Chicago	Philadelphia	New York
Phoenix	**93**	[72,76]	[72,76]	[69,74]	[59,68]	[55,64]	[48,55]	[46,56]	[34,47]	[31,41]
San Antonio	[44,47]	**35**	[35,37]	[34,35]	[29,32]	[27,31]	[24,26]	[23,27]	[17,22]	[16,19]
Las Vegas	[64,68]	[51,54]	**51**	[48,52]	[42,47]	[40,44]	[35,38]	[33,39]	[25,33]	[22,28]
Los Angeles	[145,156]	[117,120]	[114,122]	**113**	[97,106]	[91,100]	[80,85]	[75,87]	[57,73]	[52,63]
Houston	[72,83]	[58,64]	[58,64]	[56,61]	**51**	[48,49]	[40,44]	[40,43]	[30,36]	[27,32]
Dallas	[50,58]	[40,45]	[40,45]	[39,43]	[36,36]	**33**	[28,31]	[28,30]	[21,25]	[19,22]
San Diego	[46,52]	[38,40]	[37,41]	[36,39]	[32,35]	[30,33]	**26**	[25,28]	[19,24]	[17,20]
Chicago	[108,132]	[87,102]	[86,101]	[84,97]	[78,83]	[73,78]	[61,70]	**63**	[48,54]	[43,48]
Philadelphia	[58,79]	[47,61]	[46,61]	[45,58]	[42,50]	[40,47]	[33,42]	[34,39]	**28**	[24,27]
New York	[129,169]	[105,130]	[102,130]	[102,123]	[92,107]	[87,100]	[74,88]	[75,84]	[60,66]	**55**
Totals	[813,937]	[655,719]	[648,720]	[636,686]	[578,595]	[543,560]	[459,493]	[450,484]	[341,407]	[309,353]

Note: Diagonal entries are observed lethal encounters (totalling 548 encounters). Off-diagonal entries are 90% confidence intervals for counterfactual values of lethal encounters $\hat{Y}_t^*(i, i, z_j)$ in equation (2.19), which replace characteristics of agency i in the rows with that of agency j in the columns; see the text in Section 2.5.1 for additional details. In this case, we replace only observed determinants of police use of deadly force.

A.3.2 Additional results from Section 2.5.2

In this section, I report additional results from the application to firm misallocation in Section 2.5.2.

First, for completeness, Figures A.3, A.4 and A.5 report results for allocative efficiency in the sense of $d \log \mathrm{TFP}$ in equation (2.21) for the remaining size quartiles. Second, whereas in the text I have focused on a baseline measurement model with uncorrelated measurements (but unrestricted heteroskedasticity), I report here results further allowing for weak dependence.

In particular, given the unbalanced nature of the panel, I allow for richer forms of m-dependence as a larger number of measurements becomes available. Let T_i denote the number of such periods for firm i. I choose $m = 1$ for $T_i = 3$, $m = 2$ if $T_i \in \{4, 5\}$, $m = 3$ if $T_i \in \{5, 6, 7, 8\}$ and $m = 4$ if $T_i \in \{9, 10\}$. In other words, for firms that enter in 1991 and remain active throughout the panel, we allow for unrestricted dependence in measurement errors over up to four year horizons. It is easy to see that these choices satisfy the order condition in Assumption 2.1, sometimes with equality. Note that for chapter2 relative to allocative efficiency not only the confidence intervals but also the point estimates might change as a consequence, see equation (A.2).

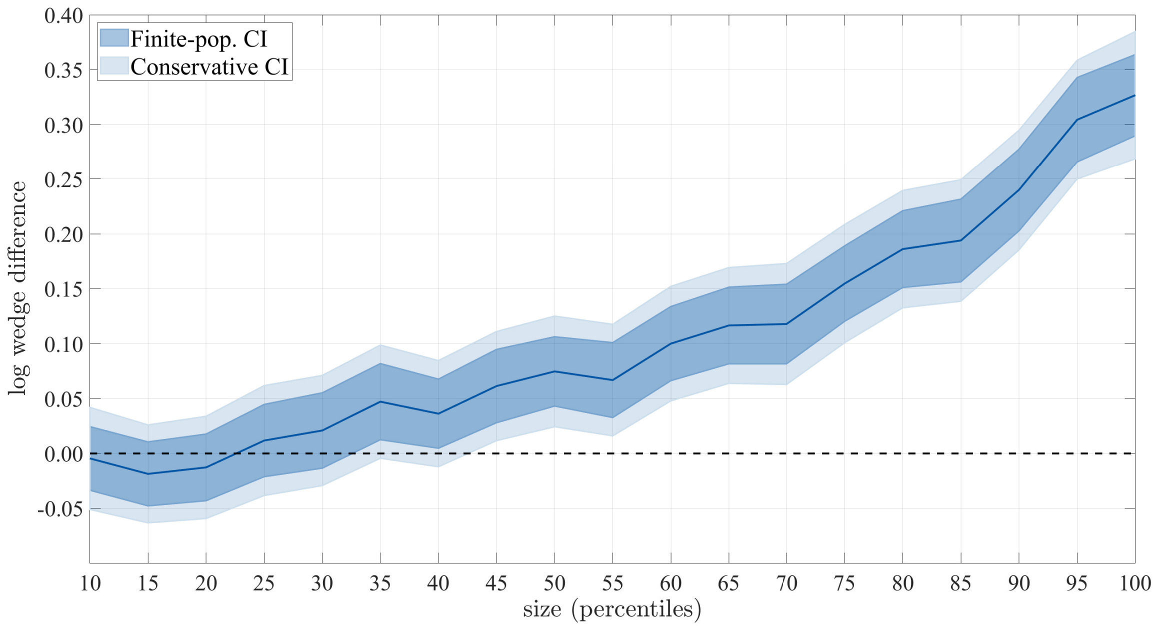


FIGURE A.1. Labor wedges across the distribution of firm size at entry (relative to 5th percentile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. Results allowing for m-dependence in measurements, see the text for additional details.

Figures A.1 and A.2 are the counterparts to Figures 2.4 and 2.5. In qualitative terms, the results are similar to those reported in Chapter 2, although the finite-population confidence intervals tend to be wider. This reinforces the main message in Section 2.5.2 emphasizing the need to account for measurement-based uncertainty while leaving a minor role for sampling-based uncertainty — even if the analyst treats the population as a negligible fraction of an infinite superpopulation.

It is also important to stress that sensitivity to the baseline assumption of uncorrelated measurements might suggest either a restrictive notion of weak dependence or misspecification of the underlying measurement system. Regarding the former, the m-dependence restrictions above are only marginally rejected when

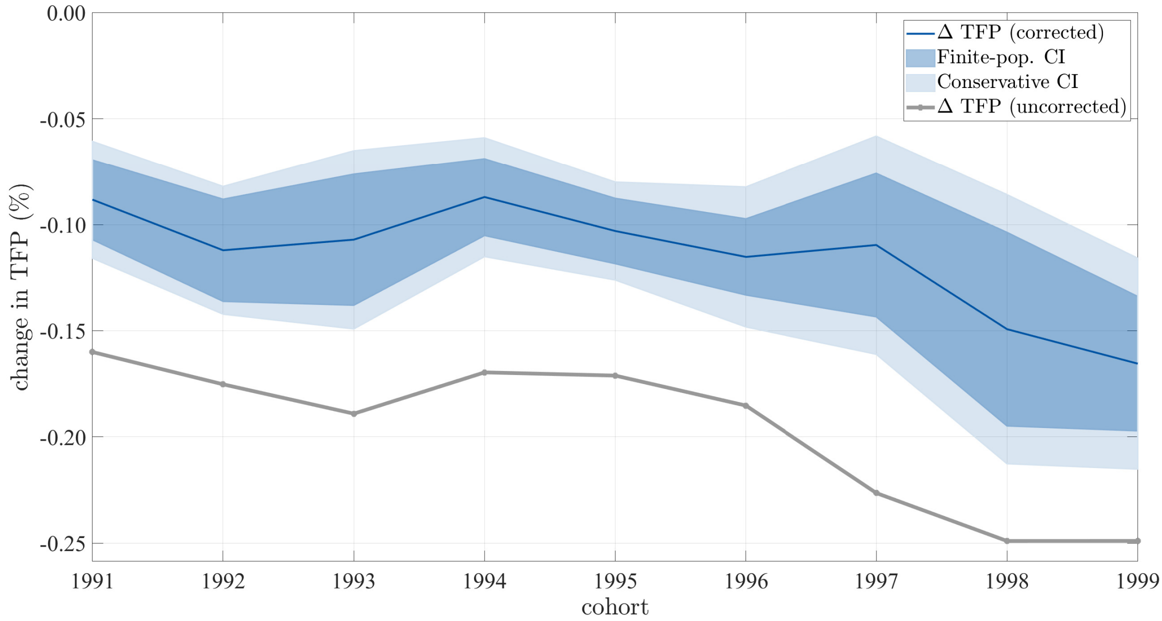
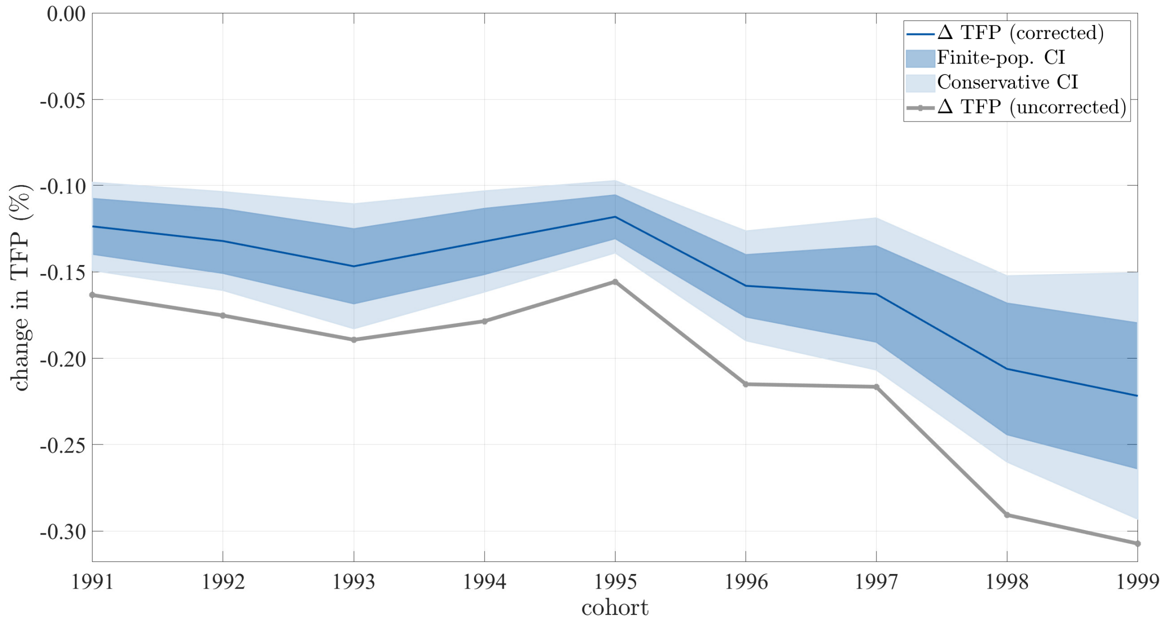
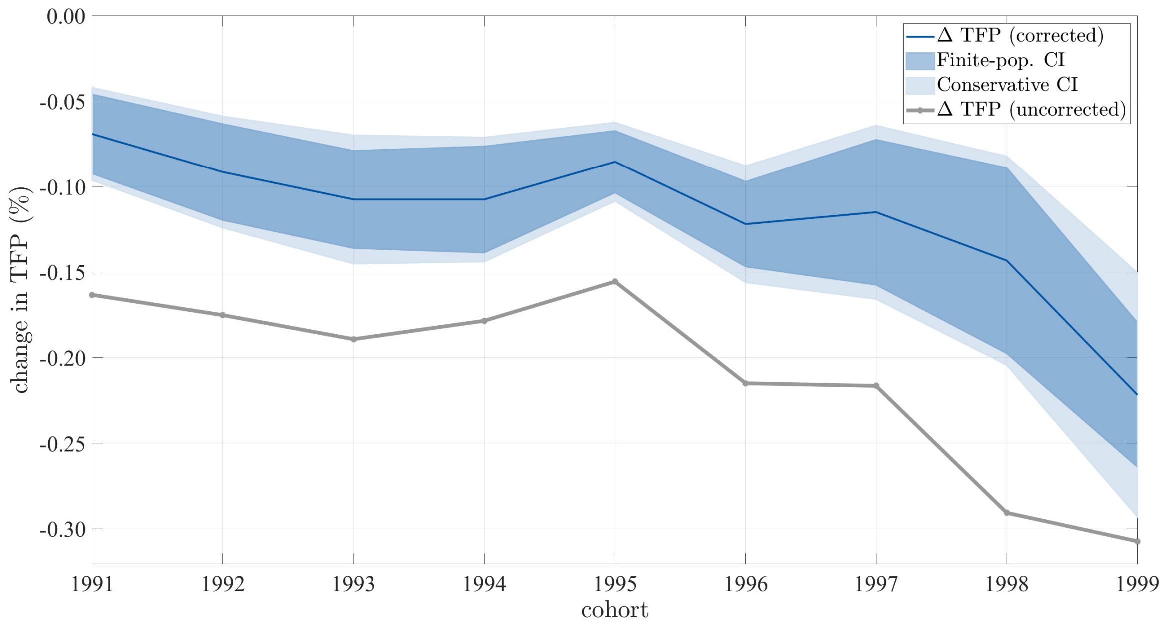


FIGURE A.2. Evolution of allocative efficiency as in equation (2.21) for each cohort (within firms in the bottom size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. Results allowing for m-dependence in measurements, see the text for additional details.

I implement a test of covariance structures along the lines of Remark 2.4. An alternative is to consider richer measurement equations for θ_i beyond the benchmark model. This seems particularly promising in this context, where there might be a time-varying systematic component in labor-related distortions or persistent, predictable variation in MRPL beyond what is captured in equation (2.20). All of these can be framed within the class of models discussed in Section 3.3.

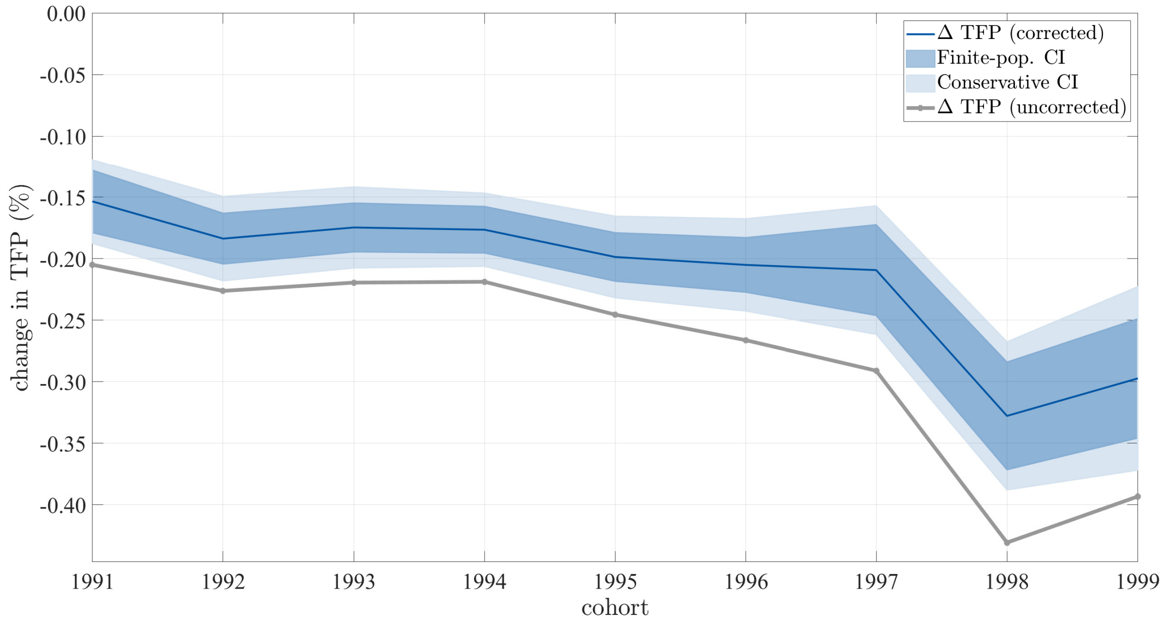


(a) Uncorrelated measurements.

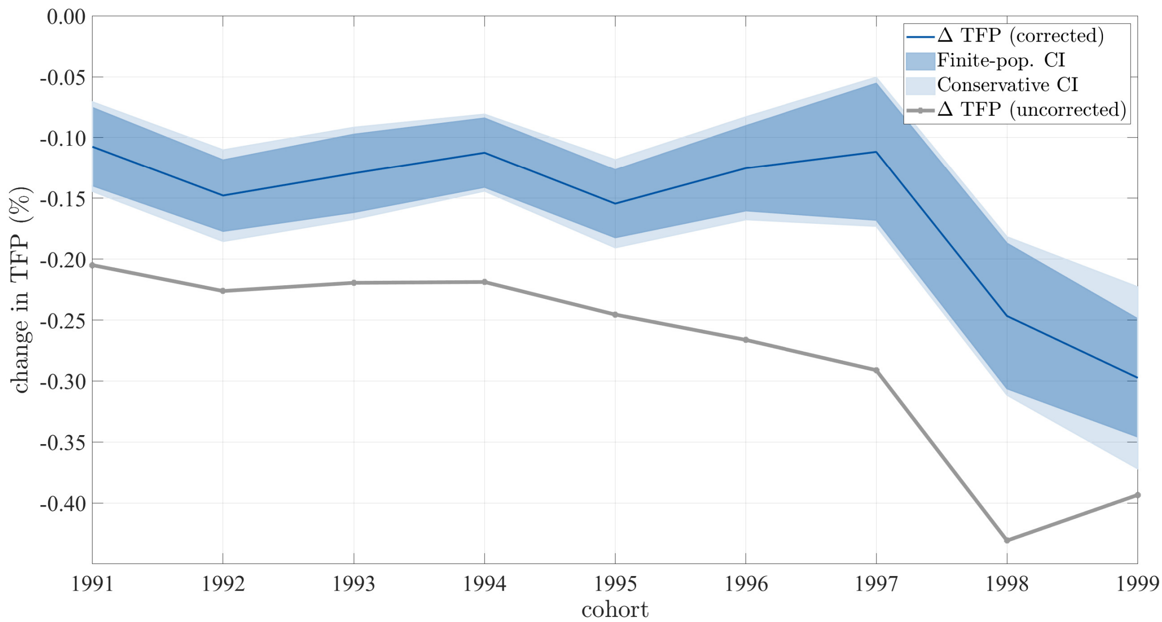


(b) m-dependent measurements.

FIGURE A.3. Evolution of allocative efficiency as in equation (2.21) for each cohort (within firms in the second size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. See the text for details on dependence over measurements.

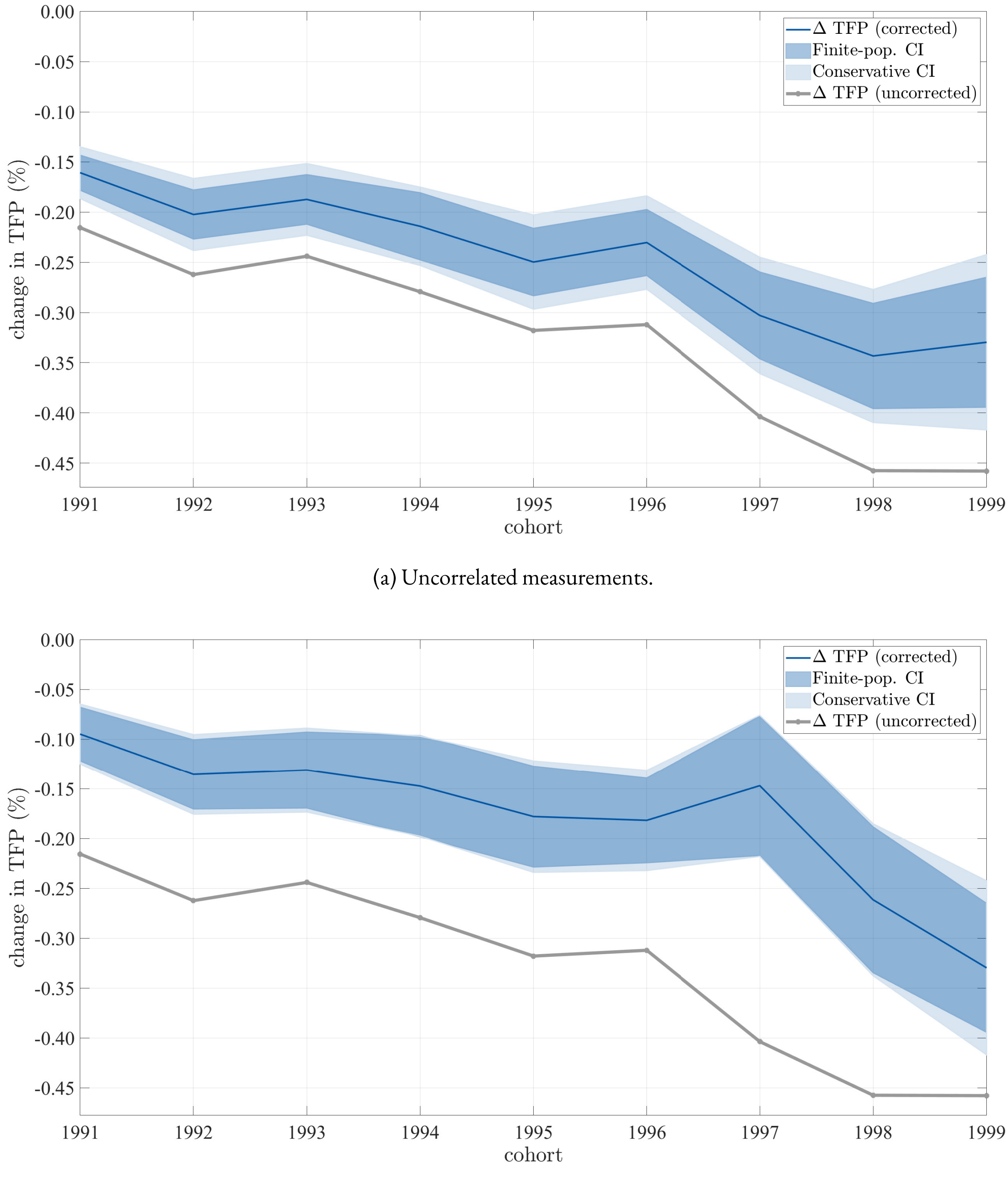


(a) Uncorrelated measurements.



(b) m-dependent measurements.

FIGURE A.4. Evolution of allocative efficiency as in equation (2.21) for each cohort (within firms in the third size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. See the text for details on dependence over measurements.



(a) Uncorrelated measurements.

(b) m-dependent measurements.

FIGURE A.5. Evolution of allocative efficiency as in equation (2.21) for each cohort (within firms in the upper size quartile). 95% confidence bands (finite-population and conservative) are displayed together with the point estimates. See the text for details on dependence over measurements.

APPENDIX B

APPENDIX TO CHAPTER 3

B.1 Additional proofs

We adopt the following notation in the proofs below. We use P_N, E_N, Var_N, Cov_N to denote probability, expectation, variance and covariance given $\{\theta_i, s_i\}_{i=1}^N$ (we insert a subindex κ or κ_T when necessary).

With a slight abuse of nomenclature we sometimes call Loève's inequality to the statement $|\sum_{i=1}^m X_i|^r \leq c_r \sum_{i=1}^m |X_i|^r$ (with $c_r = 1$ if $r \leq 1$ and $c_r = m^{r-1}$ otherwise) where $X_1, \ldots, X_m$ are random variables and not just to $E\left[|\sum_{i=1}^m X_i|^r\right] \leq c_r \sum_{i=1}^m E\left[|X_i|^r\right]$ (which is implied by the former). See Davidson (1994, Theorem 9.28).

Without loss of generality we assume $\kappa \geq 0$. We also define the scaling function $g(\kappa) = \max\{1, \kappa\}$ and note that $g(\kappa)/\kappa = g(\kappa^{-1})$. In Proposition 3.1

$$\frac{V(h,\kappa)}{g(\kappa^2/N)} = \frac{\sum_{\ell=0}^{\infty}\left\{\iota_\ell(h)\bar{\beta}_\ell^2 E_N\left[X_t^2 X_{t+h-\ell}^2\right] + \bar{\gamma}_\ell^2 E_N\left[X_t^2 Z_{t+h-\ell}^2\right]\right\}}{g(\kappa^2/N)} + \frac{\sum_{i=1}^N \sum_{\ell=0}^{\infty} \hat{s}_i^2 \delta_{i\ell}^2 E_N\left[X_t^2 u_{i,t+h-\ell}^2\right]}{N g(N/\kappa^2)}$$

is bounded below by $\underline{CM}^2 > 0$ and above by $3C^4 M_4 < \infty$ for any κ (and h). The same applies to $V(h,\kappa)/g(\kappa^2/N)$ in Proposition 3.2. In Proposition 3.3, $\text{tr}\{\boldsymbol{V}(h,\kappa)\}/g(\kappa^2/N)$ is bounded below by $(a_0^2+1)\underline{CM}^2 > 0$ and above by $6(p+1)(a_0^2+1)C^4 M_4 < \infty$.

Proposition 3.1

Parts (A), (B) and (C) of the proof of Proposition 3.1 in Appendix A.1 are established in Lemmas B.1, B.2 and B.3 below. Lemmas B.4 and B.5 provide auxiliary results for Lemma B.1, while B.6 and B.7 do the same for B.2. At all times, we make Assumptions 3.1, 3.2 and 3.3 and we fix h and $p \geq h$ as $T, N \to \infty$ (note we do not need $T/N \to 0$ here).

Lemma B.1 (Asymptotic normality of the score).

$$\frac{\sum_{t=1}^{T-h} X_t \xi_t(h,\kappa_T)}{\sqrt{(T-h)V(h,\kappa_T)}} \xrightarrow[P_{\kappa_T}]{\text{d}} N(0,1).$$

Proof. The argument relies on the martingale representation:

$$\sum_{t=1}^{T-h} \frac{X_t \xi_t(h, \kappa_T)}{\sqrt{(T-h)V(h, \kappa_T)}} = \sum_{t=1}^{T} \chi_{T,t}(h, \kappa_T)$$

where we have defined

$$\chi_{T,t}(h, \kappa) = \frac{X_t \Xi_{X,t}(h, \kappa) + Z_t \Xi_{Z,t}(h) + (\kappa_T/N) \sum_{i=1}^{N} u_{it} \Xi_{U,it}(h)}{\sqrt{(T-h)V(h, \kappa_T)}}$$

with

$$\begin{aligned}
\Xi_{X,t}(h, \kappa) &= \sum_{\ell=1}^{h} \mathbb{1}\{t-\ell \geq 1\} \bar{\beta}_{h-\ell} X_{t-\ell} + \sum_{\ell=p+1}^{\infty} \mathbb{1}\{t \leq T-h\} \bar{\beta}_{h+\ell} X_{t-\ell} \\
&\quad + \sum_{\ell=0}^{\infty} \mathbb{1}\{t \leq T-h\} \left[\bar{\gamma}_{h+\ell} Z_{t-\ell} + \frac{\kappa}{N} \sum_{i=1}^{N} \hat{s}_i \delta_{i,h+\ell} u_{i,t-\ell} \right], \\
\Xi_{Z,t}(h) &= \sum_{\ell=1}^{h} \mathbb{1}\{t-\ell \geq 1\} \bar{\gamma}_{h-\ell} X_{t-\ell}, \\
\Xi_{U,it}(h) &= \sum_{\ell=1}^{h} \mathbb{1}\{t-\ell \geq 1\} \hat{s}_i \delta_{i,h-\ell} X_{t-\ell}.
\end{aligned}$$

Under Assumption 3.2, it can be readily verified that $\{\chi_{T,t}(h, \kappa_T)\}_{t=1}^{T}$ is a martingale difference array adapted to the natural filtration $\{\mathcal{F}_{T,t}\}_{t=1}^{T}$,

$$\mathcal{F}_{T,t} = \sigma\left(\{X_\tau, Z_\tau, \{u_{i\tau}\}_{i=1}^{N}\}_{\tau \leq t}, \{\theta_i, s_i\}_{i=1}^{N} \right),$$

that is, $\chi_{T,t}(h, \kappa_T)$ is $\mathcal{F}_{T,t}$-measurable and $E_{\kappa_T}\left[\chi_{T,t}(h, \kappa_T) \middle| \mathcal{F}_{T,t-1}\right] = 0$.

By construction, $\sum_{t=1}^{T} E_{\kappa_T}\left[\chi_{T,t}(h, \kappa_T)^2\right] = 1$ and we can show (Lemmas B.4 and B.5)

$$\sum_{t=1}^{T} \chi_{T,t}(h, \kappa_T)^2 \xrightarrow[P_{\kappa_T}]{\text{p}} 1 \text{ and } \lim_{T\to\infty} \sum_{t=1}^{T} E_{\kappa_T}\left[\chi_{T,t}(h, \kappa_T)^4\right] = 0.$$

By Davidson (1994, Theorems 23.11, 23.16 and 24.3), the Lemma follows. □

Lemma B.2 (Consistency of the standard error).

$$\frac{\hat{V}(h)}{V(h, \kappa_T)} \xrightarrow[P_{\kappa_T}]{\text{p}} 1.$$

Proof. Since $V(h, \kappa_T) > 0$ holds P_{κ_T}-a.s., it suffices to show that

$$\frac{\hat{V}(h) - V(h, \kappa_T)}{g(\kappa_T^2/N)} \xrightarrow[P_{\kappa_T}]{\text{p}} 0.$$

Write

$$\frac{\hat{V}(h) - V(h,\kappa_T)}{g(\kappa_T^2/N)} = D_{T,1}(h,\kappa_T) + D_{T,2}(h,\kappa_T),$$

where we have defined

$$D_{T,1}(h,\kappa_T) = \sum_{t=1}^{T-h} \frac{\left(X_t^2\xi_t(h,\kappa_T)^2 - E_{\kappa_T}\left[X_t^2\xi_t(h,\kappa_T)^2 \middle| \{\theta_i, s_i\}_{i=1}^N\right]\right)}{(T-h)g(\kappa_T^2/N)},$$

$$D_{T,2}(h,\kappa_T) = \sum_{t=1}^{T-h} \left[\frac{\left(N^{-1}\sum_{i=1}^N \hat{x}_{it}(h)\hat{\xi}_{it}(h)\right)^2 - X_t^2\xi_t(h,\kappa_T)^2}{(T-h)g(\kappa_T^2/N)}\right].$$

Next, using $(x^2 - y^2) = (x-y)(x+y)$ and the Cauchy-Schwarz inequality,

$$\left|D_{T,2}(h,\kappa_T)\right| \leq \sqrt{D_{T,2}^-(h,\kappa_T)}\sqrt{D_{T,2}^+(h,\kappa_T)},$$

with

$$D_{T,2}^-(h,\kappa_T) = \sum_{t=1}^{T-h} \frac{\left[\left(N^{-1}\sum_{i=1}^N \hat{x}_{it}(h)\hat{\xi}_{it}(h)\right) - X_t\xi_t(h,\kappa_T)\right]^2}{(T-h)g(\kappa_T^2/N)},$$

$$D_{T,2}^+(h,\kappa_T) = \sum_{t=1}^{T-h} \frac{\left[\left(N^{-1}\sum_{i=1}^N \hat{x}_{it}(h)\hat{\xi}_{it}(h)\right) + X_t\xi_t(h,\kappa_T)\right]^2}{(T-h)g(\kappa_T^2/N)}.$$

Adding and subtracting $X_t\xi_t(h,\kappa_T)$ within the squares in $D_{T,2}^+(h,\kappa_T)$ and applying Loève's inequality,

$$D_{T,2}^+(h,\kappa_T) \leq 2D_{T,2}^-(h,\kappa_T) + 8|D_{T,1}(h,\kappa_T)| + \frac{8V(h,\kappa_T)}{g(\kappa_T^2/N)}.$$

We can show (Lemmas B.6 and B.7) that $D_{T,1}(h,\kappa_T) = o_{P_{\kappa_T}}(1)$ and $D_{T,2}^-(h,\kappa_T) = o_{P_{\kappa_T}}(1)$. Given that $V(h,\kappa_T)/g(\kappa_T^2/N)$ is bounded P_{κ_T}-a.s., $D_{T,2}^+(h,\kappa_T) = O_{P_{\kappa_T}}(1)$ which implies $D_{T,2}(h,\kappa_T) = o_{P_{\kappa_T}}(1)$ and the Lemma follows. □

Lemma B.3 (Negligibility of the reminder).

$$R_T(h,\kappa_T) \xrightarrow[P_{\kappa_T}]{\text{p}} 0.$$

Proof. Let $\bar{x}_t(h) = (X_{t-1} - \bar{X}_1(h), \ldots, X_{t-p} - \bar{X}_p(h))'$ where $\bar{X}_\ell(h) = (T-h)^{-1}\sum_{t=1}^{T-h} X_{t-\ell}$. Since either $\hat{s}_i$ was demeaned or time effects were not included as controls,

$$\hat{\pi}(h)'W_{it} = \hat{\pi}_{0,i}(h) + \sum_{\ell=1}^{p} \hat{\pi}_{X,\ell}(h)\hat{s}_i X_{t-\ell} = \hat{s}_i\left(\bar{X}_0(h) + \hat{\pi}_X(h)'\bar{x}_t(h)\right),$$

where $\{\hat{\pi}_{0,i}(h)\}, \pi_X(h) = (\hat{\pi}_{X,1}(h), \ldots, \hat{\pi}_{X,p}(h))'$ are the coefficients from the regression of $s_i X_t$ on unit fixed effects and p lags of $\hat{s}_i X_t$. Furthermore, it is readily seen that $\hat{\pi}_X(h)$ are also the coefficients in a regression of X_t on $\bar{x}_t(h)$,

$$\hat{\pi}_X(h) = \left[\sum_{t=1}^{T-h} \bar{x}_t(h)\bar{x}_t(h)'\right]^{-1} \sum_{t=1}^{T-h} \bar{x}_t(h) X_t.$$

Note that $E\left[X_{t-\ell}\right] = E\left[X_{t-\ell}X_t\right] = 0$ and that $\text{Var}\left(\sum_{t=1}^{T-h} X_{t-\ell}\right)$, $\text{Var}\left(\sum_{t=1}^{T-h} X_{t-\ell}X_t\right)$ are bounded by a constant (M_2 and M_4, respectively) times $(T-h)$ under Assumptions 3.1, 3.2 and 3.3. Also note that $(T-h)^{-1}\sum_{t=1}^{T-h} \bar{x}_t(h)\bar{x}_t(h)' = E\left[X_t^2\right] \times I_p + o_{P_{\kappa_T}}(1)$. All of this is independent of κ_T. It follows that

$$\bar{X}_0(h) = O_{P_{\kappa_T}}\left((T-h)^{-1/2}\right), \qquad \hat{\pi}_X(h) = O_{P_{\kappa_T}}\left((T-h)^{-1/2}\right).$$

Write

$$R_T(h,\kappa_T) = -\frac{\bar{X}_0(h)\sum_{t=1}^{T-h}\xi_t(h,\kappa_T)}{\sqrt{(T-h)V(h,\kappa_T)}} - \frac{\hat{\pi}_X(h)'\sum_{t=1}^{T-h}\bar{x}_t(h)\xi_t(h,\kappa_T)}{\sqrt{(T-h)V(h,\kappa_T)}}.$$

To obtain $R_T(h,\kappa_T) = o_{P_{\kappa_T}}(1)$, we show $\left\{(T-h)V(h,\kappa_T)\right\}^{-1/2}\sum_{t=1}^{T}\xi_t(h,\kappa_T) = O_{P_{\kappa_T}}(1)$ and that $\left\{(T-h)V(h,\kappa_T)\right\}^{-1/2}\sum_{t=1}^{T}\bar{x}_t(h)\xi_t(h,\kappa_T) = O_{P_{\kappa_T}}(1)$. We do so by direct calculation.

First,

$$\begin{aligned}
E_{N,\kappa_T}\left[\left(\sum_{t=1}^{T-h}\xi_t(h,\kappa_T)\right)^2\right] &= E_N\left[\left(\sum_{t=1}^{T-h}\sum_{\ell=0}^{\infty}\iota_\ell(h)\bar{\beta}_\ell X_{t+h-\ell}\right)^2\right] + E_N\left[\left(\sum_{t=1}^{T-h}\sum_{\ell=0}^{\infty}\bar{\gamma}_\ell Z_{t+h-\ell}\right)^2\right] \\
&\quad + \frac{\kappa_T^2}{N^2}E_N\left[\left(\sum_{t=1}^{T-h}\sum_{i=1}^{N}\sum_{\ell=0}^{\infty}\hat{s}_i\delta_{i\ell}u_{i,t+h-\ell}\right)^2\right] \\
&\leq 2(T-h)\left[\left(\sum_{\ell=0}^{\infty}\iota_\ell(h)|\bar{\beta}_\ell|\right)^2 E_N\left[X_t^2\right] + \left(\sum_{\ell=0}^{\infty}|\bar{\gamma}_\ell|\right)^2 E_N\left[Z_t^2\right]\right. \\
&\quad \left. + \frac{\kappa_T^2}{N^2}\sum_{i=1}^{N}\left(\sum_{\ell=0}^{\infty}|\hat{s}_i\delta_{i\ell}|\right)^2 E_N\left[u_{it}^2\right]\right] \\
&\leq (T-h)\times 2(2+\kappa_T^2/N)C^4 M_2,
\end{aligned}$$

where the last line uses Assumption 3.3(i)–(iv).[1] By iterated expectations and Chebyshev's inequality, for any

[1]We also used the fact that for any linear process $\omega_t = \sum_{\ell=0}^{\infty}\varphi_\ell\varepsilon_{t-\ell}$ where $\{\varphi_\ell\}$ are absolutely summable and $\{\varepsilon_t\}$ is white noise with $E\left[\varepsilon_t\right] = 0$ and $E\left[\varepsilon_t^2\right] = 1$,

$$E\left[\left(\sum_{t=1}^{T}\omega_t\right)^2\right] = \sum_{m=-(T-1)}^{T-1}(T-|m|)\sum_{\ell=0}^{\infty}\varphi_\ell\varphi_{\ell+|m|} \leq T\sum_{\ell=0}^{\infty}|\varphi_\ell|\sum_{m=-\infty}^{\infty}|\varphi_{\ell+|m|}| \leq 2T\left(\sum_{\ell=0}^{\infty}|\varphi_\ell|\right)^2.$$

$\varepsilon > 0$,

$$P_{\kappa_T}\left(\left|\frac{\sum_{t=1}^T \xi_t(h,\kappa_T)}{\sqrt{(T-h)V(h,\kappa_T)}}\right| > \varepsilon\right) = E_{\kappa_T}\left[P_{N,\kappa_T}\left(\left|\frac{\sum_{t=1}^T \xi_t(h,\kappa_T)}{\sqrt{(T-h)V(h,\kappa_T)}}\right| > \varepsilon\right)\right]$$
$$\leq \frac{1}{\varepsilon^2} E_{\kappa_T}\left[\frac{2(2+\kappa_T^2/N)C^4 M_2}{V(h,\kappa_T)}\right] \leq \frac{1}{\varepsilon^2}\frac{6C^4 M_2}{\underline{\text{CM}}^2} < \infty,$$

where the bound on $(2+\kappa_T^2/N)/V(h,\kappa_T) = ((2+\kappa_T^2/N)/g(\kappa_T^2/N)) \times (g(\kappa_T^2/N)/V(h,\kappa_T))$ uses $(2+\kappa)/g(\kappa) \leq 3$ and $V(h,\kappa_T)/g(\kappa_T^2/N) \geq \underline{\text{CM}}^2$.

Similarly for any $k = 1, \dots, p$,

$$E_{N,\kappa_T}\left[\left(\sum_{t=1}^{T-h} X_{t-k}\xi_t(h,\kappa_T)\right)^2\right] \leq (T-h)\left[\sum_{\ell=0}^{\infty} \iota_\ell(h)\bar{\beta}_\ell^2 E_N\left[X_{t-k}^2 X_{t+h-\ell}^2\right] + \sum_{\ell=0}^{\infty}\bar{\gamma}_\ell^2 E_N\left[X_{t-k}^2 Z_{t+h-\ell}^2\right]\right.$$
$$\left. + \frac{\kappa_T^2}{N^2}\sum_{i=1}^N\sum_{\ell=0}^{\infty}\hat{s}_i^2\delta_{i\ell}^2 E_N\left[X_{t-k}^2 u_{i,t+h-\ell}^2\right] + 2\sum_{\ell=1}^{h+k}\iota_{h+k-\ell}(h)\iota_{h+k+\ell}(h)|\bar{\beta}_{h+k-\ell}\bar{\beta}_{h+k+\ell}| E_N\left[X_{t-k}^2 X_{t-k-\ell}^2\right]\right]$$
$$\leq (T-h) \times (4+\kappa_T^2/N)C^4 M_4,$$

where we used the autocovariances of $X_{t-k}\xi_t(h,\kappa_T)$ and Assumption 3.3(i)–(iv) again. By iterated expectations and Chebyshev, for any $\varepsilon > 0$,

$$P_{\kappa_T}\left(\left|\frac{\sum_{t=1}^T X_{t-r}\xi_t(h,\kappa_T)}{\sqrt{(T-h)V(h,\kappa_T)}}\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}E_{\kappa_T}\left[\frac{(4+\kappa_T^2/N)C^4 M_4}{V(h,\kappa_T)}\right] \leq \frac{1}{\varepsilon^2}\frac{5C^4 M_4}{\underline{\text{CM}}^2} < \infty.$$

Thus, $R_T(h,\kappa_T) = o_{P_{\kappa_T}}(1)$ and the Lemma follows. □

Lemma B.4. *Under the conditions of Lemma B.1,*

$$\sum_{t=1}^T \chi_{T,t}(h,\kappa_T)^2 \xrightarrow[P_{\kappa_T}]{\text{p}} 1.$$

Proof. We show $\text{Var}_{N,\kappa_T}\left(\sum_{t=1}^T \chi_{T,t}(h,\kappa_T)^2\right) \leq \bar{V}/(T-h)$ for a constant $\bar{V}$ independent of κ_T. Since $E_{N,\kappa_T}\left[\sum_{t=1}^T \chi_{T,t}(h,\kappa_T)^2\right] = 1$, by iterated expectations and Chebyshev's inequality, for any $\varepsilon > 0$,

$$P_{\kappa_T}\left(\left|\sum_{t=1}^T \chi_{T,t}(h,\kappa_T)^2 - 1\right| > \varepsilon\right) = E_{\kappa_T}\left[P_{N,\kappa_T}\left(\left|\sum_{t=1}^T \chi_{T,t}(h,\kappa_T)^2 - 1\right| > \varepsilon\right)\right]$$
$$\leq \frac{\bar{V}}{\varepsilon^2(T-h)} \to 0.$$

As argued at the beginning of the section, $V(h,\kappa)/g(\kappa^2/N)$ is bounded away from zero and infinity uniformly over κ. Thus, it suffices to show

$$\mathrm{Var}_{N,\kappa_T}\left(\sum_{t=1}^{T}\frac{V(h,\kappa_T)\chi_{T,t}(h,\kappa_T)^2}{g(\kappa_T^2/N)}\right)\leq\frac{\bar{V}}{T-h},$$

P_{κ_T}-a.s., for some constant $\bar{V}$ independent of κ_T. We do this by a direct calculation. Define $\tilde{\chi}_{T,t}(h,\kappa_T) = \chi_{T,t}(h,\kappa_T)\left\{(T-h)V(h,\kappa_T)/g(\kappa_T^2/N)\right\}^{1/2}$ so that

$$\begin{aligned}
g\left(\frac{\kappa_T}{\sqrt{N}}\right)\tilde{\chi}_{T,t}(h,\kappa_T) &= X_t\Xi_{X,t}(h,\kappa)+Z_t\Xi_{Z,t}(h)+\frac{\kappa_T}{N}\sum_{i=1}^{N}u_{it}\Xi_{U,it}(h)\\
&= \underbrace{\sum_{\ell=1}^{\infty}b_{t,\ell}X_tX_{t-\ell}}_{\equiv g(\kappa_T/\sqrt{N})\zeta_{1,t}}+\underbrace{\sum_{\ell=0}^{\infty}c_{t,\ell}X_tZ_{t-\ell}}_{\equiv g(\kappa_T/\sqrt{N})\zeta_{2,t}}+\underbrace{\sum_{\ell=1}^{h}\tilde{c}_{t,\ell}Z_tX_{t-\ell}}_{\equiv g(\kappa_T/\sqrt{N})\zeta_{3,t}}\\
&\quad+\underbrace{\frac{\kappa_T}{N}\sum_{i=1}^{N}\sum_{\ell=0}^{\infty}d_{it,\ell}X_tu_{i,t-\ell}}_{\equiv g(\kappa_T/\sqrt{N})\zeta_{4,t}}+\underbrace{\frac{\kappa_T}{N}\sum_{i=1}^{N}\sum_{\ell=1}^{h}\tilde{d}_{it,\ell}u_{it}X_{t-\ell}}_{\equiv g(\kappa_T/\sqrt{N})\zeta_{5,t}}
\end{aligned} \tag{B.1.1}$$

for some $\{b_{t,\ell},c_{t,\ell},\tilde{c}_{t,\ell},\{d_{it,\ell},\tilde{d}_{it,\ell}\}_{i=1}^N\}$ that depend on $\{\theta_i,s_i\}_{i=1}^N$ (and h). Note that the coefficients depend on t only via the indicator functions $\mathbb{1}\{t-\ell\leq 1\}$ and $\mathbb{1}\{t\leq T-h\}$. We define $\{b_\ell,c_\ell,\tilde{c}_\ell,\{d_{i,\ell},\tilde{d}_{i,\ell}\}_{i=1}^N\}$ as the coefficients we would get by setting the indicators to one. This implies $|b_{t,\ell}|\leq|b_\ell|$, $|c_{t,\ell}|\leq|c_\ell|$, and so on. By Assumption 3.3(iv), $|b_\ell|,|c_\ell|,|\tilde{c}_\ell|,|d_{i\ell}|,|\tilde{d}_{i\ell}|\leq\bar{C}_\ell$ almost surely for finite constants $\bar{C}_\ell$ such that $\bar{C}=\sum_{\ell=1}^{\infty}\bar{C}_\ell<\infty$ (in fact, we can take $\bar{C}\leq C^2$ independent of h).

Consider the variance

$$\mathrm{Var}_{N,\kappa_T}\left(\sum_{t=1}^{T}\frac{V(h,\kappa_T)\chi_{T,t}(h,\kappa_T)^2}{g(\kappa_T^2/N)}\right)=\frac{\sum_{t=1}^{T}\sum_{\tau=1}^{T}\Gamma_T(t,\tau)}{(T-h)^2}$$

where (omitting the dependence on h, κ_T and $\{\theta_i,s_i\}_{i=1}^N$)

$$\Gamma_T(t,\tau)=\mathrm{Cov}_{N,\kappa_T}\left(\tilde{\chi}_{T,t}(h,\kappa_T)^2,\tilde{\chi}_{T,\tau}(h,\kappa_T)^2\right).$$

Expanding the square of $\tilde{\chi}_{T,t}(h,\kappa_T)$ and using the linearity of the covariance we can express $\Gamma_T(t,\tau)$ as the sum of covariances $\Gamma_{T,k_1k_2k_3k_4}(t,\tau)=\mathrm{Cov}_{N,\kappa_T}\left(\zeta_{k_1,t}\zeta_{k_2,t},\zeta_{k_3,\tau}\zeta_{k_4,\tau}\right)$ where k_1,k_2,k_3,k_4 range over the five terms in (B.1.1). Moreover, if $k_1=k_2$, $\Gamma_{T,k_1k_2k_3k_4}(t,\tau)$ can only be non-zero if $k_3=k_4$, while if $k_1\neq k_2$, only if either $k_1=k_3$ and $k_2=k_4$ or $k_1=k_4$ and $k_2=k_3$. Then, by the triangle inequality,

$$\begin{aligned}
\left|\Gamma_T(t,\tau)\right| &= \left|\sum_{k_1=1}^{5}\sum_{k_2=1}^{5}\sum_{k_3=1}^{5}\sum_{k_4=1}^{5}\Gamma_{T,k_1k_2k_3k_4}(t,\tau)\right|\\
&\leq\sum_{k_1=1}^{5}\sum_{k_3=1}^{5}\left|\Gamma_{T,k_1k_1k_3k_3}(t,\tau)\right|+2\sum_{k_1=1}^{5}\sum_{k_2=1}^{5}\left|\Gamma_{T,k_1k_2k_1k_2}(t,\tau)\right|.
\end{aligned} \tag{B.1.2}$$

We begin with $\sum_{t=1}^{T}\sum_{\tau=1}^{T}\Gamma_{T,k_1k_1k_3k_3}(t,\tau)$. Consider $k_1 = k_3 = 1$:

$$
\begin{aligned}
g\left(\frac{\kappa_T^4}{N^2}\right)\left|\Gamma_{T,1111}(t,\tau)\right| &= \left|\mathrm{Cov}_N\left(\left(\sum_{\ell=1}^{\infty} b_{t,\ell}X_tX_{t-\ell}\right)^2, \left(\sum_{\ell=1}^{\infty} b_{\tau,\ell}X_\tau X_{\tau-\ell}\right)^2\right)\right| \\
&= \left|\sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty}\sum_{\ell_3=1}^{\infty}\sum_{\ell_4=1}^{\infty} b_{t,\ell_1}b_{t,\ell_2}b_{\tau,\ell_3}b_{\tau,\ell_4}\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}X_{t-\ell_2}, X_\tau^2X_{\tau-\ell_3}X_{\tau-\ell_4}\right)\right| \\
&\leq \sum_{\ell_1=1}^{\infty}\sum_{\ell_3=1}^{\infty} b_{\ell_1}^2 b_{\ell_3}^2\left|\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}^2, X_\tau^2X_{\tau-\ell_3}^2\right)\right| \\
&\quad + 2\sum_{\ell_1=1}^{\infty}\sum_{\ell_2\neq\ell_1} |b_{\ell_1}b_{\ell_2}b_{\ell_1+\tau-t}b_{\ell_2+\tau-t}|\left|\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}X_{t-\ell_2}, X_\tau^2X_{t-\ell_1}X_{t-\ell_2}\right)\right|.
\end{aligned}
$$

The inequality uses the fact that by Assumption 3.2, $\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}X_{t-\ell_2}, X_\tau^2X_{\tau-\ell_3}X_{\tau-\ell_4}\right)$ can only be non-zero if $\ell_1 = \ell_2$ and $\ell_3 = \ell_4$ or, with $\ell_1 \neq \ell_2$, if either $\ell_3 = \ell_1 + \tau - t$ and $\ell_4 = \ell_2 + \tau - t$ or $\ell_3 = \ell_2 + \tau - t$ and $\ell_4 = \ell_1 + \tau - t$.[2] We also use $|b_{t,\ell}| \leq |b_\ell|$.

For the first double sum, now summing over t and τ,

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{\tau=1}^{T}\sum_{\ell_1=1}^{\infty}\sum_{\ell_3=1}^{\infty} b_{\ell_1}^2 b_{\ell_3}^2\left|\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}^2, X_\tau^2X_{\tau-\ell_3}^2\right)\right| \\
&\qquad \leq 2T\sum_{m=0}^{T-1}\sum_{\ell_1=1}^{\infty}\sum_{\ell_3=1}^{\infty}\bar{C}_{\ell_1}^2\bar{C}^2\left|\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}^2, X_{t-m}^2X_{t-m-\ell_3}^2\right)\right| \\
&\qquad \leq 2T\bar{C}^2\sum_{\ell_1=1}^{\infty}\bar{C}_{\ell_1}^2\left(\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\left|\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}^2, X_{t-m}^2X_{t-m-\ell_3}^2\right)\right|\right) \\
&\qquad \leq 2T\bar{C}^2\bar{K}\sum_{\ell_1=1}^{\infty}\bar{C}_{\ell_1}^2 \leq 2T\bar{C}^4\bar{K}
\end{aligned}
$$

for some constant $\bar{K}$ that can be shown to exist as by Assumption 3.3(iii) the fourth-order cumulants of X_t^2 conditional on $\{\theta_i, s_i\}_{i=1}^N$ are absolutely summable.

Turning to the second double sum, by Assumption 3.2, since $\ell_1 \neq \ell_2$,

$$
\left|\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}X_{t-\ell_2}, X_\tau^2X_{\tau-\ell_1}X_{\tau-\ell_2}\right)\right| = \left|E_N\left[X_t^2X_\tau^2X_{t-\ell_1}^2X_{t-\ell_2}^2\right]\right| \leq E_N\left[X_t^8\right] \leq M_8,
$$

where M_8 is the moment bound from Assumption 3.3(i). Then,

$$
2\sum_{t=1}^{T}\sum_{\tau=1}^{T}\sum_{\ell_1=1}^{\infty}\sum_{\ell_2\neq\ell_1} |b_{\ell_1}b_{\ell_2}b_{\ell_1+\tau-t}b_{\ell_2+\tau-t}|\left|\mathrm{Cov}_N\left(X_t^2X_{t-\ell_1}X_{t-\ell_2}, X_\tau^2X_{t-\ell_1}X_{t-\ell_2}\right)\right|
$$

[2] This is similar to the proof of Montiel Olea and Plagborg-Møller (2021, Lemma A.6)

$$
\begin{aligned}
&\leq 4TM_8 \sum_{m=0}^{T-1} \sum_{\ell_1=1}^{\infty} \sum_{\ell_2 \neq \ell_1} |b_{\ell_1} b_{\ell_2} b_{\ell_1+m} b_{\ell_2+m}| \\
&\leq 4TM_8 \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} |b_{\ell_1}||b_{\ell_2}| \left(\sum_{m=0}^{\infty} |b_{\ell_1+m}||b_{\ell_2+m}| \right) \\
&\leq 4TM_8 \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} |b_{\ell_1}||b_{\ell_2}| \left(\sum_{m_1=1}^{\infty} |b_{m_1}|^2 \sum_{m_2=1}^{\infty} |b_{m_2}|^2 \right)^{1/2} \\
&\leq 4T\bar{C}^4 M_8,
\end{aligned}
$$

where the second inequality increases the range of summation over ℓ_2 and m, the third uses Cauchy-Schwarz and the fourth follows from Assumption 3.3(iv).

Putting these calculations together and using $g(\kappa) \geq 1$,

$$
\frac{\sum_{t=1}^{T} \sum_{\tau=1}^{T} \left| \Gamma_{T,1111}(t,\tau) \right|}{(T-h)^2} \leq \frac{T \times 2\bar{C}^4(\bar{K} + 2M_8)}{g(\kappa_T^4/N^2)(T-h)^2} \leq \frac{2\bar{C}^4(\bar{K} + 2M_8)}{(1 - h/T)(T-h)}.
$$

In fact, the same bound works for $\sum_{t=1}^{T} \sum_{\tau=1}^{T} \left| \Gamma_{T,k_1 k_1 k_3 k_3}(t,\tau) \right|$ for any $k_1, k_3 \in \{1, 2, 3\}$.

Next consider $k_1 = k_3 = 4$:

$$
\begin{aligned}
g\left(\frac{\kappa_T^4}{N^2}\right) \frac{\left|\Gamma_{T,4444}(t,\tau)\right|}{(\kappa_T^4/N^4)} &= \left| \mathrm{Cov}_N\left(\left(\sum_{i=1}^{N} \sum_{\ell=1}^{\infty} d_{it,\ell} X_t u_{i,t-\ell} \right)^2, \left(\sum_{i=1}^{N} \sum_{\ell=1}^{\infty} d_{i\tau,\ell} X_\tau u_{i,\tau-\ell} \right)^2 \right) \right| \\
&= \left| \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{i_3=1}^{N} \sum_{i_4=1}^{N} \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} \sum_{\ell_3=1}^{\infty} \sum_{\ell_4=1}^{\infty} d_{i_1 t,\ell_1} d_{i_2 t,\ell_2} d_{i_3 \tau,\ell_3} d_{i_4 \tau,\ell_4} \right. \\
&\qquad \left. \times \mathrm{Cov}_N\left(X_t^2 u_{i_1,t-\ell_1} u_{i_2,t-\ell_2}, X_\tau^2 u_{i_3,\tau-\ell_3} u_{i_4,\tau-\ell_4} \right) \right| \\
&\leq \sum_{i_1=1}^{N} \sum_{i_3=1}^{N} \left| \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} \sum_{\ell_3=1}^{\infty} \sum_{\ell_4=1}^{\infty} d_{i_1 t,\ell_1} d_{i_1 t,\ell_2} d_{i_3 \tau,\ell_3} d_{i_3 \tau,\ell_4} \right. \\
&\qquad \left. \times \mathrm{Cov}_N\left(X_t^2 u_{i_1,t-\ell_1} u_{i_1,t-\ell_2}, X_\tau^2 u_{i_3,\tau-\ell_3} u_{i_3,\tau-\ell_4} \right) \right| \\
&\quad + \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \left| \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} \sum_{\ell_3=1}^{\infty} \sum_{\ell_4=1}^{\infty} d_{i_1 t,\ell_1} d_{i_2 t,\ell_2} d_{i_1 \tau,\ell_3} d_{i_2 \tau,\ell_4} \right. \\
&\qquad \left. \times \mathrm{Cov}_N\left(X_t^2 u_{i_1,t-\ell_1} u_{i_2,t-\ell_2}, X_\tau^2 u_{i_1,\tau-\ell_3} u_{i_2,\tau-\ell_4} \right) \right| \\
&\quad + \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \left| \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} \sum_{\ell_3=1}^{\infty} \sum_{\ell_4=1}^{\infty} d_{i_1 t,\ell_1} d_{i_2 t,\ell_2} d_{i_2 \tau,\ell_3} d_{i_1 \tau,\ell_4} \right. \\
&\qquad \left. \times \mathrm{Cov}_N\left(X_t^2 u_{i_1,t-\ell_1} u_{i_2,t-\ell_2}, X_\tau^2 u_{i_2,\tau-\ell_3} u_{i_1,\tau-\ell_4} \right) \right|.
\end{aligned}
$$

The inequality uses the fact that $\mathrm{Cov}_N\left(X_t^2 u_{i_1,t-\ell_1} u_{i_1,t-\ell_2}, X_\tau^2 u_{i_3,\tau-\ell_3} u_{i_3,\tau-\ell_4}\right)$ can only be non-zero if $i_1 = i_2$ and $i_3 = i_4$, or $i_1 = i_3$ and $i_2 = i_4$, or $i_1 = i_4$ and $i_2 = i_3$.

Summing over t and τ and applying to each of the three summands on the right hand side the same steps as the case $k_1 = k_3 = 1$,

$$\frac{\sum_{t=1}^{T}\sum_{\tau=1}^{T}\left|\Gamma_{T,4444}(t,\tau)\right|}{(T-h)^2} \le \frac{3N^2 \times \kappa_T^4/N^4 \times 2\bar{C}^4(\bar{K}+2M_8)}{g(\kappa_T^4/N^2)(1-h/T)(T-h)} \le \frac{6\bar{C}^4(\bar{K}+2M_8)}{(1-h/T)(T-h)}.$$

Repeating the calculation for the remaining cases (and noting that this bound is three times larger than the one we computed for $k_1 = k_3 = 1$) we conclude that $6\bar{C}^4(\bar{K}+2M_8)/(1-h/T)(T-h)$ works for any $k_1, k_3 \in \{1,2,3,4,5\}$. By similar reasoning, the bound also works for $\sum_{t=1}^{T}\sum_{\tau=1}^{T}\Gamma_{T,k_1k_2k_1k_2}(t,\tau)$ whenever $k_1 \neq k_2$. We then get

$$\frac{\sum_{t=1}^{T}\sum_{\tau=1}^{T}\Gamma_T(t,\tau)}{(T-h)^2} \le \frac{\bar{V}}{(T-h)},$$

where $\bar{V} = 75 \times 6\bar{C}^4(\bar{K}+2M_8)/(1-h/T)$ does not depend on κ_T (75 is the number of covariances in (B.1.2)). This establishes $\sum_{t=1}^{T}\chi_{T,t}(h,\kappa_T)^2 = 1 + o_{P_{\kappa_T}}(1)$. □

Lemma B.5. *Under the conditions of Lemma B.1,*

$$\lim_{T\to\infty}\sum_{t=1}^{T} E_{\kappa_T}\left[\chi_{T,t}^4\right] = 0.$$

Proof. Using the notation of Lemma B.4 and Loève's inequality,

$$E_N\left[\tilde{\chi}_{T,t}(h,\kappa_T)^4\right] \le 5^3\sum_{k=1}^{5} E_N\left[\zeta_{k,t}^4\right]. \tag{B.1.3}$$

Each of the five terms in (B.1.3) is under Assumption 3.3(i)–(iv) bounded by a constant that does not depend on κ_T. For $k = 1$,

$$\begin{aligned}
g\left(\frac{\kappa_T^4}{N^2}\right) E_N\left[\zeta_{1,t}^4\right] &= E_N\left[\left(\sum_{\ell=1}^{\infty} b_{t,\ell}X_tX_{t-\ell}\right)^4\right] \\
&\le \sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty}\sum_{\ell_3=1}^{\infty}\sum_{\ell_4=1}^{\infty}\left|b_{t,\ell_1}b_{t,\ell_2}b_{t,\ell_3}b_{t,\ell_4}\right|\left|E_N\left[X_t^4X_{t-\ell_1}X_{t-\ell_2}X_{t-\ell_3}X_{t-\ell_4}\right]\right| \\
&\le M_8\sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty}\sum_{\ell_3=1}^{\infty}\sum_{\ell_4=1}^{\infty}\left|b_{\ell_1}b_{\ell_2}b_{\ell_3}b_{\ell_4}\right| \le M_8\left(\sum_{\ell=1}^{\infty}|b_\ell|\right)^4 \le M_8\bar{C}^4,
\end{aligned}$$

where $\bar{C}$ is the constant we defined in the first part. The same bound works for $k = 2$ and $k = 3$ in (B.1.3).

For $k = 4$,

$$g\left(\frac{\kappa_T^4}{N^2}\right)\frac{E_N\left[\zeta_{4,t}^4\right]}{(\kappa_T^4/N^4)} = E_N\left[\left(\sum_{i=1}^{N}\sum_{\ell=1}^{\infty} d_{it,\ell}X_t u_{i,t-\ell}\right)^4\right]$$
$$\leq \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\sum_{i_4=1}^{N}\sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty}\sum_{\ell_3=1}^{\infty}\sum_{\ell_4=1}^{\infty} |d_{i_1t,\ell_1}d_{i_2t,\ell_2}d_{i_3t,\ell_3}d_{i_4t,\ell_4}|$$
$$\times\left|E_N\left[X_t^4 u_{i_1,t-\ell_1}u_{i_2,t-\ell_2}u_{i_3,t-\ell_3}u_{i_4,t-\ell_4}\right]\right|$$
$$\leq 3\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty} |d_{i_1t,\ell_1}^2 d_{i_2t,\ell_2}^2|\left|E_N\left[X_t^4 u_{i_1,t-\ell_1}^2 u_{i_2,t-\ell_2}^2\right]\right|$$
$$\leq 3N^2 M_8 \sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty} d_{\ell_1}^2 d_{\ell_2}^2 \leq 3N^2 M_8 \bar{C}^4,$$

where the second inequality uses that for $E_N\left[X_t^4 u_{i_1,t-\ell_1}u_{i_2,t-\ell_2}u_{i_3,t-\ell_3}u_{i_4,t-\ell_4}\right]$ to be non-zero we need $i_1 = i_2$ and $i_3 = i_4$, or $i_1 = i_3$ and $i_2 = i_4$, or $i_1 = i_4$ and $i_2 = i_3$ because of Assumptions 3.1(ii) and 3.2. The same bound applies to $k = 5$ in (B.1.3).

Putting these bounds together,

$$\sum_{t=1}^{T} E_N\left[\chi_{T,t}(h,\kappa_T)^4\right] = \frac{\sum_{t=1}^{T} E_N\left[\tilde{\chi}_{T,t}(h,\kappa_T)^4\right] g(\kappa_T^4/N^2)}{(T-h)^2 V(h,\kappa_T)^2} \leq \frac{9M_8\bar{C}^4 g(\kappa_T^4/N^2)}{(1-h/T)(T-h)V(h,\kappa_T^2)^2}.$$

Since $V(h,\kappa_T)^2/g(\kappa_T^4/N^2) \geq \underline{\mathrm{CM}}^2 > 0$, we conclude $\sum_{t=1}^{T} E_{\kappa_T}\left[\chi_{T,t}(h,\kappa_T)^4\right] = o(1)$ by iterated expectations where the convergence is uniform over κ_T. ◻

Lemma B.6. *Under the conditions of Lemma B.2,*

$$\sum_{t=1}^{T-h}\frac{X_t^2\xi_t(h,\kappa_T)^2 - E_{\kappa_T}\left[X_t^2\xi_t(h,\kappa_T)^2\middle|\{\theta_i,s_i\}_{i=1}^N\right]}{(T-h)g(\kappa_T^2/N)} \xrightarrow[P_{\kappa_T}]{\mathrm{p}} 0.$$

Proof. The proof is analogous to that of Lemma B.4. We will show that for a constant $\bar{V}$ independent of κ_T, $\mathrm{Var}_{N,\kappa_T}\left(\sum_{t=1}^{T} X_t^2\xi_t(h,\kappa_T)^2/g(\kappa_T^2/N)\right) \leq \bar{V}(T-h)$. By iterated expectations and Chebyshev's inequality it will follow that, for any $\varepsilon > 0$,

$$P_{\kappa_T}\left(\left|\sum_{t=1}^{T}\frac{X_t^2\xi_t(h,\kappa_T)^2 - E_{N,\kappa_T}\left[X_t^2\xi_t(h,\kappa_T)^2\right]}{(T-h)g(\kappa_T^2/N)}\right| > \varepsilon\right) \leq \frac{\bar{V}}{\varepsilon^2(T-h)} \to 0.$$

We can write

$$X_t\xi_t(h,\kappa_T) = \sum_{\ell=0}^{\infty}\iota_\ell(h)\bar{\beta}_\ell X_t X_{t+h-\ell} + \sum_{\ell=0}^{\infty}\bar{\gamma}_\ell X_t Z_{t+h-\ell} + \frac{\kappa_T}{N}\sum_{i=1}^{N}\sum_{\ell=0}^{\infty}\hat{s}_i\delta_{i\ell}X_t u_{i,t-\ell}$$

$$= \underbrace{\sum_{\ell=0}^{\infty} b_\ell X_t X_{t+h-\ell}}_{\equiv g(\kappa_T/\sqrt{N})\zeta_{1,t}} + \underbrace{\sum_{\ell=0}^{\infty} c_\ell X_t Z_{t+h-\ell}}_{\equiv g(\kappa_T/\sqrt{N})\zeta_{2,t}} + \underbrace{\frac{\kappa_T}{N}\sum_{i=1}^{N}\sum_{\ell=0}^{\infty} d_{i\ell} X_t u_{i,t+h-\ell}}_{\equiv g(\kappa_T/\sqrt{N})\zeta_{3,t}}. \tag{B.1.4}$$

for some coefficients $\{b_\ell, c_\ell, \{d_{i\ell}\}_{i=1}^N\}$ that depend on $\{\theta_i, s_i\}_{i=1}^N$ (and h). By Assumption 3.3(iv), we have $|b_\ell|, |c_\ell|, |d_{i\ell}| \leq C_\ell$ almost surely for some positive finite constants C_ℓ such that $C = \sum_{\ell=1}^{\infty} C_\ell < \infty$. Note that the coefficients, constants and variables $\zeta_{1,t}, \zeta_{2,t}, \zeta_{3,t}$ are different from the ones in the proof of Lemma B.4.

Consider the variance

$$\mathrm{Var}_{N,\kappa_T}\left(\sum_{t=1}^{T} \frac{X_t^2 \xi_t(h,\kappa_T)^2}{g(\kappa_T^2/N)}\right) = \sum_{t=1}^{T-h}\sum_{\tau=1}^{T-h} \Gamma_T(t,\tau)$$

where (omitting the dependence on h, κ_T and $\{\theta_i, s_i\}_{i=1}^N$)

$$\Gamma_T(t,\tau) = \mathrm{Cov}_{N,\kappa_T}\left(\frac{X_t^2\xi_t(h,\kappa_T)^2}{g(\kappa_T/\sqrt{N})}, \frac{X_\tau^2\xi_\tau(h,\kappa_T)^2}{g(\kappa_T/\sqrt{N})}\right).$$

As in the proof of Lemma B.4, we expand the square of $X_t^2\xi_t(h,\kappa_T)^2$ to express $\Gamma_T(t,\tau)$ as the sum of covariances $\Gamma_{T,k_1k_2k_3k_4}(t,\tau) = \mathrm{Cov}_{N,\kappa_T}\left(\zeta_{k_1,t}\zeta_{k_2,t}, \zeta_{k_3,\tau}\zeta_{k_4,\tau}\right)$ where k_1, k_2, k_3, k_4 range over the three terms in (B.1.4). If $k_1 = k_2$, $\Gamma_{T,k_1k_2k_3k_4}(t,\tau)$ can only be non-zero if $k_3 = k_4$, while if $k_1 \neq k_2$, only if either $k_1 = k_3$ and $k_2 = k_4$ or $k_1 = k_4$ and $k_2 = k_3$. Then,

$$\begin{aligned} \left|\Gamma_T(t,\tau)\right| &= \sum_{k_1=1}^{3}\sum_{k_2=1}^{3}\sum_{k_3=1}^{3}\sum_{k_4=1}^{3}\left|\Gamma_{T,k_1k_2k_3k_4}(t,\tau)\right| \\ &= \sum_{k_1=1}^{3}\sum_{k_3=1}^{3}\left|\Gamma_{T,k_1k_1k_3k_3}(t,\tau)\right| + 2\sum_{k_1=1}^{3}\sum_{k_2=1}^{3}\left|\Gamma_{T,k_1k_2k_1k_2}(t,\tau)\right|. \end{aligned} \tag{B.1.5}$$

By calculations similar to that of Lemma B.4, for any $k_1, k_2, k_3 \in \{1,2,3\}$,

$$\begin{aligned} \sum_{t=1}^{T-h}\sum_{\tau=1}^{T-h}\left|\Gamma_{T,k_1k_1k_3k_3}(t,\tau)\right| &\leq 6C^4(\bar{K}+2M_8)\times(T-h), \\ \sum_{t=1}^{T-h}\sum_{\tau=1}^{T-h}\left|\Gamma_{T,k_1k_2k_1k_2}(t,\tau)\right| &\leq 6C^4(\bar{K}+2M_8)\times(T-h). \end{aligned}$$

We therefore arrive at

$$\sum_{t=1}^{T}\sum_{\tau=1}^{T}\Gamma_T(t,\tau) \leq \bar{V}(T-h),$$

with $\bar{V} = 27 \times 6C^4(\bar{K}+2M_8)$ independent of κ_T (27 is the number of terms in (B.1.5)). Hence, $\{(T-h)g(\kappa_T^2/N)\}^{-1}\sum_{t=1}^{T}\left(X_t^2\xi_t(h,\kappa_T)^2 - E_{N,\kappa_T}\left[X_t^2\xi_t(h,\kappa_T)^2\right]\right) = o_{P_{\kappa_T}}(1)$. □

Lemma B.7. *Under the conditions of Lemma B.2,*

$$\sum_{t=1}^{T-h} \frac{\left[\left(N^{-1}\sum_{i=1}^{N}\hat{x}_{it}(h)\hat{\xi}_{it}(h)\right) - X_t\xi_t(h,\kappa_T)\right]^2}{(T-h)g(\kappa_T^2/N)} \xrightarrow[P_{\kappa_T}]{\text{p}} 0.$$

Proof. We begin by writing

$$\hat{x}_{it}(h) = \hat{s}_i(X_t - \hat{X}_t(h)), \quad \hat{X}_t(h) = \bar{X}_0(h) + \hat{\pi}_X(h)'\bar{x}_t(h), \tag{B.1.6}$$

with $\bar{X}_0(h)$, $\hat{\pi}_X(h)$ and $\bar{x}_t(h) = (X_{t-1} - \bar{X}_1(h), \ldots, X_{t-p} - \bar{X}_p(h))'$ as in the proof of Lemma B.3. As argued, $\bar{X}_0(h) = O_{P_{\kappa_T}}\left((T-h)^{-1/2}\right)$ and $\hat{\pi}_X(h) = O_{P_{\kappa_T}}\left((T-h)^{-1/2}\right)$.

Next, we write $\hat{\eta}(h)'W_{it} = \hat{\eta}_{0,i}(h) + \hat{\eta}_X(h)'\bar{x}_t(h)\hat{s}_i$ and $\eta_{X,ih} = (\beta_{i,h+1}, \ldots, \beta_{i,h+p})'$ so that

$$\hat{\xi}_{it}(h) - \xi_{it}(h,\kappa_T) = \left(\mu_i - \hat{\eta}_{0,i}(h) + \sum_{\ell=1}^{p}\beta_{i,h+\ell}\bar{X}_\ell(h)\right) + (\beta_{ih} - \hat{\beta}(h)\hat{s}_i)X_t + (\eta_{X,ih} - \hat{\eta}_X(h)\hat{s}_i)'\bar{x}_t(h)$$

and we note

$$\begin{aligned}\begin{pmatrix}\hat{\beta}(h)\\ \hat{\eta}_X(h)\end{pmatrix} &= \left[\sum_{t=1}^{T-h}\begin{pmatrix}X_t - \bar{X}_0(h)\\ \bar{x}_t(h)\end{pmatrix}\begin{pmatrix}X_t - \bar{X}_0(h)\\ \bar{x}_t(h)\end{pmatrix}'\right]^{-1}\sum_{t=1}^{T-h}\begin{pmatrix}X_t - \bar{X}_0(h)\\ \bar{x}_t(h)\end{pmatrix}\hat{Y}_{t+h}\\ &= \begin{pmatrix}\tilde{\beta}(h)\\ \tilde{\eta}_X(h)\end{pmatrix} + \left[\sum_{t=1}^{T-h}\begin{pmatrix}X_t - \bar{X}_0(h)\\ \bar{x}_t(h)\end{pmatrix}\begin{pmatrix}X_t - \bar{X}_0(h)\\ \bar{x}_t(h)\end{pmatrix}'\right]^{-1}\sum_{t=1}^{T-h}\begin{pmatrix}X_t - \bar{X}_0(h)\\ \bar{x}_t(h)\end{pmatrix}\frac{\xi_t(h,\kappa_T)}{(N^{-1}\sum_{i=1}^{N}\hat{s}_i^2)}\end{aligned}$$

where $\hat{Y}_{t+h} = (\sum_{i=1}^{N}\hat{s}_i^2)^{-1}\sum_{i=1}^{N}\hat{s}_i Y_{i,t+h}$ and $\tilde{\eta}_X(h) = (\sum_{i=1}^{N}\hat{s}_i^2)^{-1}\sum_{i=1}^{N}\hat{s}_i\eta_{X,ih}$. Since the least squares denominator matrix when scaled by $(T-h)^{-1}$ converges to $E\left[X_t^2\right] \times I_{p+1}$ in probability uniformly over κ_T, the calculations in Lemma 3.3 imply that

$$\begin{aligned}\frac{(N^{-1}\sum_{i=1}^{N}\hat{s}_i^2)(\hat{\beta}(h) - \tilde{\beta}(h))}{g(\kappa_T/\sqrt{N})} &= O_{P_{\kappa_T}}\left((T-h)^{-1/2}\right),\\ \frac{(N^{-1}\sum_{i=1}^{N}\hat{s}_i^2)(\hat{\eta}_X(h) - \tilde{\eta}_X(h))}{g(\kappa_T/\sqrt{N})} &= O_{P_{\kappa_T}}\left((T-h)^{-1/2}\right).\end{aligned}$$

Because W_{it} includes unit effects, $\sum_{i=1}^{N}\hat{x}_{it}(h)(\hat{\eta}_{0,i}(h) - \mu_i + \sum_{\ell=1}^{p}\beta_{i,h+\ell}\bar{X}_\ell(h)) = 0$ and,

$$\begin{aligned}N^{-1}\sum_{i=1}^{N}\hat{x}_{it}(h)(\hat{\xi}_{it}(h) - \xi_{it}(h,\kappa_T)) &= \left(N^{-1}\sum_{i=1}^{N}\hat{s}_i^2\right)(\tilde{\beta}(h) - \hat{\beta}(h))X_t(X_t - \hat{X}_t(h))\\ &+ \left(N^{-1}\sum_{i=1}^{N}\hat{s}_i^2\right)(\tilde{\eta}_X(h) - \hat{\eta}_X(h))'\bar{x}_t(h)(X_t - \hat{X}_t(h)).\end{aligned} \tag{B.1.7}$$

To prove the Lemma, add and subtract $N^{-1}\sum_{i=1}^{N}\hat{x}_{it}(h)\xi_{it}(h,\kappa_T)$ within the squares and use Loève's inequality to obtain

$$\sum_{t=1}^{T-h} \frac{\left[\left(N^{-1}\sum_{i=1}^{N}\hat{x}_{it}(h)\hat{\xi}_{it}(h)\right) - X_t\xi_t(h,\kappa_T)\right]^2}{(T-h)g(\kappa_T^2/N)} \leq 2D_{T,2}^{\pi}(h,\kappa_T) + 2D_{T,2}^{\eta}(h,\kappa_T),$$

where

$$D_{T,2}^{\pi}(h,\kappa_T) = \sum_{t=1}^{T-h} \frac{\left[N^{-1}\sum_{i=1}^{N}(\hat{s}_i X_t - \hat{x}_{it}(h))\xi_{it}(h,\kappa_T)\right]^2}{(T-h)g(\kappa_T^2/N)},$$

$$D_{T,2}^{\eta}(h,\kappa_T) = \sum_{t=1}^{T-h} \frac{\left[N^{-1}\sum_{i=1}^{N}\hat{x}_{it}(h)(\hat{\xi}_{it}(h) - \xi_{it}(h,\kappa_T))\right]^2}{(T-h)g(\kappa_T^2/N)}.$$

Inserting (B.1.6) into the first term and using Loève's inequality,

$$D_{T,2}^{\pi}(h,\kappa_T) \leq 2\left[\bar{X}_0(h)^2 \frac{\sum_{t=1}^{T-h}\xi_t(h,\kappa_T)^2}{(T-h)g(\kappa_T^2/N)} + \|\hat{\pi}_X(h)\|^2 \frac{\sum_{t=1}^{T-h}\|\bar{x}_t(h)\xi_t(h,\kappa_T)\|^2}{(T-h)g(\kappa_T^2/N)}\right],$$

where $\|\cdot\|$ is Euclidean norm. From calculations similar to those in Lemma B.3,

$$\frac{\sum_{t=1}^{T-h}\xi_t(h,\kappa_T)^2}{(T-h)g(\kappa_T^2/N)} = O_{P_{\kappa_T}}(1) \text{ and } \frac{\sum_{t=1}^{T-h}\|\bar{x}_t(h)\xi_t(h,\kappa_T)\|^2}{(T-h)g(\kappa_T^2/N)} = O_{P_{\kappa_T}}(1),$$

which allows us to conclude that $D_{T,2}^{\pi}(h,\kappa_T) = o_{P_{\kappa_T}}(1)$.

Inserting (B.1.7) into the second term and using Loève's inequality,

$$D_{T,2}^{\eta}(h,\kappa_T) \leq 2\left[\left(\frac{(N^{-1}\sum_{i=1}^{N}\hat{s}_i^2)(\tilde{\beta}(h) - \hat{\beta}(h))}{g(\kappa_T/\sqrt{N})}\right)^2 \frac{\sum_{t=1}^{T-h}X_t^2(X_t - \hat{X}_t(h))^2}{T-h}\right.$$
$$\left. + \left\|\left(\frac{(N^{-1}\sum_{i=1}^{N}\hat{s}_i^2)(\tilde{\eta}_X(h) - \hat{\eta}_X(h))}{g(\kappa_T/\sqrt{N})}\right)\right\|^2 \frac{\sum_{t=1}^{T-h}\|\bar{x}_t(h)(X_t - \hat{X}_t(h))\|^2}{T-h}\right].$$

Under Assumption 3.3(i), we can show that $(T-h)^{-1}\sum_{t=1}^{T-h}X_t^2(X_t - \hat{X}_t(h))^2 = O_{P_{\kappa_T}}(1)$ and $(T-h)^{-1}\sum_{t=1}^{T-h}\|x_t(h)(X_t - \hat{X}_t(h))\|^2 = O_{P_{\kappa_T}}(1)$. Thus, $D_{T,2}^{\eta}(h,\kappa_T) = o_{P_{\kappa_T}}(1)$. □

Proposition 3.2

Parts (A), (B) and (C) of the proof of Proposition 3.2 in Appendix A.1 are established in Lemmas B.8, B.9 and B.10 below. The argument closely resembles the proof of Proposition 3.1 and, therefore, in order to conserve space we only sketch the steps. Again, we adopt Assumptions 3.1, 3.2 and 3.3, we fix p and assume $h_T/T \leq \phi < 1$ as $T, N \to \infty$.

Lemma B.8 (Asymptotic normality of the score).

$$\frac{\sum_{t=1}^{T-h_T} X_t \xi_t(h_T, \kappa_T)}{\sqrt{(T-h_T)V(h_T,\kappa_T)}} \xrightarrow[P_{\kappa_T}]{\text{d}} N(0,1).$$

Proof. The proof given for Lemma B.1 goes through with the following adjustment: we can remove the terms $\bar{\beta}_\ell, \bar{\gamma}_\ell, \delta_{i\ell}$ from $\Xi_{X,t}(h,\kappa)$ whenever $\ell > h$. That is, we set

$$\Xi_{X,t}(h,\kappa) = \sum_{\ell=1}^{h} \mathbb{1}\{t-\ell \geq 1\} \bar{\beta}_{h-\ell} X_{t-\ell} + \mathbb{1}\{t \leq T-h\} \left[\bar{\gamma}_h Z_t + \frac{\kappa}{N} \sum_{i=1}^{N} \hat{s}_i \delta_{ih} u_{it} \right].$$

The calculations in Lemmas B.4 and B.5 apply with the same adjustment. In Lemma B.4, $\bar{V} \leq 75 \times 6 C^8 (\bar{K} + 2M_8)/(1-\phi)$, which does not depend on κ_T or h_T. Similarly, in Lemma B.5, $\sum_{t=1}^{T} E_N \left[\chi_{T,t}(h_T,\kappa_T)^4 \right] \leq 9 M_8 C^8 / (1-\phi)^2 \underline{\mathrm{CM}}^2 T$, which tends to zero as $T \to \infty$ uniformly over κ_T and h_T. □

Lemma B.9 (Consistency of the standard error).

$$\frac{\hat{V}(h_T)}{V(h_T,\kappa_T)} \xrightarrow[P_{\kappa_T}]{\text{p}} 1.$$

Proof. The proofs of Lemma B.2 and auxiliary Lemma B.6 go through without change. To establish the equivalent to Lemma B.7 in this context, define $\bar{x}_t(h_T)$ as in its proof and let $\bar{y}_{it}(h_T) = (\hat{Y}_{i,t-1}(h_T), \ldots, \hat{Y}_{i,t-p}(h_T))$ with $\hat{Y}_{i,t-\ell}(h_T)$ the residual from regressing $g(\kappa_T)^{-1} Y_{i,t-\ell}$ on unit and time effects. We can write

$$\hat{\pi}(h_T)' W_{it} = \hat{s}_i \bar{X}_0(h_T) + \hat{s}_i \hat{\pi}_X(h_T)' \bar{x}_t(h_T) + \hat{\pi}_Y(h_T)' \bar{y}_{it}(h_T),$$
$$\hat{\eta}(h_T)' W_{it} = \hat{\eta}_{0,i}(h_T) + \hat{s}_i \hat{\eta}_X(h_T)' \bar{x}_t(h_T) + \hat{\eta}_Y(h_T)' \bar{y}_{it}(h_T).$$

Scaling $Y_{i,t-\ell}$ by $g(\kappa_T)^{-1}$ leaves the least square predictions $\hat{\pi}(h_T)' W_{it}$ and $\hat{\eta}(h_T)' W_{it}$ unchanged, but it helps bound them in probability uniformly over κ_T.

Calculations similar to those in Lemma B.3 deliver

$$\begin{pmatrix} \bar{X}_0(h_T) \\ \hat{\pi}_X(h_T) \\ \hat{\pi}_Y(h_T) \end{pmatrix} = O_{P_{\kappa_T}}\left((T-h_T)^{-1/2} \right),$$

$$g\left(\frac{\kappa_T}{\sqrt{N}}\right)^{-1} \begin{pmatrix} (\hat{\beta}(h_T) - \tilde{\beta}(h_T)) \\ (\hat{\eta}_X(h_T) - \tilde{\eta}_X(h_T)) \\ (\hat{\eta}_Y(h_T) - \tilde{\eta}_Y(h_T)) \end{pmatrix} = O_{P_{\kappa_T}}\left((T-h_T)^{-1/2} \right),$$

where $\tilde{\eta}_X(h_T) = (\tilde{B}_1(h_T), \ldots, \tilde{B}_p(h_T))'$ and $\tilde{\eta}_Y(h_T) = g(\kappa_T)(A_1(h_T), \ldots, A_p(h_T))'$ with $A_\ell(h)$ and $\tilde{B}_\ell(h)$ as defined in the proof of Proposition 3.2 in Appendix A.1.

The rest of the proof follows the steps of Lemma B.7. The convergence is uniform in both κ_T and h_T because $T - h_T \leq (1-\phi)T$ with $\phi < 1$. □

Lemma B.10 (Negligibility of the remainder).

$$R_T(h_T, \kappa_T) \xrightarrow[P_{\kappa_T}]{\text{p}} 0.$$

Proof. We begin by defining $\bar{x}_t(h_T)$ and $\bar{y}_{it}(h)$ as in Lemma B.9, by writing

$$\hat{\pi}(h_T)' W_{it} = \hat{s}_i \bar{X}_0(h_T) + \hat{s}_i \hat{\pi}_X(h_T)' \bar{x}_t(h_T) + \hat{\pi}_Y(h_T)' \bar{y}_{it}(h_T),$$

and by noting again that

$$\begin{pmatrix} \bar{X}_0(h_T) \\ \hat{\pi}_X(h_T) \\ \hat{\pi}_Y(h_T) \end{pmatrix} = O_{P_{\kappa_T}}\left((T - h_T)^{-1/2}\right).$$

Next, we write $r_{it}(h_T) = (\beta_{ih} - \bar{\beta}(h)\hat{s}_i)X_t + \sum_{\ell=1}^{p}(B_{i\ell}(h) - \bar{B}_\ell(h)\hat{s}_i)X_{t-\ell}$ and

$$R_T(h_T, \kappa_T) = -\frac{\bar{X}_0(h_T) \sum_{t=1}^{T-h_T} \xi_t(h_T, \kappa_T)}{\sqrt{(T - h_T) V(h_T, \kappa_T)}} - \frac{\hat{\pi}_X(h_T)' \sum_{t=1}^{T-h_T} \bar{x}_t(h_T) \xi_t(h_T, \kappa_T)}{\sqrt{(T - h_T) V(h_T, \kappa_T)}}$$
$$- \frac{\hat{\pi}_Y(h_T)' \sum_{i=1}^{N} \sum_{t=1}^{T-h_T} \bar{y}_{it}(h_T)(r_{it}(h_T) + \xi_{it}(h_T, \kappa_T))}{N\sqrt{(T - h_T) V(h_T, \kappa_T)}}$$

The rest of the argument mimics the proof of Lemma B.3. □

Proposition 3.3

Parts (A), (B) and (C) of the proof of Proposition 3.3 in Appendix A.1 are stated in Lemmas B.11, B.12 and B.13 below. The proofs are virtually identical to their counterparts in Proposition 3.1 with some minor differences. Here we make Assumptions 3.4 and we hold h and $p \geq h$ fixed as $T, N \to \infty$.

Lemma B.11 (Asymptotic normality of the score).

$$\frac{\sum_{t=1}^{T-h} \boldsymbol{\lambda}' \boldsymbol{X}_t^* \xi_t(h, \kappa_T)}{\sqrt{(T - h)\boldsymbol{\lambda}' \boldsymbol{V}(h, \kappa_T) \boldsymbol{\lambda}}} \xrightarrow[P_{\kappa_T}]{\text{d}} N(0, 1).$$

Proof. The arguments given for Lemma B.1 and auxiliary Lemmas B.4 and B.5 apply with the obvious change in notation. □

Lemma B.12 (Consistency of the standard error and OLS denominator).

$$\frac{\boldsymbol{\lambda}' \hat{\boldsymbol{V}}^{\text{IV}}(h) \boldsymbol{\lambda}}{\boldsymbol{\lambda}' \boldsymbol{V}(h, \kappa_T) \boldsymbol{\lambda}} \xrightarrow[P_{\kappa_T}]{\text{p}} 1 \text{ and } \hat{\boldsymbol{J}}^{\text{IV}}(h) \xrightarrow[P_{\kappa_T}]{\text{p}} \boldsymbol{J}.$$

Proof. The first part follows from arguments analogous to those given for Lemma B.2 and auxiliary Lemmas B.6 and B.7 (with obvious notational changes). For the second part, note $\text{Var}_{N,\kappa_T}\left(X_t^* \tilde{X}_t\right) \leq \bar{V}/(T - h)$ for some constant $\bar{V}$ independent of κ_T under Assumption 3.4(ii), so that $\left\|\hat{\boldsymbol{J}}^{\mathrm{IV}}(h) - \boldsymbol{J}\right\| = o_{P_{\kappa_T}}(1)$ follows from iterated expectations and Chebyshev's inequality. □

Lemma B.13 (Negligibility of the remainder).

$$R_T(h, \kappa_T) \xrightarrow[P_{\kappa_T}]{\mathrm{p}} 0.$$

Proof. For any $\boldsymbol{\lambda} \neq 0_{(p+1)\times 1}$, by the same calculations as in Lemma B.3,

$$\frac{\sum_{t=1}^{T-h} \boldsymbol{\lambda}' \boldsymbol{X}_t^*}{(T-h)} = O_{P_{\kappa_T}}\left((T-h)^{-1/2}\right) \text{ and } \frac{\sum_{t=1}^{T-h} \xi_t(h, \kappa_T)}{\sqrt{(T-h)\boldsymbol{\lambda}' \boldsymbol{V}(h, \kappa_T)\boldsymbol{\lambda}}} = O_{P_{\kappa_T}}(1)\,.$$

Since $\hat{\boldsymbol{J}}^{\mathrm{IV}}(h) = \boldsymbol{J} + o_{P_{\kappa_T}}(1)$ by the second part of Lemma B.12, the result follows. □

B.2 Details of simulation study

Here we complement Section 3.4 with additional details. First, we describe how we simulate the heterogeneity. Second, we specify the calibration of our DGPs. Third and last, we present further simulation results.

Simulation of observable and unobservable heterogeneity. A primary feature is the correlation between s_i and $\{\beta_i, \gamma_i.\delta_i\}$.[3] We begin by drawing the vector

$$(s_i, s_{\gamma,i}, s_{\delta,i})' \sim N\left(1_{3\times1}, (1-\rho)I_3 + \rho 1_{3\times3}\right)$$

for some $\rho \neq 0$. Next, we set a very large $\bar{L}$ and compute

$$\beta_{i\ell} = s_i\breve{\beta}_{i\ell}, \quad \gamma_{i\ell} = s_{\gamma,i}\breve{\gamma}_{i\ell}, \quad \delta_{i\ell} = s_{\delta,i}\breve{\delta}_{i\ell},$$

where $\{\breve{\beta}_{i\ell}, \breve{\gamma}_{i\ell}, \breve{\delta}_{i\ell}\}_{\ell=0}^{\bar{L}}$ are obtained by (a) drawing the roots of ARMA polynomials from Beta distributions, (b) computing their MA(∞) representations, (c) truncating them at $\bar{L}$, and (d) normalizing them so that $\sum_{\ell=0}^{\bar{L}}\breve{\beta}_{i\ell}^2 = \sum_{\ell=0}^{\bar{L}}\breve{\gamma}_{i\ell}^2 = \sum_{\ell=0}^{\bar{L}}\breve{\delta}_{i\ell}^2 = 1$.[4]

To generate time-varying heterogeneity we set $s_{it} = s_i + \zeta_{it}$ with $\zeta_{it} \sim N(0,1)$, i.i.d. over units and time, and independent of s_i and everything else. This ensures s_{it} remains exogenous with respect to aggregate and idiosyncratic shocks.

Finally, in the VAR DGP, we set

$$B_{i\ell} = s_i\breve{B}_{i\ell}, \quad C_{i0} = s_{\gamma,i}, \quad D_{i0} = s_{\delta,i}.$$

where $\{\breve{B}_{i\ell}\}_{\ell=0}^{\bar{L}}$ are obtained in the same way as $\{\breve{\beta}_{i\ell}\}_{\ell=0}^{\bar{L}}$ above.

Our method does not satisfy Assumption 3.3(iv), although responses are bounded with sufficiently high probability that it does not seem to make a difference.

DGP calibration. In the general DGP, we set $\rho = 0.5$, and generate $\{\breve{\beta}_{i\ell}, \breve{\gamma}_{i\ell}, \breve{\delta}_{i\ell}\}_{\ell=0}^{\bar{L}}$ from ARMA(4, 2) processes with expected roots $(0.7, 0.3, 0.2, 0.1)$ and $(0, 0)$ for $\breve{\beta}_{i\ell}$, $(0.7, 0.2, 0.1, -0.2)$ and $(0.2, -0.2)$ for $\breve{\gamma}_{i\ell}$, and $(0.9, 0.3, 0.1, 0.1)$ and $(0.5, 0.2)$ for $\breve{\delta}_{i\ell}$. We draw each root as Beta$(\bar{\lambda}\nu, (1-\bar{\lambda})\nu)$ where $\bar{\lambda}$ is the mean listed above and $\nu = 10$, and we truncate polynomials at $\bar{L} = 2T$ lags.

In the LP-IV case, we use a similar method for $\{b_\ell, c_\ell\}_{\ell=0}^{\bar{L}}$. We obtain b_ℓ from an ARMA(1, 1) with roots 0.3 and -0.2, and c_ℓ from an ARMA(2, 2) with roots $(0.4, 0.2)$ and $(0.1, -0.1)$. We also set $a_0 = 10$ to be

[3] Instead, μ_i (and m_i in the VAR setup) does not play a big role and we simply draw it as $N(0,1)$.

[4] The advantage of this representation is that it separates the scale and persistence. For example, if X_t is white noise with unit variance conditional on $\{\beta_{i\ell}\}_{\ell=0}^{\bar{L}}$, the variance of $\sum_{\ell=0}^{\bar{L}}\beta_{i\ell}X_{t-\ell}$ is $\sum_{\ell=0}^{\bar{L}}\beta_{i\ell}^2 = s_i^2$ while the ratio of long-run variance to variance of $\sum_{\ell=0}^{\bar{L}}\beta_{i\ell}X_{t-\ell}$ (a measure of persistence) is

$$\frac{\left(\sum_{\ell=0}^{\bar{L}}\beta_{i\ell}\right)^2}{\sum_{\ell=0}^{\bar{L}}\beta_{i\ell}^2} = \frac{\left(\sum_{\ell=0}^{\bar{L}}\breve{\beta}_{i\ell}\right)^2}{\sum_{\ell=0}^{\bar{L}}\breve{\beta}_{i\ell}^2},$$

which does not depend on s_i.

safely above standard weak IV thresholds.

Finally, for the VAR DGP, we draw $\{\breve{B}_{i\ell}\}_{\ell=0}^{p}$ from an MA(2) with roots $(0.8, -0.5)$ and $\nu = 10$, and we set $\{A_\ell\}_{\ell=1}^{p}$ to an AR(2) with roots $(1 - 5/T, 0.5)$.

The mean and quantiles of responses for each horizon can be seen in Figure B.2.1.

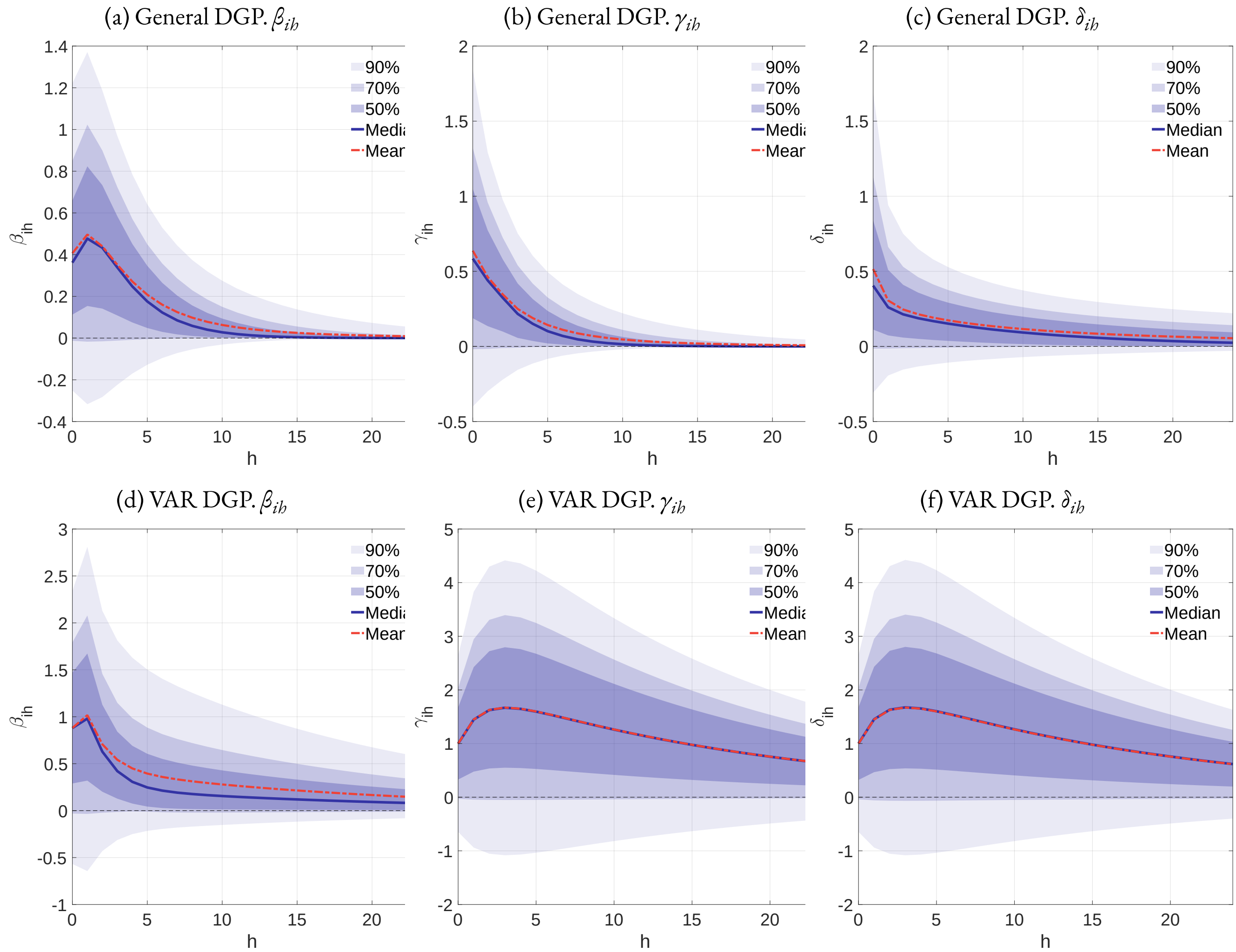


FIGURE B.2.1. Distributions of impulse responses for general and VAR DGPs.

Additional results. Figure B.2.2 presents coverage rates of 90% confidence intervals in the general DGP with $T = 100$ for panel LPs on X_t (panels (a)-to-(c)) and on $s_{it}X_t$ (panels (d)-to-(f)).[5] As mentioned in Chapter 3, the estimands are different: LPs on X_t recover the mean impulse response while LPs on $s_{it}X_t$ recover their projection on s_{it}. Yet, the observations we made about inference from Section 3.4 are unchanged. In particular, t-LAHR inference dominates all the alternatives in delivering correct coverage for the nonparametric panel local projection estimand.

[5]For panel LPs on X_t time effects are excluded from the vector of controls. Otherwise, the estimation and inference procedures are the same as in Figure 3.1 in Chapter 3.

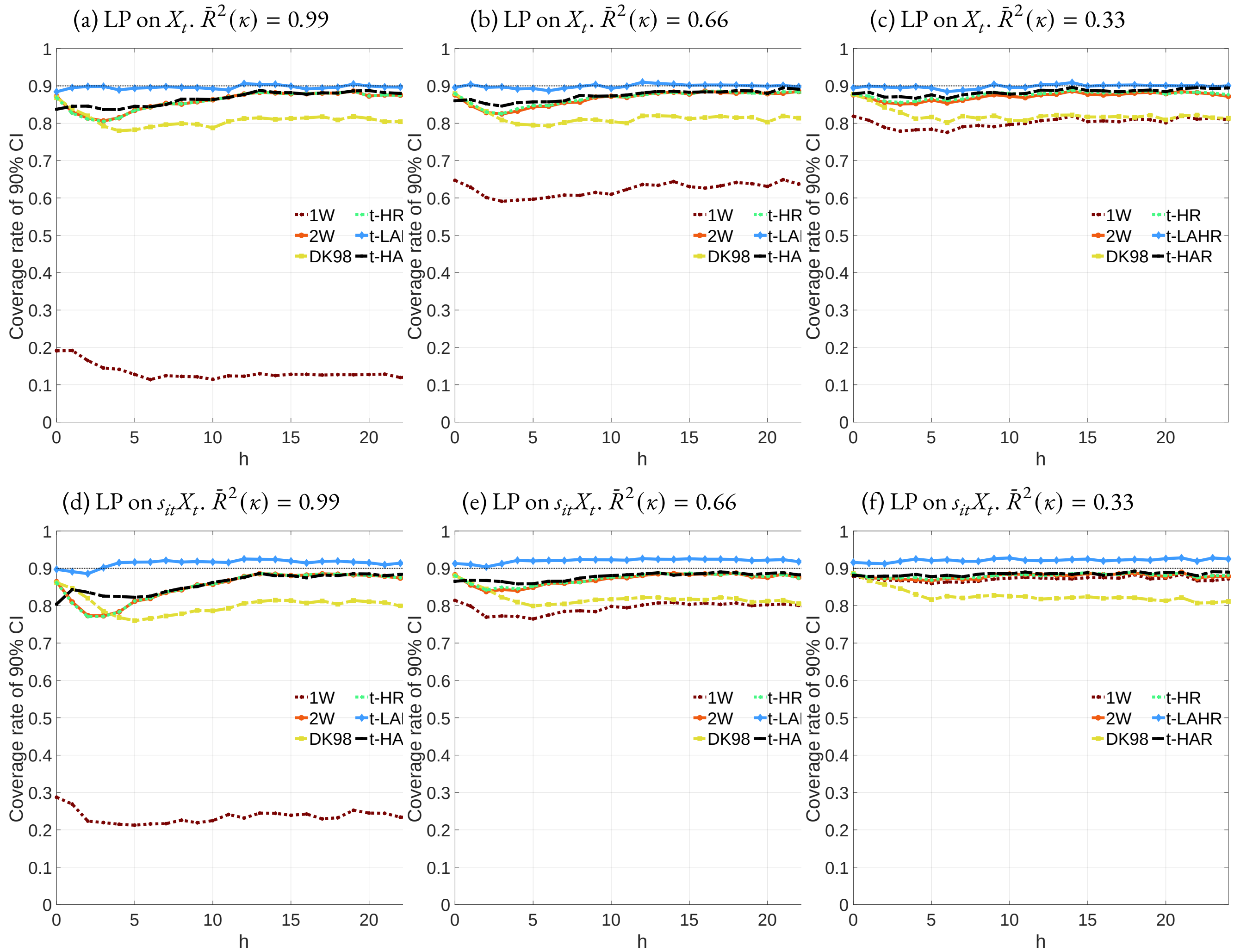


FIGURE B.2.2. Coverage rates of 90% confidence intervals for $T = 100$.

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

B.3 A survey of empirical applications

Below, we survey relevant empirical applications by the method used to calculate standard errors. The list reflects the recent surge in applications (with the oldest paper dated 2018) and includes both published work and working papers. We have aimed to make the list comprehensive, but it is possible that some might have been inadvertently omitted. When different methods were used, we favored the one used in the main specification and the one used in estimation of dynamic effects (non-zero horizons). We classified as one-way clustering (within units) applications that cluster at a higher level of aggregation than primary units; say, at the industry (or industry-time) level when units are firms. While allowing for sector-level shocks, these still rule out economy-wide spatial dependence. See the Introduction for additional details.

By method

Two-way clustering (within units and time)	Ippolito, Ozdagli, and Perez-Orive (2018), Jeenas (2019), Ottonello and Winberry (2020), Amberg, Jansson, Klein, and Rogantini Picco (2022), Palazzo and Yamarthy (2022), Paz (2022), Belifemine, Couturier, and Jamilov (2023), Cascaldi-Garcia, Vukotić, and Zubairy (2023), Drechsel (2023), Durante, Ferrando, and Vermeulen (2022), Duval, Furceri, Lee, and Tavares (2023), Ferreira, Ostry, and Rogers (2023), González, Nuño, Thaler, and Albrizio (2023), Lakdawala and Moreland (2023), Singh, Suda, and Zervou (2023), Thürwächter (2023), Zhou (2023), Anderson and Cesa-Bianchi (2024), Berthold, Cesa-Bianchi, Di Pace, and Haberis (2024), Caglio, Darst, and Kalemli-Özcan (2024), Camêlo (2024), Gulyas, Meier, and Ryzhenkov (2024), Paranhos (2024), Lakdawala and Moreland (forthcoming)
Clustering within units	Wu (2018), Ozdagli (2018), Crouzet and Mehrotra (2020), Singh, Suda, and Zervou (2022), Albrizio, González, and Khametshin (2023), Andersen, Johannesen, Jørgensen, and Peydró (2023), Camara and Ramirez Venegas (2023), Ghomi (2023), Indarte (2023), Bardóczy, Bornstein, Maggi, and Salgado (2024), Jeenas (2024), Jeenas and Lagos (2024), Lo Duca, Moccero, and Parlapiano (2024), Paranhos (2024), Ruzzier (2024)
Driscoll and Kraay (1998) standard errors	Holm, Paul, and Tischbirek (2021), Bahaj, Foulis, Pinter, and Surico (2022), Cloyne, Ferreira, Froemel, and Surico (2023), Fagereng, Gulbrandsen, Holm, and Natvik (2023), Gorea, Kryvtsov, and Kudlyak (2023), Bilal and Känzig (2024), Cao, Hegna, Holm, Juelsrud, König, and Riiser (2024)
Clustering within time	Gürkaynak, Karasoy-Can, and Lee (2022)

APPENDIX C

APPENDIX TO CHAPTER 4

C.1 Data

C.1.1 India – sample selection and validation

This section describes sample selection and validation of the subjective expectations data in detail. For the most part, it mirrors the analysis in Section 1 in Attanasio and Augsburg (2016). However, we update some of the criteria used and review additional sample selection decisions needed.

Tables C.1.1 and C.1.2 can be understood as an extended version of Tables 2 and 3 (respectively) reported in Attanasio and Augsburg (2016).

Table C.1.1 details basic sample selection steps and resulting sample sizes. For instance, we exclude observations with missing income or at least one reported probability. We also report the number of households for which either the elicited subjective lower or upper bound on future income is missing, although we do not exclude these from the final sample.[1] We also exclude from the sample some extreme reports of current household income under the category "implausible income", which are likely to correspond to survey measurement error.[2]

The total number of unique households drops to 930 and 877 in the first and second rounds, respectively. Relative to these, we study bunching of the reported probabilities at the 0%, 50% and 100% marks in Table C.1.2, and note substantial bunching at 50% for the midpoint (especially in the first round).

Keeping only households present in both rounds implies a balanced panel with 789 unique households. We further drop households who report at least one probability as equal to zero or one or whose elicited subjective *cdf* is not *strictly* monotonic. This further reduces the final sample size to $N = 770$ households.

[1] A total of 1,041 households were originally interviewed in the first round. We drop five observations who were asked about monthly rather than yearly income. In the remaining rows, minor differences with respect to those reported in Attanasio and Augsburg (2016) are due to slightly different/updated criteria.

[2] In particular, these correspond to reports outside the range $[0.5 \times rmin, 2 \times rmax]$, where *rmin* and *rmax* are the reported lower and upper bound on subjective income distributions (respectively), and are replaced by r_A and r_C, respectively, if missing. Unlike in the Colombian data, as reported in Appendix C.1.2, using looser or more strict "cutoffs" does not lead to substantial changes in sample sizes.

We present robustness checks keeping those households in Appendix C.4.1.

TABLE C.1.1. India: response rates and sample sizes

	Round 1	Round 2
Total number of observations	1036	947
Missing income	2	1
Missing either Min or Max	11	11
Missing at least one probability	22	11
Wrong — direction	3	5
Wrong — violation of monotonicity	9	19
Implausible thresholds	63	2
Implausible income	22	41
Available observations	926	873
Balanced panel (robustness)	789	
At least one probability is 0 or 1	9	11
At least two prob. are equal	3	2
Balanced panel (final)	770	

Note. "Wrong — direction" refers to households that in a given survey report $\Pr(y_{t+1} \geq r_C) > \Pr(y_{t+1} \geq r_B) > \Pr(y_{t+1} \geq r_A)$, among those with no missing probabilities. "Wrong — violation of monotonicity" refers to weak monotonicity violations. "Implausible thresholds" refers to households for which r_A, r_B or r_C is missing, among those who report no missing probabilities, households for which $r_B < r_A$ or $r_C < r_B$, and households with implausibly large interval differences between $r_B - r_A$ and $r_C - r_B$. "Implausible income" refers to households that report income outside $[0.5 \times rmin, 2 \times rmax]$, where *rmin* and *rmax* are replaced by r_A and r_C, respectively, if missing.

TABLE C.1.2. India: bunching

	Round 1	Round 2
Threshold A (Lower)		
0%	6	10
50%	1	20
100%	2	0
Threshold B (Midpoint)		
0%	0	0
50%	503	142
100%	2	0
Threshold C (Higher)		
0%	0	0
50%	44	102
100%	3	1
Available observations	926	873

Note. The table shows the number of respondents who reported 0%, 50% and 100% probabilities in each survey round.

C.1.2 Colombia – sample selection and validation

We repeat the same analysis as for the Indian data (Appendix C.1.1) here. Tables C.1.3 and C.1.4 summarize sample selection decisions and validation of the expectations data and bunching, respectively.

As shown in Table C.1.3, of the original sample of 11,462 households interviewed during the 2002 baseline survey, 10,743 were re-interviewed in 2003 and 9,463 in 2005/06, of which 9,221 provided information on household income during the first survey round and 7,517 during the second. The fraction of households with missing reported probabilities or extreme values household income is also larger than in India.[3] Relative to the analysis for India, here we include an additional step on "missing covariates", where we exclude a few households for which covariates used in the main analysis (village, number of income sources and the proportion of income from farming sources) were missing.

The remaining rows in Table C.1.3 and Table C.1.4 correspond to the validation exercise on the subjective expectations data, similar to the one performed for India in Appendix C.1.1 and in Attanasio and Augsburg (2016). The final balanced sample we use in the main analysis has $N = 2,230$ unique households, after excluding those who report probabilities equal to zero or one or answers that violate strict monotonicity of (subjective) *cdf* s. We report our main results keeping these observations in Appendix C.4.2. Finally, Tables C.1.5 and C.1.6 report differences in observable characteristics between households in the final dataset and those available but excluded at after validation.

Since the subjective expectations data in Colombia has not been used before, we now elaborate on validation:

- *Logical response errors*. Table C.1.3 shows that in the first survey round 350 households provided answers that violated monotonicity and 37 provided responses that adhered to monotonicity but were "inverted", in that probabilities were non-decreasing, rather than non-increasing.[4] These figures imply violations of around 4% of those that gave responses to the probabilities, somewhat higher than in the Indian context where the logical response error was around 1%, but comparable to other studies. Dominitz and Manski (1997a), for example, report violations for almost 5% of their sample, and almost twice that number when including respondents where prompting happened (in that their responses would have initially been classified as logical response errors, but changed responses after having been prompted; such prompting was not allowed in either of the contexts considered here).

- *Bunching of percentages*. Table C.1.4 reports on the extent to which households bunch at the 0%, 50%, or 100% probability marks for the different thresholds. In the first round, there is substantial bunching at 0% for the lowest and 100% for the highest thresholds (around 14% of responses, compared to a negligible amount in India), which suggests some households might not have understood the prompts correctly or that the elicited expected income range is not accurate. There is also apparent bunching at 50% for the midpoint, a common feature with subjective probability data also present in the Indian data. There is a more muted presence of these issues in the second wave data.

[3] In the Colombian data, using looser or more strict rules for "implausible income" (described in Table C.1.3) leads to substantially larger and smaller sample sizes, respectively. For instance, using an interval given by $[0.2 \times rmin, 5 \times rmax]$ allows us to keep approximately around 200 additional unique households. Naturally, repeating the analysis with this rather permissive rule tends to exaggerate the features (nonlinear persistence, skewness, etc) we study.

[4] Recall that, as described in Figure 4.1, households were asked to report probabilities of the form $\Pr(y_{t+1} \geq r)$.

Even though these data display a higher degree of logical response errors and bunching than in India, we conclude that the responses provided in the Colombian data conform for the most part to the basic probability laws, and seem to suggest substantial coherency and variability, in that most respondents appear to have understood the instructions and provided thoughtful responses.

TABLE C.1.3. Colombia: response rates and sample sizes

	Round 1	Round 2
Total number of observations	10743	9463
Missing income	1522	1946
Missing either Min or Max	1294	958
Missing at least one probability	1361	964
Wrong — direction	37	33
Wrong — violation of monotonicity	350	291
Implausible thresholds	24	21
Implausible income	633	390
Available observations	7262	6295
Missing covariates	38	25
Balanced panel (robustness)	4420	
At least one probability is 0 or 1	2005	1434
At least two elicited probabilities are equal	866	600
Balanced panel (final)	2230	

Note. "Wrong — direction" refers to households that in a given survey report $\Pr(y_{t+1} \geq r_C) > \Pr(y_{t+1} \geq r_B) > \Pr(y_{t+1} \geq r_A)$, among those with no missing probabilities. "Wrong — violation of monotonicity" refers to weak monotonicity violations. "Implausible thresholds" refers to households for which r_A, r_B or r_C is missing, among those who report no missing probabilities, households for which $r_B < r_A$ or $r_C < r_B$, and households with implausibly large interval differences between $r_B - r_A$ and $r_C - r_B$. "Implausible income" refers to households that report income outside $[0.5 \times rmin, 2 \times rmax]$, where *rmin* and *rmax* are replaced by r_A and r_C, respectively, if missing.

TABLE C.1.4. Colombia: bunching

	Round 1	Round 2
Threshold A (Lower)		
0%	1041	1036
50%	712	500
100%	83	23
Threshold B (Midpoint)		
0%	201	184
50%	806	762
100%	274	88
Threshold C (Higher)		
0%	85	36
50%	549	524
100%	1139	545
Available observations	7262	6295

Note. The table shows the number of respondents who reported 0%, 50% and 100% probabilities in each survey round.

TABLE C.1.5. Colombia: covariate balance (wave 1)

	(0)		(1)		(0)-(1)
Variable	*N*	Mean	*N*	Mean	
Number of adults	3670	2.73	2230	2.69	0.03
Number of female adults	3670	1.39	2230	1.38	0.01
Number of kids	3670	3.11	2230	3.25	-0.15^{***}
Log income	3670	12.74	2230	12.73	0.02
Rural household	3661	0.46	2225	0.45	0.01
Household head:					
Age	3657	44.30	2228	43.60	0.70^{**}
Some primary education	3612	0.43	2213	0.45	−0.02
Some secondary education	3612	0.15	2213	0.15	−0.01
Primary source of income:					
Laborer/employee	3464	0.29	2107	0.27	0.02
Domestic employee	3464	0.06	2107	0.05	0.00
Day laborer	3464	0.21	2107	0.23	−0.02
Self-employment	3464	0.39	2107	0.38	0.00
Partner in farm/plot	3464	0.06	2107	0.07	−0.01
Proportion of regular income	3664	0.79	2230	0.79	0.00
Health shocks	3670	0.13	2230	0.12	0.01
Other shocks	3670	0.14	2230	0.13	0.01

Note. (1) refers to observations in the final sample and (0) to available observations excluded from the final sample (just before "Balanced panel (robustness)" in Table C.1.3). First wave only. Robust standard errors; ***=.01, **=.05, *=.1.

TABLE C.1.6. Colombia: covariate balance (wave 2)

	(0)		(1)		(0)-(1)
Variable	*N*	Mean	*N*	Mean	
Number of adults	3118	2.80	2230	2.69	0.11^{***}
Number of female adults	3118	1.43	2230	1.38	0.05^{**}
Number of kids	3118	3.18	2230	3.25	−0.07
Log income	3118	12.73	2230	12.76	−0.02
Rural household	3106	0.47	2225	0.45	0.02
Household head:					
Age	3096	46.42	2216	45.39	1.03^{***}
Some primary education	3027	0.45	2165	0.45	0.01
Some secondary education	3027	0.15	2165	0.18	-0.03^{***}
Primary source of income:					
Laborer/employee	2990	0.31	2162	0.32	−0.01
Domestic employee	2990	0.05	2162	0.04	0.01^{**}
Day laborer	2990	0.26	2162	0.26	0.00
Self-employment	2990	0.32	2162	0.32	0.00
Partner in farm/plot	2990	0.06	2162	0.06	0.00
Proportion of regular income	3115	0.90	2230	0.91	-0.02^{***}
Health shocks	3118	0.15	2230	0.13	0.02^{*}
Other shocks	3118	0.22	2230	0.18	0.04^{***}

Note. (1) refers to observations in the final sample and (0) to available observations excluded from the final sample (just before "Balanced panel (robustness)" in Table C.1.3). Second wave only. Robust standard errors; ***=.01, **=.05, *=.1.

C.2 Methodological appendix

Recall the flexible model in equation (4.9), which we reproduce below for convenience:

$$\ell_{jit} = \beta_0(r_{jit}) + \beta_1(r_{jit})\psi\left(y_{it}\right) + \beta_2(r_{jit})\eta_i + \varepsilon_{jit}. \tag{C.2.1}$$

We now provide further details on parameterization (Section C.2.1), estimation (Sections C.2.2 and C.2.3) and implementation (Section C.2.4).

C.2.1 Specification details

We view model (C.2.1) as a sequence of approximating parameter spaces — or sieves — for the nonparametric model in (4.3)-(4.4); see Chen (2007) for a technical review of the method of sieves.

In particular, we parameterize model (C.2.1) as

$$\ell_{jit} = \sum_{\tau=1}^{K_0} \beta_{0,\tau} h_\tau(r_{jit}) + \sum_{\tau=1}^{K_1}\sum_{\kappa=1}^{K_y} \beta_{1,\tau,\kappa} h_\tau(r_{jit}) g_\kappa(y_{it}) + \sum_{\tau=1}^{K_2} \beta_{2,\tau} h_\tau(r_{jit})\eta_i + \varepsilon_{jit}, \tag{C.2.2}$$

where $g_\kappa(y)$ are Hermite polymomials (we omit the constant term, i.e., g_1 is linear in y) and $h_\tau(r)$ are basis functions of natural cubic splines, with K_s knots ($K_s \geq 2$) for $s = \{0, 1, 2\}$.[5] For the ease of notation, we use L instead of K_s in Chapter 4, and explicitly refer to $\beta_0(\cdot)$, $\beta_1(\cdot)$ and/or $\beta_2(\cdot)$. We normalize $\beta_{0,1} = 0$ and $\beta_{2,1} = 1$ to accommodate the level and scale of the fixed effects η_i. This implies there are $K_0 + K_1K_y + K_2 - 2$ target parameters in model (C.2.3). The baseline specification we use for nonlinear models in Sections 4.5.3 and 4.5.4 sets $K_0 = K_1 = 3$, $K_y = 1$ and $K_2 = 1$ (additive fixed effects) or $K_2 = 2$ (interacted fixed effects).

It is often useful to rewrite (C.2.3) in vector notation. If $h_{1:K_s}(r)$ is used for to indicate the $1 \times K_s$ array obtained by horizontal concatenation of the elements in $\{h_\tau(r)\}_{\tau=1}^{K_S}$, let us define

$$\begin{aligned} D_{jit} &= (h_{2:K_0}(r_{jit}), h_{1:K_1}(r_{jit})g_1(y_{it}), \ldots, h_{1:K_1}(r_{jit})g_{K_y}(y_{it}))' \\ \beta_{0,1} &= (\beta_{0,2:K_0}, \beta_{1,1:K_1,1}, \ldots, \beta_{1,1:K_1,K_y})', \end{aligned}$$

and similarly $H_{jit} = h_{2:K_2}(r_{jit})'$ and $\beta_2 = \beta'_{2,2:K_2}$. We then stack observations (t, j) vertically for each unit to

[5] Cubic natural splines are piece-wise cubic polynomials that are twice continuously differentiable and restricted to be linear beyond the boundary knots. Differentiability is crucial to compute densities and quantile-based measures of nonlinear persistence, as in (4.14). In particular, let τ_k for $k \in \{1, \ldots, K_s\}$ index the knots in increasing order, which we place at the $k/(K_s+1)$th quantiles of the empirical distribution of r_{jit}. The following K_s basis functions can be used to represent the spline model:

$$\left[1, r_{jit}, d_1(r_{jit}) - d_{K_s-1}(r_{jit}), \ldots, d_{K_s-2}(r_{jit}) - d_{K_s-1}(r_{jit})\right],$$

so that $K_s = 2$ corresponds to a linear spline, and for $K_s > 2$ we have

$$d_k(r) = \frac{(r-\tau_k)^3\, \mathbb{1}\{r \geq \tau_k\} - \left(r - \tau_{K_s}\right)^3 \mathbb{1}\left\{r \geq \tau_{K_s}\right\}}{\tau_{K_s} - \tau_k}$$

for $k \in \{1, \ldots, K_s - 2\}$.

obtain

$$\ell_i = D_i\beta_{0,1} + \left(1_{TJ} + H_i\beta_2\right)\eta_i + \varepsilon_i, \tag{C.2.3}$$

where 1_A is a vector of ones of size A and where $\ell_i = (\ell_{1i1}, \ell_{2i1}, \ell_{3i1}, \ell_{1i2}, \ell_{2i2}, \ell_{3i2})'$ and so on.

C.2.2 Estimation: additive fixed effects

Model (C.2.3) is then a series regression model on the sequence of parameter sets defined by $(K_{0,1,2}, K_y)$, with the additional twist that η_i is unobserved.

When η_i enters additively (set $K_2 = 1$), given the conditional mean assumption $E\left[\varepsilon_i \middle| r_i, y_i, \eta_i\right] = 0$, the model in (C.2.3) is a static fixed-effects regression that can be estimated using the within-group estimator. In other words, let $\tilde{\ell}_{jit} = \ell_{jit} - (TJ)^{-1}\sum_{(j,t)}\ell_{jit}$ denote variables in deviations with respect to means (recall that $T = 2$ and $J = 3$ here) and use $\tilde{\ell}_i$ for the corresponding vectors. Equation (C.2.3) then becomes $\tilde{\ell}_i = \tilde{D}_i\beta_{0,1} + \tilde{\varepsilon}_i$ and an estimate of $\beta_{0,1}$ can be obtained via least squares.

Penalization. When considering models where (K_0, K_1, K_y) grow large, regularized estimators might be attractive. We explore implementations that penalize the nonlinear sieve terms and might prove useful in richer setups where even more flexible specifications might be feasible. A simple implementation via a ridge penalty $\lambda > 0$ allows us to maintain the simplicity of the estimation method and an explicit solution that recovers the linear model as $\lambda \longrightarrow \infty$.

C.2.3 Estimation: interacted fixed effects

As noted in Section 4.3.5, treating $\{\eta_i\}_{i=1}^n$ as parameters to be estimated jointly with $\beta_{0,1}$ and β_2 results in an incidental parameters problem that precludes fixed-T consistent estimation. Here we generalize the linear instrumental variables (IV) strategy introduced in the main text, and discuss a method-of-moments approach in greater generality at the end.

Recall that we use $\tilde{\ell}_{jit} = \ell_{jit} - \bar{\ell}_i$ and so on as notation for variables in deviations with respect to means:

$$\begin{aligned}\tilde{\ell}_{jit} &= \tilde{D}_{jit}\beta_{0,1} + \tilde{H}_{jit}\eta_i\beta_2 + \tilde{\varepsilon}_{jit},\\ \bar{\ell}_i &= \bar{D}_i\beta_{0,1} + (1 + \bar{H}_i\beta_2)\eta_i + \bar{\varepsilon}_i.\end{aligned}$$

We consider the case $K_2 = 2$, so that effectively $H_{jit} = r_{jit}$. We look for linear transformations of the model that do not depend on η_i but still allow us to estimate β_2. Note that

$$\begin{aligned}H_{jit}\bar{\ell}_i - \bar{H}_i\ell_{jit} &= H_{jit}\left(\bar{D}_i\beta_{0,1} + (1 + \bar{H}_i\beta_2)\eta_i + \bar{\varepsilon}_i\right) - \bar{H}_i\left(D_{jit}\beta_{0,1} + (1 + H_{jit}\beta_2)\eta_i + \varepsilon_{jit}\right)\\ &= \left(H_{jit}\bar{D}_i - \bar{H}_iD_{jit}\right)\beta_{0,1} + \tilde{H}_{jit}\eta_i + H_{jit}\bar{\varepsilon}_i - \bar{H}_i\varepsilon_{jit},\end{aligned}$$

where note that the first element in $H_{jit}\bar{D}_i - \bar{H}_iD_{jit}$ is zero. We solve for the term involving η_i and plug it in back in the model in deviations,

$$\tilde{\ell}_{jit} = \tilde{D}_{jit}\beta_{0,1} + \left(H_{jit}\bar{\ell}_i - \bar{H}_i\ell_{jit}\right)\beta_2 - \left(H_{jit}\bar{D}_i + \bar{H}_iD_{jit}\right)\gamma + \xi_{jit} \tag{C.2.4}$$

where $\xi_{jit} = \tilde{\varepsilon}_{jit} - H_{jit}\beta_2\bar{\varepsilon}_i + \bar{H}_i\beta_2\varepsilon_{jit}$ and $\gamma = -\beta_{0,1}\beta_2$, a generalization of the simple model considered in equation (4.19) in Section (4.3.5). We need at least one instrument for $H_{jit}\bar{\ell}_i - \bar{H}_i\ell_{jit}$. Note that

$$E\left[H_{jit}\bar{\ell}_i - \bar{H}_i\ell_{jit}\middle|r_i, y_i\right] = \left(H_{jit}\bar{D}_i - \bar{H}_iD_{jit}\right)\beta_{0,1} + \tilde{H}_{jit}E\left[\eta_i\middle|r_i, y_i\right],$$

and thus any predictor of η_i (conditional on the included regressors) is possibly a valid instrument. We thus consider a set of $K_0 + K_1K_y - 1$ instruments given by $Z_{jit} = \tilde{H}_{jit}\bar{D}_i$; this corresponds to five instruments in the baseline specification used in Section 4.5.4. We then propose to estimate (C.2.4) by TSLS.

Note that the restriction $\gamma = -\beta_{0,1}\beta_2$ does not need to be imposed for consistent estimation. If one is willing to impose the restrictions, this can be done expost via minimum distance estimation or exante via nonlinear GMM. We briefly explored the latter, which is exposed to similar numerical/convergence problems as those of the more general nonlinear method-of-moments estimator described below.

A method-of-moments estimator. The IV estimator developed here retains the simplicity and linearity of the within-group estimator even as we move on to more flexible models, and we have found it to be a reliable approach. A general method-of-moments approach for the parameters in equation (C.2.3) is as follows.

Let $B_i(\beta_2)$ and $Q_i(\beta_2)$ denote the generalized between- and within-group transformations, respectively, defined as

$$B_i(\beta_2) = \left(\left(1_{TJ} + H_i\beta_2\right)'\left(1_{TJ} + H_i\beta_2\right)\right)^{-1}\left(1_{TJ} + H_i\beta_2\right)',$$
$$Q_i(\beta_2) = I_{TJ} - \left(1_{TJ} + H_i\beta_2\right)B_i(\beta_2),$$

where I_A is the identity matrix of size $A \times A$. Back to equation (C.2.3), we note that the generalized within-group residuals are mean independent of the regressors:

$$E\left[Q_i(\beta_2)\left(\ell_i - D_i\beta_{0,1}\right)\middle|r_i, y_i\right] = 0.$$

A nonlinear GMM estimator inspired in Chamberlain (1992) and Arellano and Bonhomme (2012) is then available exploiting these conditional moment restrictions. In some sense, the IV strategy proposed above is a transparent way to finding such informative restrictions for the interacted fixed-effects term.

C.2.4 Additional details on implementation

Here we discuss how to we compute the objects of interest after estimation of the model parameters in equation (4.9) (reproduced here as equation (C.2.1)) and shed light on some additional details beyond those discussed in Section 4.3.5.

Standardizing the data. When estimating flexible models, we first standardize the data as follows:

$$\check{r}_{jit} = \frac{r_{jit} - \bar{r}}{\bar{\sigma}_r}$$

$$\check{y}_{it} = \frac{y_{it} - \bar{y}}{\bar{\sigma}_y},$$

where $\bar{x}$ and $\bar{\sigma}_x$ are measures of location and scale for variable x; we use the median and the IQR respectively in our implementation. This helps standardize the range of variation of the data across units. Importantly, we need to undo these transformations before reporting the final output: letting $\check{q}_{it}(\tau)$ denote the τth conditional quantile on the standardized data, we need to calculate

$$q_{it}(\tau) = \bar{r} + \bar{\sigma}_r \check{q}_{it}(\tau).$$

Estimation in growth rates. In Section 4.3.5, we note that redefining $s_{jit} = r_{jit} - y_{it}$ is equivalent to estimating predictive distributions for growth rates and argue that this is a convenient transformation. Note that we are still interested in the range of values of r (or $\check{r}$): this entails careful adjustment of the support grid of conditional distribution functions in implementation. On a related note, the fact that the argument of the conditional distribution now depends on y has to be taken into account when computing numerical derivatives below.

Details on computing quantile-based measures of dispersion, skewness and persistence. Given estimates $(\hat{\beta}'_{0,1}, \hat{\beta}'_2)'$, the target summaries in Section 4.3.4 can be computed in three steps:

1. Obtain predicted probabilities.

   Given reference conditioning values $(\bar{y}, \bar{\eta})$ (usually a quantile of interest) and for r in a given grid r_{grid}, we calculate fitted probabilities $\hat{p} = \hat{F}(r, \bar{y}, \bar{\eta})$, which we collect in $\hat{p}_{\text{rgrid}}$. When non-monotonic, we follow Chernozhukov et al. (2010) in sorting the original estimated curve into a monotone rearranged one.

2. Recover conditional quantiles.

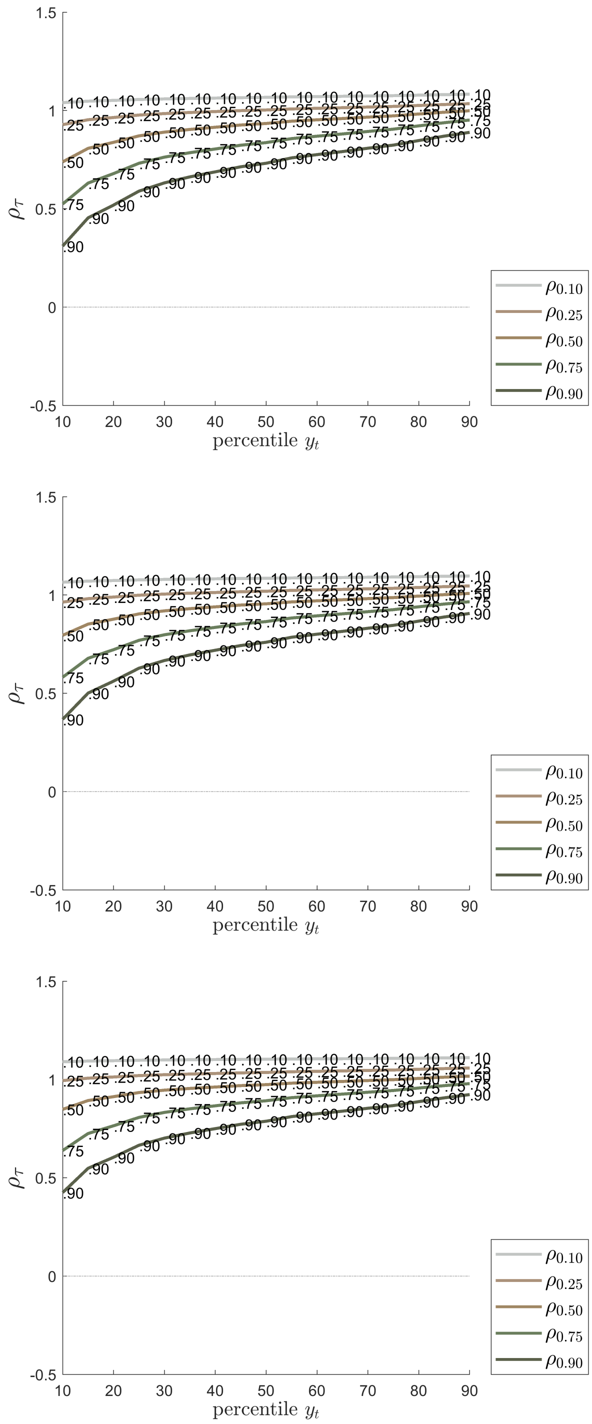
   This involves inverting the fitted conditional distribution function to obtain conditional quantiles for a given τ, which we denote $\hat{r}(\tau)$; see also equation (4.10) in the main text. We resort to interpolation within $\hat{p}_{\text{rgrid}}$.[6]

3. Compute target summaries.

   Quantile-based measures of dispersion and skewness as in equations (4.12) and (4.13) are then readily available (note the comment above on standardizing the data). Regarding nonlinear persistence, we calculate the derivatives in equation (4.15) mumerically. Note that this requires recalculating predicted probabilities along the lines of step 1, so as to condition on $\hat{r}(\tau)$.

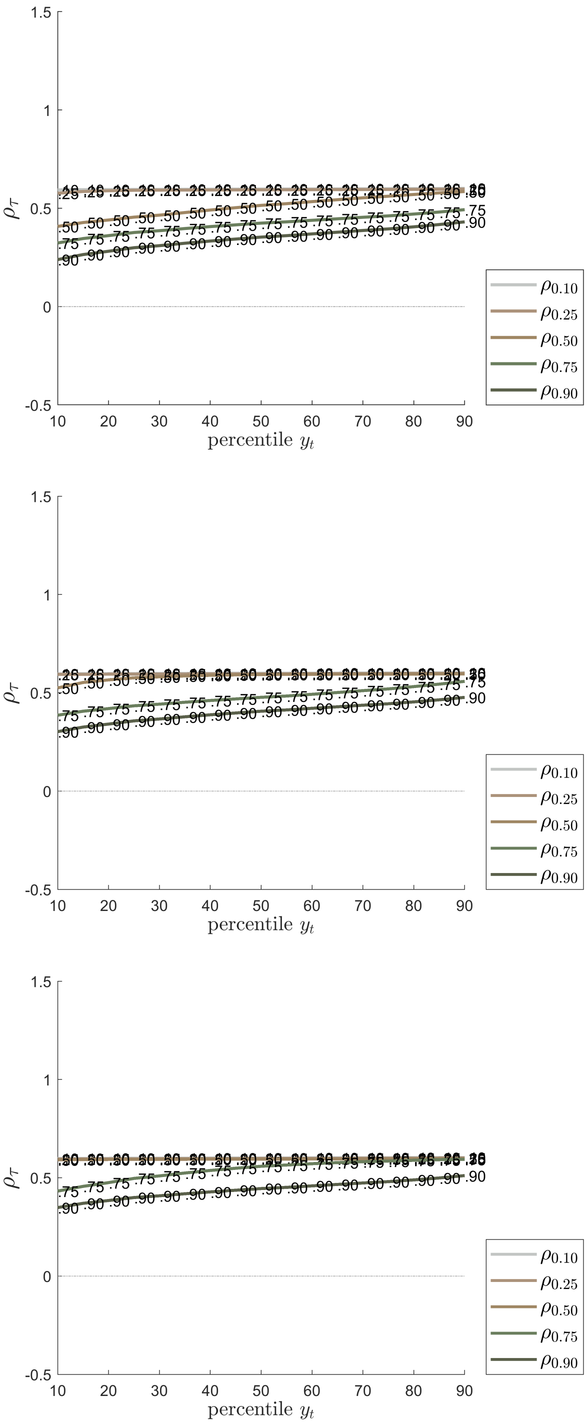
[6]Alternative methods are bracketing or root-finding alogorithms, which solve for $\hat{r}(\tau)$ in $\tau = \hat{F}(\hat{r}(\tau), \bar{y}, \bar{\eta})$. These are model-based approaches that impose the (estimated) logit structure, which is problematic when it is non-monotonic. In fact, these algorithms impose implicit rearrangement methods that are starting-value dependent.

C.3 Documenting heterogeneity



Note. Reference values for $\bar{\eta}$ correspond to its 10th, 50th and 90th percentiles, respectively. See Figure C.3.3 for pointwise confidence bands.

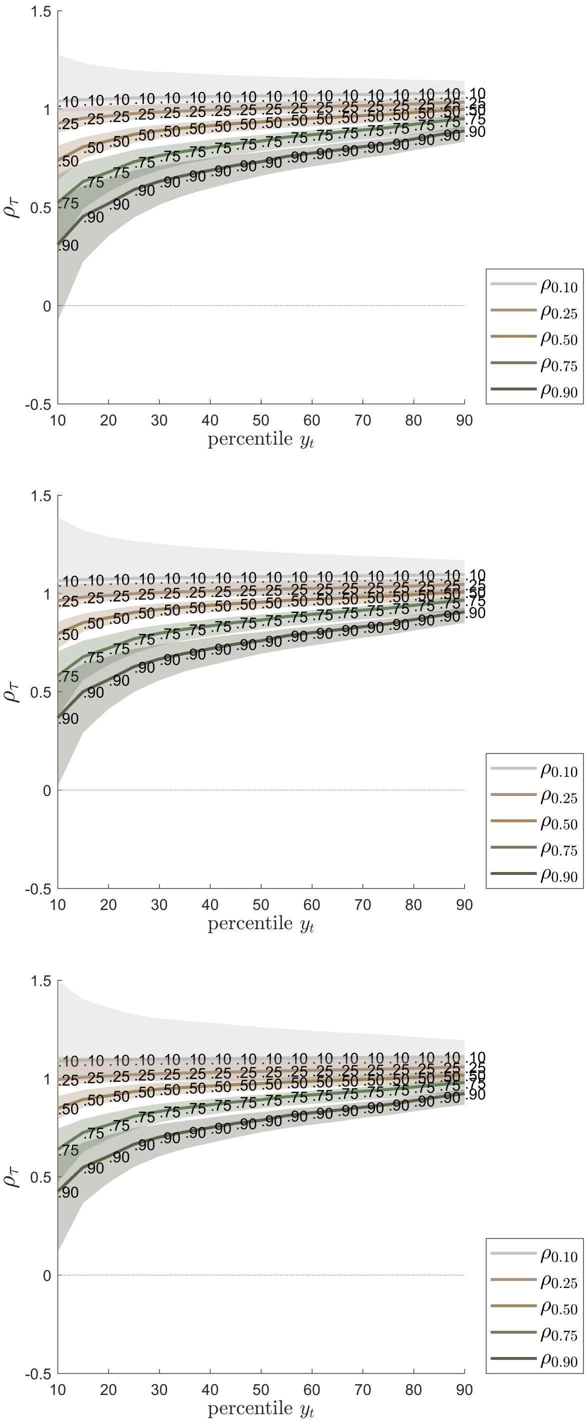
FIGURE C.3.1. India — nonlinear persistence at different reference values of η_i



Note. Reference values for $\bar{\eta}$ correspond to its 10th, 50th and 90th percentiles, respectively. See Figure C.3.4 for pointwise confidence bands.

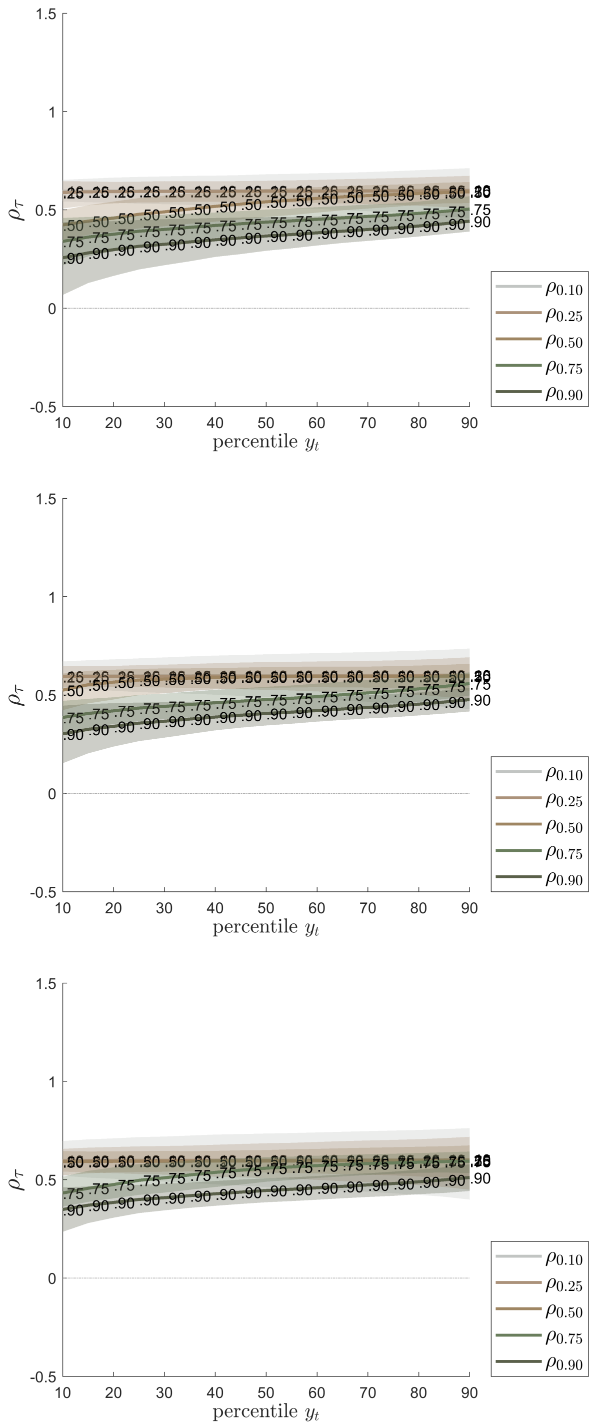
FIGURE C.3.2. Colombia — nonlinear persistence at different reference values of η_i

C.3.1 Confidence bands



Note. Reference values for $\bar{\eta}$ correspond to its 10th, 50th and 90th percentiles, respectively. 90% pointwise confidence bands; block bootstrap with 1000 repetitions.

FIGURE C.3.3. India — nonlinear persistence (with confidence bands)



Note. Reference values for $\bar{\eta}$ correspond to its 10th, 50th and 90th percentiles, respectively.

FIGURE C.3.4. Colombia — nonlinear persistence (with confidence bands)

C.4 Robustness: sample selection and modeling choices

In Sections C.4.1 and C.4.2, we report counterparts to (a subset of) the empirical results reported in the main text when including elicited probabilities that equal zero or one (applying the transformation below) and those which violate strict monotonicity (two reported cumulative probabilities are equal for the same household). This entails minimal sample selection in the Indian data (see Table C.1.1) but is substantial in the Colombian data (see Table C.1.3). We find very similar results in both qualitative and quantitative terms for our target summary objects in both datasets. This is remarkable for the Colombian data, where using these transformations essentially imply doubling the sample size (from 2,230 to 4,420 unique households) and introducing substantial additional elicitation error (as measured by residual variances).

Keeping zero/one probabilities The logit transformation in equation (4.2) in Section 4.3 restricts observed, elicited probabilities p_{jit} to lie strictly between zero and one. We suggest here an alternative transformation — which still maps these probabilities to the real line — that allows us to keep these observations:

$$\ell_{jit} = \operatorname{logit}\left(\check{p}_{jit}\right), \quad \check{p}_{jit} = \frac{p_{jit} + \frac{1}{2m}}{1 + \frac{J}{2m}}. \tag{C.4.1}$$

This is a generalization of the modified logit transformation of Cox and Snell (1970, p. 32) for binary data. In a context where p_{jit} are noisy measurements due to possible rounding and randomness in the elicitation process, the adjustment m can be interpreted as a measure of the accuracy of the elicitation such that $m = O\left(1/\sigma_\varepsilon^2\right)$. In particular, elicitations errors ε_{jit} in (4.3) can be seen as capturing sampling uncertainty from a hypothetical random sample of size m; see Arellano, Bonhomme, De Vera, Hospido, and Wei (2022, Online Appendix F) for additional details in the context of subjective expectations data and a Bayesian interpretation of (C.4.1).[7]

[7] Below, we set the regularization parameter m in equation (C.4.1) to $m = 10$ in both cases. Other reasonable choices lead to the same conclusions.

C.4.1 India: larger samples

	No FE	FE
ρ	0.96	0.93
	(0.94, 0.99)	(0.90, 0.96)
σ	0.58	0.33
	(0.53, 0.62)	(0.31, 0.35)
$IQR_{0.75}$	1.26	0.72
	(1.17, 1.36)	(0.67, 0.77)
$IQR_{0.90}$	2.53	1.44
	(2.34, 2.72)	(1.35, 1.53)
σ^2_η		0.18
		(0.15, 0.23)
σ^2_η village		0.12
		(0.11, 0.16)
σ^2_ε	1.05	0.98
	(1.03, 1.08)	(0.95, 1.02)

Note. The table reports results for the linear model in (4.7) using the data for India. Results correspond to the alternative sample described in Section C.4. Specifications include survey round dummies. In parenthesis we report 90% block bootstrap CI.

TABLE C.4.1. India — linear model (robustness sample)

	y_{p10}	y_{p50}	y_{p90}
$IQR_{0.75}$	0.83	0.63	0.54
	(0.75, 0.93)	(0.57, 0.66)	(0.48, 0.58)
$IQR_{0.90}$	1.69	1.29	1.10
	(1.56, 1.91)	(1.21, 1.37)	(0.99, 1.21)
$SK_{0.90}$	−0.04	−0.11	−0.15
	(−0.16, 0.04)	(−0.21, −0.05)	(−0.29, −0.05)
$\rho_{\tau 0.25}$	0.97	1.02	1.04
	(0.92, 1.04)	(0.98, 1.06)	(0.99, 1.08)
$\rho_{\tau 0.50}$	0.81	0.95	1.00
	(0.74, 0.87)	(0.92, 0.98)	(0.97, 1.02)
$\rho_{\tau 0.75}$	0.59	0.86	0.95
	(0.42, 0.72)	(0.82, 0.89)	(0.91, 0.98)
σ^2_η		0.19	
		(0.16, 0.23)	
σ^2_η village		0.11	
		(0.11, 0.16)	
σ^2_ε		0.95	
		(0.91, 0.99)	

Note. The table reports results for India for the flexible model with additive fixed effects in (4.20). Results correspond to the alternative sample described in Section C.4. Specifications include survey round dummies. In parenthesis we report 90% block bootstrap CI.

TABLE C.4.2. India — flexible model (additive fixed effects)

C.4.2 Colombia: larger samples

	No FE	FE
ρ	0.72	0.51
	(0.69, 0.75)	(0.48, 0.54)
σ	0.82	0.56
	(0.80, 0.85)	(0.55, 0.58)
$IQR_{0.75}$	1.80	1.24
	(1.75, 1.86)	(1.21, 1.27)
$IQR_{0.90}$	3.61	2.48
	(3.49, 3.72)	(2.41, 2.55)
σ_η^2		0.54
		(0.51, 0.58)
σ_η^2 village		0.10
		(0.10, 0.13)
σ_ε^2	1.75	1.34
	(1.71, 1.78)	(1.30, 1.37)

Note. The table reports results for the linear model in (4.7) using the data for Colombia. Results correspond to the alternative sample described in Section C.4. Specifications include survey round and month and interview dummies. In parenthesis we report 90% block bootstrap CI.

TABLE C.4.3. Colombia — linear model (robustness sample)

	y_{p10}	y_{p50}	y_{p90}
$IQR_{0.75}$	1.36	1.18	1.10
	(1.31, 1.42)	(1.14, 1.21)	(1.06, 1.14)
$IQR_{0.90}$	2.67	2.37	2.22
	(2.57, 2.78)	(2.31, 2.44)	(2.15, 2.29)
$SK_{0.90}$	0.08	0.05	0.01
	(0.04, 0.13)	(0.01, 0.08)	(−0.02, 0.04)
$\rho_{\tau 0.25}$	0.60	0.63	0.65
	(0.55, 0.64)	(0.59, 0.67)	(0.58, 0.70)
$\rho_{\tau 0.50}$	0.45	0.60	0.63
	(0.41, 0.51)	(0.56, 0.63)	(0.59, 0.66)
$\rho_{\tau 0.75}$	0.34	0.50	0.59
	(0.28, 0.40)	(0.46, 0.53)	(0.55, 0.62)
σ_η^2		0.53	
		(0.50, 0.57)	
σ_η^2 village		0.10	
		(0.10, 0.13)	
σ_ε^2		1.33	
		(1.29, 1.36)	

Note. The table reports results for Colombia for the flexible model with additive fixed effects in (4.20). Results correspond to the alternative sample described in Section C.4. Specifications include survey round and month and interview dummies. In parenthesis we report 90% block bootstrap CI.

TABLE C.4.4. Colombia — flexible model (additive fixed effects)

C.5 Questionnaires

Figure C.5.1 reports the original questionnaire for India. The questions on elicitation of subjective expectations follow those on income and income components and correspond to section 6 of the household survey.[8] Figure C.5.2 shows the original questionnaire for Colombia (in Spanish).

INTERVIEWER: Add all income sources in the shaded column to calculate yearly income of the household.

5.	READ OUT CALCULATED YEARLY INCOME and ask: Is this a typical yearly income for your household?	1. yes 2. no, it is higher than typical 3. no, it is lower	
6.	IF NO: What would be a typical yearly income for your household?	(Rs.)	

IF ONLY INCOME SOURCE IS FROM DAIRY ACTIVITY (7) >> GO TO SECTION 7. ELSE, go on to question 7.

7.	Imagine that you have a very good year, every member of working age in the household managed to have work, and there were no droughts or anything the like. What would be the maximum amount of income your household would receive in such a situation in one year?	Y	(Rs.)	
8.	Now imagine the total opposite: the harvest is bad, animals get sick, finding work is not possible. What would be the yearly income of your household in such a situation?	X	(Rs.)	

INTERVIEWER: Calculate the following values:

Expected Income (threshold B):	B = (X+Y)/2	
Threshold A:	A = (B+X)/2	
Threshold C:	C = (B+Y)/2	

INTERVIEWER: Explain the rainfall question to the respondent (See extra Sheet)

R.1	So, what do you think how likely it is that it will rain *tomorrow*?	
R.2	So, what do you think how likely it is that it will rain within the *coming week*?	
R.3	So, what do you think how likely it is that it will rain within the *coming month*?	

9.	How likely do you think it is that your yearly income in the coming year will be higher than _________(A) Rupees?	
10.	How likely do you think it is that your yearly income in the coming year will be higher than _________(B) Rupees?	
12.	How likely do you think it is that your yearly income in the coming year will be higher than _________(C) Rupees?	

FIGURE C.5.1. India — questionnaire

[8] To be precise, this is the second-round version of the questionnaire. In the first-round version, five households were instead asked about monthly — rather than yearly — income. Importantly, the wording of the questions is unchanged, and the data included an identifier for these households, which are not part of the original sample in Table C.1.1.

FORMULARIO TIPO 1 No. MÓDULO 6 4

631	¿El mes pasado recibió algún ingreso por concepto de trabajo, diferente al de su ocupación u oficio principal?	Si 1 ○ → ¿Cuánto recibió? $	No 2 ○
632	ENTREVISTADORA: *Verifique la edad de __________ en **604** y marque de acuerdo con la respuesta registrada.*	10 a 24 años 1 ○ → 635 25 y más años 2 ○	
633	¿El mes pasado recibió dinero por concepto de pensión de jubilación, sustitución pensional, invalidez o vejez?	Si 1 ○ → ¿Cuánto recibió? $	No 2 ○
634	¿El mes pasado recibió dinero por concepto de arriendos o intereses?	Si 1 ○ → ¿Cuánto recibió? $	No 2 ○
635	¿El mes pasado recibió dinero por otras fuentes diferentes al trabajo? (por ejemplo, venta o empeño de un bien)	Si 1 ○ → ¿Cuánto recibió? $	No 2 ○
636	ENTREVISTADORA: *Verifique en el "Reporte de Seguimiento". La persona objeto de este módulo es:*	Jefe del núcleo familiar seleccionado 1 ○ Cónyuge del jefe del núcleo familiar seleccionado 2 ○ Otro 3 ○ → E	
637	ENTREVISTADORA: *¿La persona objeto de este módulo debe aplicar "expectativas de ingreso"? Tenga en cuenta que esta sección, **aplica sólo a una persona del núcleo familiar seleccionado**.*	Si 1 ○ No 2 ○ → E	

D. EXPECTATIVAS DE INGRESO

ENTREVISTADORA: *Lea a su entrevistad@ el siguiente texto:*

"Ahora vamos a realizar un pequeño juego que consiste en lo siguiente: Aquí tenemos una regla que tiene una escala de 0 a 100. Queremos que la utilice para indicarnos qué tan seguro está Usted, de que alguna situación se va a presentar en el futuro, por ejemplo, si le preguntamos: ¿Qué tan seguro está de que mañana va a llover?.

1. Si Usted esta totalmente seguro que va a llover nos indica el punto 100 de la regla.
2. Si Usted está totalmente seguro de que no va a llover nos indica el punto 0 de la regla.
3. Y si Usted no está seguro de lo que va a ocurrir, pero cree que hay una alta probabilidad de que llueva se colocaría más cerca del 100 que del 0.
4. Y si cree que hay una alta probabilidad que no va a llover se colocaría más cerca del 0 que del 100.

"Ahora muéstreme en la regla qué tan seguro está de que mañana va a llover" ***(Que él indique con un lápiz)***.

638	Ahora suponga que el próximo mes los miembros de su familia que quieren trabajar, consiguen un trabajo bueno. (*Si tiene parcela, decir también: Imagine además que Usted obtiene una buena cosecha*). ¿Cuánto dinero cree que ganaría o le entraría en ese mes al hogar?	X	$	NS/NR ○ → E
639	Suponga ahora todo lo contrario, que tienen muy poco trabajo el próximo mes (*Si tiene parcela, decir también: Suponga que la cosecha salió mal*), y que sólo viven de eso y de lo que la gente les da, y que la gente les da muy poco. ¿Cuánto dinero cree que recibiría en ese mes el hogar?	Y	$	NS/NR ○ → E
640	ENTREVISTADORA: *Promedie las dos posibilidades (X y Y), y calcule el ingreso esperado del hogar. Mencione la cifra al entrevistado, diciendo **"entonces el ingreso promedio sería" (Z).***	Z	(X+Y)/2 $	
641	ENTREVISTADORA: *Calcule el valor de ingreso **M**, a partir del ingreso promedio.*	M	(Z+X)/2 $	
642	ENTREVISTADORA: *Calcule el valor de ingreso **P**, a partir del ingreso promedio.*	P	(Z+Y)/2 $	

		A	B	C
643	Ahora vamos a jugar con la regla. Usted debe responder señalándome un punto en la regla, y la pregunta es la siguiente: ¿Qué tan seguro está Usted que el ingreso del hogar va a estar entre $_________ y $_________? ENTREVISTADORA: *Si no entiende, repítale el ejemplo de la lluvia.*	Entre X y M %	Entre X y Z %	Entre X y P %

ENTREVISTADORA: *Compruebe que la respuesta de C sea mayor que la de B y la de B mayor que la de A. Si no es así vuelva y repítale el ejemplo de la lluvia.*

1256556123 SEI; ARD-234 /DD-06/JUL-03

FIGURE C.5.2. Colombia — questionnaire

BIBLIOGRAPHY

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020): "Sampling-based vs design-based uncertainty in regression analysis." *Econometrica*, 88, 265–296.

——— (2023): "When Should You Adjust Standard Errors for Clustering?" *Quarterly Journal of Economics*, 138, 1–35.

Abadie, A., G. W. Imbens, and F. Zheng (2014): "Inference for misspecified models with fixed regressors." *Journal of the American Statistical Association*, 109, 1601–1614.

Abowd, J. M. and D. Card (1989): "On the covariance structure of earnings and hours changes." *Econometrica*, 57, 411–445.

Adamopoulos, T., L. Brandt, J. Leight, and D. Restuccia (2022): "Misallocation, Selection, and Productivity: A Quantitative Analysis With Panel Data From China." *Econometrica*, 90, 1261–1282.

Ahn, D., S. Choi, D. Gale, and S. Kariv (2014): "Estimating Ambiguity Aversion in a Portfolio Choice Experiment." *Quantitative Economics*, 5, 195–223.

Albrizio, S., B. González, and D. Khametshin (2023): "A Tale of Two Margins: Monetary Policy and Capital Misallocation." Working paper.

Almuzara, M. (2020): "Heterogeneity in Transitory Income Risk." Working paper.

Altonji, J.G., H. D. and I. Vidangos (2023): "Individual earnings and family income: Dynamics and distribution." *Review of Economic Dynamics*, 49, 225–250.

Alvarez, J. and M. Arellano (2022): "Robust likelihood estimation of dynamic panel data models." *Journal of Econometrics*, 226, 21–61.

Amberg, N., T. Jansson, M. Klein, and A. Rogantini Picco (2022): "Five Facts about the Distributional Income Effects of Monetary Policy Shocks." *AER: Insights*, 4, 289–304.

Andersen, A. L., N. Johannesen, M. Jørgensen, and J.-L. Peydró (2023): "Monetary Policy and Inequality." *Journal of Finance*, 78, 2945–2989.

Anderson, G. and A. Cesa-Bianchi (2024): "Crossing the Credit Channel: Credit Spreads and Firm Heterogeneity." *American Economic Journal: Macroeconomics*, 16, 417–446.

Andreoni, J. and L. Samuelson (2006): "Building Rational Cooperation." *Journal of Economic Theory*, 127, 117–154.

Andrews, D., X. Cheng, and P. Guggenberger (2020): "Generic results for establishing the asymptotic size of confidence sets and tests," *Journal of Econometrics*, 218, 496–531.

Andrews, D. 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.

Arellano, M. (1987): "Computing Robust Standard Errors for Within-Group Estimators." *Oxford Bulletin of Economics and Statistics*, 49, 431–434.

——— (2003): *Panel Data Econometrics*, Oxford University Press.

——— (2014): "Uncertainty, persistence and heterogeneity: A panel data perspective." *Journal of the European Economic Association*, 12, 1127–1153.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and consumption dynamics: A nonlinear panel data framework." *Econometrica*, 95, 693–734.

Arellano, M. and S. Bonhomme (2012): "Identifying distributional characteristics in random coefficients panel data models." *Review of Economic Studies*, 79, 987–1020.

——— (2016): "Nonlinear panel data estimation via quantile regressions." *Econometrics Journal*, 19, C61–C94.

Arellano, M., S. Bonhomme, M. De Vera, L. Hospido, and S. Wei (2022): "Income Risk Inequality: Evidence from Spanish Administrative Records." *Quantitative Economics*, 13, 1747–1801.

Arkhangelsky, D. and V. Korovkin (2023): "On Policy Evaluation with Aggregate Time-Series Shocks." Working Paper arXiv:1905.13660.

Aronow, P. M., D. P. Green, and D. K. K. Lee (2014): "Sharp bounds on the variance in randomized experiments," *The Annals of Statistics*, 42, 850–871.

Attanasio, O. (2009): "Expectations and Perceptions in Developing Countries: Their Measurement and Their Use." *American Economic Review: Papers and Proceedings*, 99, 87–92.

Attanasio, O. and B. Augsburg (2016): "Subjective Expectations and Income Processes in Rural India," *Economica*, 83, 416–442.

Attanasio, O., E. Battistin, and A. Mesnard (2012): "Food and Cash Transfers: Evidence from Colombia." *Economic Journal*, 122, 92–124.

Augsburg, B. (2009): *Microfinance - Greater Good or Lesser Evil?*, PhD thesis, Maastricht Graduate School of Governance, University of Maastricht.

Bahaj, S., A. Foulis, G. Pinter, and P. Surico (2022): "Employment and the Residential Collateral Channel of Monetary Policy." *Journal of Monetary Economics*, 131, 26–44.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry (2015): "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science*, 348.

Bardóczy, B., G. Bornstein, C. Maggi, and S. Salgado (2024): "Monopsony Power and the Transmission of Monetary Policy." Working paper.

Barras, L., P. Gagliardini, and O. Scaillet (2022): "Skill, Scale, and Value Creation in the Mutual Fund Industry." *Journal of Finance*, 77, 601–638.

Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997): "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics*, 112, 537–579.

Bellifemine, M., A. Couturier, and R. Jamilov (2023): "The Regional Keynesian Cross." Working paper.

Berthold, B., A. Cesa-Bianchi, F. Di Pace, and A. Haberis (2024): "The Heterogeneous Effects of Carbon Pricing: Macro and Micro Evidence." Working paper.

Bilal, A. and D. R. Känzig (2024): "The Macroeconomic Impact of Climate Change: Global vs. Local Temperature." NBER Working Paper 32450.

Bils, M., P. J. Klenow, and C. Ruane (2021): "Misallocation or Mismeasurement?" *Journal of Monetary Economics*, 124S.

Bojinov, I., A. Rambachan, and N. Shephard (2021): "Panel experiments and dynamic causal effects: A finite population perspective," *Quantitative Economics*, 12, 1171–1196.

Bonhomme, S. and A. Denis (2024): "Fixed Effects and Beyond: Bias Reduction, Groups, Shrinkage, and Factors in Panel Data." Working paper.

Bonhomme, S. and J.-M. Robin (2009): "Consistent noisy independent component analysis." *Journal of Econometrics*, 149, 12–25.

Bonhomme, S. and J. M. Robin (2010): "Generalized non-parametric deconvolution with an application to earnings dynamics." *Review of Economic Studies*, 77, 491–533.

Bonhomme, S. and U. Sauder (2011): "Recovering Distributions in Difference-in-Differences Models: A Comparison of Selective and Comprehensive Schooling." *Review of Economics and Statistics*, 93, 479–494.

Browning, M. and J. Carro (2007): "Heterogeneity and microeconometrics modeling." in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson, Cambridge University Press, vol. 2, 47–74.

Caglio, C. R., R. M. Darst, and c. Kalemli-Özcan (2024): "Collateral Heterogeneity and Monetary Policy Transmission: Evidence from Loans to SMEs and Large Firms." Working paper.

Camara, S. and S. Ramirez Venegas (2023): "The Transmission of US Monetary Policy Shocks: The Role of Investment & Financial Heterogeneity." Working paper.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference With Multiway Clustering." *Journal of Business and Economic Statistics*, 29, 238–249.

Camêlo, F. (2024): "Capital Allocation and Firm Dynamics in Small Open Economies." Working paper.

Cao, J., T. Hegna, M. B. Holm, R. Juelsrud, T. König, and M. Riiser (2024): "The Investment Channel of Monetary Policy: Evidence from Norway." Working paper.

Cascaldi-Garcia, D., M. Vukotić, and S. Zubairy (2023): "Innovation During Challenging Times." Working paper.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018a): "Alternative asymptotics and the partially linear model with many regressors," *Econometric Theory*, 34, 277–301.

——— (2018b): "Inference in Linear Regression Models with Many Covariates and Heteroscedasticity." *Journal of the American Statistical Association*, 113, 1350–1361.

Chamberlain, G. (1992): "Efficiency bounds for semiparametric regression." *Econometrica*, 60, 567–596.

Chen, C., D. Restuccia, and R. Santaeulàlia-Llopis (2022): "The effects of land markets on resource allocation and agricultural productivity." *Review of Economic Dynamics*, 45, 41–54.

——— (2023): "Land Misallocation and Productivity." *American Economic Journal: Macroeconomics*, 15, 441–465.

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models." in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, Elsevier, vol. 6B, chap. 76.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and probability curves without crossing." *Econometrica*, 78, 1093–1125.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions." *Econometrica*, 81, 2205–2268.

Chiappori, P.-A., K. Samphantharak, S. Schulhofer-Wohl, and R. M. Townsend (2014): "Heterogeneity and Risk Sharing in Village Economies." *Quantitative Economics*, 5, 1–27.

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.

Cochran, W. G. (1977): *Sampling Techniques*, John Wiley & Sons, 3rd ed.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." *Econometrica*, 80, 2543–2594.

Cox, D. R. and E. J. Snell (1970): *Analysis of Binary Data*, Cambridge University Press, second ed.

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

David, J. M. and V. Venkateswaran (2019): "The Sources of Capital Misallocation." *American Economic Review*, 109, 2531–2567.

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press, 2nd edition ed.

De Nardi, M., F. G. and G. Paz-Pardo (2020): "Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare." *Journal of the European Economic Association*, 18, 890–926.

Deeb, A. and C. de Chaisemartin (2022): "Clustering and External Validity in Randomized Controlled Trials." Working Paper arxiv:1912.01052.

Delavande, A. (2023): "Expectations in development economics." in *Handbook of Economic Expectations*, ed. by T. G. Bachmann, R. and W. van der Klaauw, 261–291.

Delavande, A., X. Giné, and D. McKenzie (2011): "Measuring subjective expectations in developing countries: A critical review and new evidence." *Journal of Development Economics*, 94, 151–163.

Delavande, A. and S. Rohwedder (2008): "Eliciting Subjective Probabilities in Internet Surveys." *Public Opinion Quarterly*, 72, 866–891.

Dominitz, J. (1998): "Earnings Expectations, Revisions, and Realizations." *Review of Economics and Statistics*, 80, 374–388.

——— (2001): "Estimation of income expectations models using expectations and realization data." *Journal of Econometrics*, 102, 165–195.

Dominitz, J. and C. F. Manski (1996): "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources*, 31, 1–26.

——— (1997a): "Perceptions of economic insecurity: evidence from the survey of economic expectations," *Public Opinion Quarterly*, 61, 261–287.

——— (1997b): "Using expectations data to study subjective income expectations." *Journal of the American Statistical Association*, 92, 855–867.

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.

Durante, E., A. Ferrando, and P. Vermeulen (2022): "Monetary Policy, Investment and Firm Heterogeneity." *European Economic Review*, 148.

Duval, R., D. Furceri, R. Lee, and M. M. Tavares (2023): "Market Power, Financial Constraints, and Monetary Transmission." *AEA Papers and Proceedings*, 113, 99–104.

Erdős, P. and A. Rényi (1959): "On the Central Limit Theorem for Samples from a Finite Population," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 4, 49–61.

Evdokimov, K. (2010): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity." Working paper.

Fagereng, A., M. A. H. Gulbrandsen, M. B. Holm, and G. J. Natvik (2023): "Fisher Effects, Monetary Policy, and Household Indebtedness." Working paper.

Ferreira, T. R., D. A. Ostry, and J. Rogers (2023): "Firm Financial Conditions and the Transmission of Monetary Policy." Working paper.

Fogarty, C. B. (2018): "Regression-assisted inference for the average treatment effect in paired experiments." *Biometrika*, 105, 994–1000.

Foresi, S. and F. Peracchi (1995): "The conditional distribution of excess returns: An empirical analysis." *Journal of the American Statistical Association*, 90, 451–466.

Freedman, D. (2008): "On regression adjustments in experiments with several treatments." *The Annals of Applied Statistics*, 2, 176–196.

Fukui, M., E. Nakamura, and J. Steinsson (2023): "The Macroeconomic Consequences of Exchange Rate Depreciations." NBER Working Paper w31279.

Fuller, W. A. (1995): "Estimation in the presence of measurement error." *International Statistical Review*, 63, 121–141.

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

García-Santana, M., E. Moral-Benito, J. Pijoan-Mas, and R. Ramos (2020): "Growing Like Spain: 1995–2007." *International Economic Review*, 61, 1–468.

Genicot, G. and D. Ray (2017): "Aspirations and Inequality." *Econometrica*, 85, 489–519.

Ghomi, M. (2023): "Monetary Policy, Economic Uncertainty, and Firms' R&D Expenditure." Working paper.

Giacomini, R., S. Lee, and S. Sarpietro (2023): "A Robust Method for Microforecasting and Estimation of Random Effects." Working Paper arXiv:2308.01596.

Gilraine, M., J. Gu, and R. McMillan (2022): "A Nonparametric Approach for Studying Teacher Impacts." Working paper.

Golosov, M. and A. Tsyvinski (2015): "Policy Implications of Dynamic Public Finance." *Annual Review of Economics*, 7, 147–171.

González, B., G. Nuño, D. Thaler, and S. Albrizio (2023): "Firm Heterogeneity, Capital Misallocation and Optimal Monetary Policy." Working paper.

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

Goncalves, S., A. M. Herrera, L. Kilian, and E. Pesavento (forthcoming): "State-dependent local projections." *Journal of Econometrics*.

Gorea, D., O. Kryvtsov, and M. Kudlyak (2023): "House Price Responses to Monetary Policy Surprises: Evidence from the U.S. Listings Data." Working paper, European Investment Bank, Bank of Canada, Federal Reserve Bank of San Francisco.

Gorodnichenko, Y., D. Revoltella, J. Svejnar, and C. T. Weiss (2021): "Resource Misallocation in European Firms: The Role of Constraints, Firm Characteristics and Managerial Decisions." Working paper.

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

Graham, B. S. and J. L. Powell (2012): "Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models." *Econometrica*, 80, 2105–2152.

Gulyas, A., M. Meier, and M. Ryzhenkov (2024): "Labor Market Effects of Monetary Policy Across Workers and Firms." *European Economic Review*, 166.

Guner, N., G. Ventura, and X. Yi (2008): "Macroeconomic Implications of Size-Dependent Policies." *Review of Economic Dynamics*, 11, 721–744.

Guvenen, F. (2009): "An empirical investigation of labor income processes." *Review of Economic Dynamics*, 12, 58–79.

Guvenen, F., F. Karahan, S. Özkan, and J. Song (2021): "What Do Data on Millions of U.S. Workers Reveal About Life-Cycle Earnings Dynamics?" *Econometrica*, 89, 2303–2339.

Gürkaynak, R., H. G. Karasoy-Can, and S. S. Lee (2022): "Stock Market's Assessment of Monetary Policy Transmission: The Cash Flow Effect." *Journal of Finance*, 77, 2027–2073.

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., J. D. Singleton, and N. Yildiz (2023): "Identification of Non-Additive Fixed Effects Models: Is the Return to Teacher Quality Homogeneous?" NBER Working Paper 31384, National Bureau of Economic Research.

Hall, R. E. and F. S. Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." *Econometrica*, 50, 461–481.

Hansen, C. B. (2007): "Asymptotic properties of a robust variance matrix estimator for panel data when T is large." *Journal of Econometrics*, 141, 597–620.

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.

Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Hansen, M. H. and W. N. Hurwitz (1943): "On the determination of optimum probabilities in sampling." *Annals of Mathematical Statistics*, 14, 333–362.

Hansen, M. H., W. N. Hurwitz, and M. A. Bershad (1961): "Measurement errors in censuses and surveys." *Bulletin of the International Statistical Institute*, 38, 359–374.

Heckman, J. (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy*, 109, 673–748.

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

Higgins, A. and K. Jochmans (2024): "Bootstrap inference for fixed-effect models." *Econometrica*, 92, 411–427.

Holm, M. B., P. Paul, and A. Tischbirek (2021): "The Transmission of Monetary Policy under the Microscope." *Journal of Political Economy*, 129, 2861–2904.

Horvitz, D. G. and D. J. Thompson (1952): "A Generalization of Sampling Without Replacement From a Finite Universe." *Journal of the American Statistical Association*, 47, 663–685.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124, 1403–1448.

Hu, Y. (2017): "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics." *Journal of Econometrics*, 200, 154–168.

Hájek, J. (1960): "Limiting Distributions in Simple Random Sampling From a Finite Population," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 361–374.

Ignatiadis, N. and S. Wager (2022): "Confidence Intervals for Nonparametric Empirical Bayes Analysis." *Journal of the American Statistical Association*, 117, 1149–1166.

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

Indarte, S. (2023): "Financial Crises and the Transmission of Monetary Policy to Consumer Credit Markets." *Review of Financial Studies*, 36, 4045–4087.

Ippolito, F., A. K. Ozdagli, and A. Perez-Orive (2018): "The Transmission of Monetary Policy through Bank Lending: The Floating Rate Channel." *Journal of Monetary Economics*, 95, 49–71.

Jeenas, P. (2019): "Monetary Policy Shocks, Financial Structure, and Firm Activity: A Panel Approach." Working paper.

——— (2024): "Firm Balance Sheet Liquidity, Monetary Policy Shocks, and Investment Dynamics." Working paper.

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

Jin, Y. and D. Rothenhäusler (2024): "Tailored inference for finite populations: conditional validity and transfer across distributions." *Biometrika*, 111, 215–230.

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.

Jovanovic, B. (1987): "Micro Shocks and Aggregate Risk." *Quarterly Journal of Economics*, 102, 395–409.

Kaplan, G. and G. L. Violante (2010): "How much consumption insurance beyond self-insurance?" *American Economic Journal: Macroeconomics*, 2, 53–87.

Klette, T. J. (1999): "Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data." *Journal of Industrial Economics*, 47, 451–476.

Kline, P., E. K. Rose, and C. R. Walters (2022): "Systemic Discrimination Among Large U.S. Employers." *Quarterly Journal of Economics*, 137, 1963–2036.

Kline, P., R. K. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components." *Econometrica*, 88, 1859–1898.

Kotlarski, I. (1967): "On characterizing the gamma and the normal distribution." *Pacific Journal of Mathematics*, 20, 69–76.

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

Lakdawala, A. and T. Moreland (2023): "The Effect of Monetary Policy on Firm-Level Uncertainty." *Economics Letters*, 232.

——— (forthcoming): "Firm-Level Uncertainty and the Transmission of Monetary Policy." *Review of Economics and Statistics*.

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.

Lehmann, E. L. (1975): *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day.

Li, X. and P. Ding (2017): "General forms of finite population central limit theorems with applications to causal inference." *Journal of the American Statistical Association*, 112, 1759–1769.

Liang, K. and S. Zeger (1986): "Longitudinal data analysis using generalized linear models." *Biometrika*, 73, 13–22.

Lo Duca, M., D. Moccero, and F. Parlapiano (2024): "The Impact of Macroeconomic and Monetary Policy Shocks on Credit Risk in the Euro Area Corporate Sector." Working paper.

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

Magnac, T., N. Pistolesi, and S. Roux (2018): "Post-Schooling Human Capital Investments and the Life Cycle of Earnings." *Journal of Political Economy*, 126, 1219–1249.

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

Manski, C. (2018): "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise." in *NBER Macroeconomics Annual*, ed. by M. Eichenbaum and J. Parker, University of Chicago Press.

Manski, C. F. (2004): "Measuring Expectations." *Econometrica*, 72, 1329–1376.

Meager, R. (2022): "Aggregating distributional treatment effects: A Bayesian hierarchical analysis of the microcredit literature." *American Economic Review*, 112, 1818–1847.

Meghir, C. and L. Pistaferri (2011): "Earnings, consumption and life cycle choices." in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4B, chap. 9.

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., B. O'Flaherty, and R. Sethi (2021): "Empirical Bayes Counterfactuals in Poisson Regression with an Application to Police Use of Deadly Force." Working paper.

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 (2024): "Double Robustness of Local Projections and Some Unpleasant VARithmetic." Working Paper NBER Working Paper 32495.

Morgan, M. G. and M. Henrion (1990): *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press.

Mukhopadhyay, P. (2001): *Topics in Survey Sampling*, vol. 153 of *Lecture Notes in Statistics*, Springer.

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

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

Newey, W. K. and D. L. McFadden (1994): "Large sample estimation and hypothesis testing." in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Elsevier, vol. 4, chap. 36.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55, 703–708.

Neyman, J. (1923/1990): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Roczniki Nauk Rolniczych*, 10, 1–51, reprinted in English in Statistical Science, 1990, 5, 463-480.

——— (1934): "On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection." *Journal of the Royal Statistical Society, Series A*, 97, 558–606.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49, 1417–1426.

Nigmatulina, D. (2023): "Sanctions and Misallocation: How Sanctioned Firms Won and Russians Lost," Working paper.

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

Ozdagli, A. K. (2018): "Financial Frictions and the Stock Price Reaction to Monetary Policy." *Review of Financial Studies*, 31, 3895–3936.

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

Palazzo, B. and R. Yamarthy (2022): "Credit Risk and the Transmission of Interest Rate Shocks." *Journal of Monetary Economics*, 130, 120–136.

Paranhos, L. (2024): "How Do Firms' Financial Conditions Influence the Transmission of Monetary Policy? A Non-parametric Perspective." Working paper.

Paz, P. (2022): "Bank Capitalization Heterogeneity and Monetary Policy." Working paper.

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

Peters, M. (2020): "Heterogeneous Markups, Growth, and Endogenous Misallocation." *Econometrica*, 88, 2037–2073.

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.

Restuccia, D. and R. Rogerson (2017): "The Causes and Costs of Misallocation." *Journal of Economic Perspectives*, 31, 151–174.

Rosenbaum, P. R. (2002): "Covariance adjustment in randomized experiments and observational studies." *Statistical Science*, 17, 286–327.

Ruzzier, G. (2024): "Specialized Banks and the Transmission of Monetary Policy: Evidence from the U.S. Syndicated Loan Market." Working paper.

Schennach, S. M. (2022): "Measurement Systems." *Journal of Economic Literature*, 60, 1223–1263.

Schulhofer-Wohl, S. (2011): "Heterogeneity and Tests of Risk Sharing." *Journal of Political Economy*, 119, 925–958.

Singh, A., J. Suda, and A. Zervou (2022): "Monetary Policy, Labor Market, and Sectoral Heterogeneity." *AEA Papers and Proceedings*, 112, 491–495.

——— (2023): "Heterogeneity in labor market response to monetary policy: small versus large firms." Working paper.

Staiger, D. and J. H. Stock (1997): "Instrumental variables regression with weak instruments." *Econometrica*, 65, 557–586.

Startz, R. and D. G. Steigerwald (2023): "Inference and extrapolation in finite populations with special attention to clustering." *Econometric Reviews*, 42, 343–357.

——— (2024): "The variance of regression coefficients when the population is finite." *Journal of Econometrics*, 240.

Stock, J. H. and M. W. Watson (2008): "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression." *Econometrica*, 76, 155–174.

——— (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.

Suri, T. (2011): "Selection and Comparative Advantage in Technology Adoption." *Econometrica*, 79, 159–209.

Thürwächter, C. (2023): "Firm Heterogeneity and Monetary Policy Transmission." Working paper.

Townsend, R. M. (1994): "Risk and Insurance in Village India." *Econometrica*, 62, 539–591.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press.

Villacorta, L. (2021): "Estimating Unobserved Country Heterogeneity in Capital-Labor Substitution: A Nonlinear Panel Data Framework." Working paper.

Wu, W. (2018): "The Credit Channel at the Zero Lower Bound through the Lens of Equity Prices." *Journal of Money, Credit and Banking*, 50, 423–452.

Xu, K.-L. (2023): "Local Projection Based Inference under General Conditions." Working paper.

Xu, R. (2021): "Potential outcomes and finite-population inference for M-estimators." *Econometrics Journal*, 24, 162–176.

Yeh, C., C. Macaluso, and B. Hershbein (2022): "Monopsony in the US Labor Market." *American Economic Review*, 112, 2099–2138.

Zhou, W. (2023): "The Firm Balance Sheet Channel of Uncertainty Shocks." Working paper.