Heterogeneous Economies and Micro Data[*]

Man Chon (Tommy) Iao
New York University

Yatheesan J. Selvakumar
New York University

Preliminary, please do not circulate

October 21, 2024

Abstract

We give sufficient conditions under which dynamic equilibrium models with heterogeneous-agents can be represented by a first-order reduced-rank vector autoregression. We exploit this result to develop an econometric framework that enables the rapid estimation of a rich class of models with evolutions of both macro and large cross-section data. In monte-carlo simulations, we show that our method using the entire cross-section delivers precision up to an order of magnitude larger than the conventional approaches. We apply our method to estimate a medium-scale HANK model with heterogeneous exposures to aggregate fluctuations. Our estimates imply that poorer households are more sensitive to aggregate fluctuations on average, but the converse is true conditional on a monetary policy shock. Through the lens of the model, we estimate that heterogeneous earnings exposures amplify the consumption response to monetary policy shocks by 40% and the output response by 20%.

Keywords: structural estimation, singular value decomposition, dynamic mode decomposition, heterogeneous-agent models, earnings heterogeneity, vector autoregressions

JEL Classification Numbers: C13, C32, E1

[*]We thank Thomas J. Sargent, Jarda Borovicka, Simon Gilchrist, Corina Boar, Virgiliu Midrigan, Juan Rubio-Ramirez, Timothy Cogley, Diego Perez, Mark Gertler, Thomas Philippon, Venky Venkateswaran, Zejin Shi and Christopher Adjaho for their helpful comments

1 Introduction

Dynamic general equilibrium models featuring rich heterogeneity have become a popular branch of the macroeconomic literature in recent years. These models have gained traction in both academic and policy circles because they speak directly to the causes and consequences of inequality.

A defining property of this class of models it that they generate *different* predictions at the individual and aggregate levels, making the use of micro data more relevant compared to their representative-agent counterparts. In relating these models to the data however, the conventional literature has focused on matching limited features of the cross-section, usually summarized by a few moments at a particular point in time.[1] An important limitation of this strategy is that it runs the risk of "throwing away" important information contained not only in the individual-level observations but also in their evolution, thereby clouding inference.

The main hurdle to fully exploiting such information is that computing the likelihood of a dataset with thousands of individuals observed over decades would render most existing methodologies completely infeasible. We make progress by developing a tractable econometric framework that approximates the likelihood of the *entire* cross-section evolution. Our key assumption is that the heterogeneous-agent dynamic model has a Gaussian linear state-space representation, where the number of states are small relative to the size of the cross-section. We provide some favorable numerical evidence for both workhorse heterogeneous-agent models with aggregate shocks and consumption data in the Consumer Expenditure Survey.

Our main theoretical result is that such a model is well-approximated by a reduced-rank first-order vector autoregression.[2] The crucial insight is that as the size of the cross-section grows large, the information in the present cross-section is enough to precisely estimate contemporaneous hidden factors, effectively substituting for information contained in the infinite past. The result implies that a general VAR(∞) representation of a state-space model collapses to a VAR(1), which occurs at the rate rate equal to the size of the cross-section.[3] Our convergence result is useful because computing the reduced-rank first-order vector autoregression is fast and scales well with the size of the cross-section, thus opening the door to rapid Bayesian estimation of the structural

[1] Examples include the average proportion of wealth held by the top 1% in a sample, or matching quasi-experimental data on the marginal propensity to consume (MPC)

[2] Sargent and Selvakumar (2024) assert and use the result to infer hidden factors of the cross-section of income and consumption in the Consumer Expenditure Survey. The present paper provides a formal proof and uses the result in a structural estimaton setting.

[3] This result draws close parallels to existing insights of Chamberlain and Rothschild (1983) Forni et al. (2000), Forni and Lippi (2001).

parameters.[4] Our method is fairly flexible to model complexity. In principle, it only requires the ability to simulate cross-sectional data from the model, though as we show in our empirical application the availablity of closed form autocovariances can improve speed and accuracy even further. We prescribe ways to compute reduced-rank first-order VAR matrices in both cases.

We apply our method to two settings. First, we present a textbook Bewley-Aiyagari-Hugget economy with aggregate uncertainty (Krusell and Smith (1998)) economy adding a tax shock that redistributes income across households. To test-drive our method, we simulate aggregate and cross-section data from this economy and compute maximum likelihood estimates, in an attempt to recover structural parameters generating the data. Through a Monte-Carlo simulation exercise, we demonstrate a sharp improvement in estimation precision compared to a benchmark that uses aggregate data and the evolution of the second moment of consumption. We show that our method is equally precise[5] at recovering the parameters of the TFP shock process, but an order of magnitude more efficient at recovering those of the tax shock process.

Second, we present an empirical application in which we study the amplification of monetary policy shocks arising from *heterogeneous earnings exposures* – the differential elasticity of household income to aggregate shocks – through the lens of a medium-scale heterogeneous-agent New Keynesian model. We encode earnings heterogeneity in the model by setting household earnings equal aggregate income multiplied by an *incidence function* that determines the earnings sensitivity of households to changes in aggregate income.[6] The incidence function takes a flexible form, indexed by two parameters that we estimate. Our specification also allows for the possibility of differential exposures conditional on a monetary policy shock, which exemplifies the appeal of our method that enables joint estimation of all the structural parameters using detailed cross-section and aggregate data.[7]

Posterior parameter estimates using our method imply impulse responses that are closer to those estimated empirically using identified monetary policy shocks of **?** compared to those implied by aggregate data only. Furthermore, our estimates imply that low productivity households are more sensitive to fluctuations in aggregate income than high productivity households, in line with the existing literature. However, we find that conditional on a monetary policy shock, earnings of

[4]This is assuming that the researcher is able to, given a vector of parameters, solve the model sufficiently rapidly. Our algorithm ensures that the bottleneck remains at the solution step, rather than the likelihood computation step.

[5]This is computed as the standard deviation of maximum likelihood estimates across the monte-carlo samples.

[6]We follow in the steps of Werning (2015), Auclert and Rognlie (2018) and Alves et al. (2020).

[7]This features is motivated in part by the literature of uneven sectoral exposures to interest rate fluctuations.

high productivity households are more sensitive to aggregate fluctuations.[8] Finally, our estimated model generates an aggregate marginal propensity to consume (MPC) of 25% in the first quarter. Compared to a benchmark economy *without* heterogeneous earnings exposures, our estimates imply a consumption amplification of 40% and output amplification of 20%.

Section 2 situates our contributions within the context of related work. Section 3 outlines the econometric framework and lays out the theoretical foundations. Section 4 pursues a Monte-Carlo simulation exercise within the context of a Krusell-Smith model. Section 5 estimates a medium-scale HANK model to analyze the effect of earnings heterogeneity on consumption amplification to a monetary policy shock. Section 6 concludes.

[8] See Dolado et al. (2021) presents empirical evidence in line with this property.

2 Related Literature

Our paper contributes to the recent literature on structural estimation of heterogeneous-agent models using both macro and cross-section data. The paper closest to us is Plagborg-Møller (2019) who compute a numerically unbiased estimate of the likelihood. Crucial to their framework is a decomposition of the likelihood into a *macro part* and a *micro part, conditional on the macro state variables*, which are treated as unobserved. Thus, the macro data play an important role in determining the aggregate states. Our approach is silent about the properties of the aggregate states, but rather infers them jointly from *both* macro and micro data. Therefore, our approach is still viable were macro data not to be used. Second, their approach is better suited to a state-space representation of the model obtained from the Reiter (2009) model solution approach. Our approach is simulation based and so its convenience is not bound to any particular solution method.

Fernández-Villaverde et al. (2023) exploit the micro and macro data jointly in a model with financial frictions. However, they assume that the underlying states are observed, while our approach infers them from the macro and micro data jointly. Chang et al. (2021) compute functional vector autoregressions where the microdata are the density of the cross-section. They estimate the feedback loop betwen aggregate time series and micro-data, but do so without the context of fully-specified equilibrium model.

After solving a heterogeneous-agent model using the sequence space approach, Auclert et al. (2021a) compute the likelihood by exploiting the solution's implied moving average representation. Doing so requires vectorizing the $N \times T$ data matrix, where the corresponding covariance matrix is $NT \times NT$. In conventional situations with only aggregate data, N is relatively small (typically less than ten), so the covariance matrix is a computationally manageable. In setting that use the full cross-sectional data, N will undoubtedley be large making the inversion of the $NT \times NT$ covariance matrix would be prohibitively slow.[9] Our paper fills the gap by providing a solution in setting in which the conventional approach becomes infeasible.

Another approach is to use the Whittle likelihood approximation in frequency domain, as in Hansen and Sargent (1981) and Christiano and Vigfusson (2003). While estimation is feasible, its quality of the approximation relies on large T, which is unrealistic for existing micro-datasets. In contrast, our method relies on large N, which seems to us a far more satisfiable requirement.

Many papers have estimated structural models using solely macro data, including Winberry

[9] In our two examples below, $N = 86$ and $N = 300$; and $T = 120$, for 30 years of quarterly data.

(2018), Auclert and Rognlie (2018), Bayer and Luetticke (2020); others have also done so with some moments of the micro-data, including Acharya et al. (2020), Mongey and Williams (2017) and Challe et al. (2017). We employ our method to estimate a benchmark HANK model using the *entire* cross-section of consumption and income from the Consumer Expenditure Survey, as well as the conventional aggregate time-series. We are not aware of other papers that pursue this empirical exercise[10]. t

3 Econometric framework

Let $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$ be a linearized fully-specified dynamic general equilibrium model that governs the vector of observables $\mathbf{y}_t \in \mathbb{R}^{N \times 1}$. Similar to the vast literature on estimation of representative-agent models, it has become standard to consider $\mathbf{y}_t$ as a vector of aggregate observables, e.g. GDP, consumption, inflation, and so on. Recent papers have also added some cross-sectional moments to the vector of observables. Examples include the proportion of (households') wealth held by the top ten percent (Bayer and Luetticke (2020), the evolution of standard deviation of (firms') sales growth (Mongey and Williams (2017), or the frequency of (firms') price adjustments (Morales-Jimenez and Stevens (2024)).

Our goal is to extend this list of observables to include the full cross-section, as well as aggregate data. To bypass difficulties of computing the likelihood arising from size of the cross-section (large N), we propose an econometric framework that well-approximates $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$ in such situations. We begin by making the following assumption on the equilibrium Markov process $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$.

3.1 Data organization

A key ingredient in our methodology is the organization of the cross-section data. Let $y_{j,t} \in \mathbb{R}$ denote an **observable** (e.g. consumption) of individual $j = 1, \ldots,$ at time $t = 1, \ldots, T$. Through the lens of the model, individuals differ in their values of the idiosyncratic state vector, $\mathbf{z} \in \mathcal{Z}$ with cardinality N.[11] To make matters concrete, Krusell and Smith (1998) differentiate individuals through their asset holdings and productivity levels; Winberry (2021) differentiate firms through their productivity, capital, stock of depreciation allowances and their current draw of fixed costs. In our baseline theory, we will assume that the econometrician has observes the individuals'

[10] Plagborg-Møller (2019) show the benefits of using the cross-section only on simulated data

[11] We assume that any continuous idiosyncratic states, e.g. productivity, have been discretized appropriately, for example via the Rouwenhorst (1995) or Tauchen (1986) method.

idiosyncratic state vector. In Section XXX we discuss how one might extend our algorithm to accommodate situations in which the states are unobserved.

We denote $y_t(\mathbf{z})$ the consumption at time t of an individual with idiosyncratic state vector $\mathbf{z}$. The vector $\mathbf{y}_t = [y_t(\mathbf{z}_1), \ldots, y_t(\mathbf{z}_N)]^\top \in \mathbb{R}^{N\times 1}$ represents the observable at time t of individuals in each model-implied state.[12]

3.2 State-space representation

We begin our theoretical argument by stating our main assumption.

Assumption 1. *$\mathcal{M}_{\theta_0}(\mathbf{y}_t)$ has a Gaussian linear state-space representation,* ***and*** *the number of observables (N) are much greater than the number of states (r)*

$$\begin{aligned} \mathbf{f}_{t+1} &= \mathbf{A}\,\mathbf{f}_t + \mathbf{C}\,\mathbf{w}_{t+1} \\ \mathbf{y}_t &= \mathbf{G}\,\mathbf{f}_t + \mathbf{v}_t, \end{aligned} \tag{1}$$

for states $\mathbf{f}_t \in \mathbb{R}^{r\times 1}$, and observables $\mathbf{y}_t \in \mathbb{R}^N = N \times 1$; where shocks $\mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{r\times r})$, measurement errors $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and $\mathbf{w}_s \perp \mathbf{v}_\tau$ for all s, τ; here $\mathbf{A} \in \mathbb{R}^{r\times r}$, $\mathbf{C} \in \mathbb{R}^{r\times r}$ and $\mathbf{G} \in \mathbb{R}^{N\times r}$ and $\mathbf{R} \in \mathbb{R}^{N\times N}$.

That $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$ has a state-space representation is not controversial, it in fact forms the basis of many state-of-the-art solution methods.[13] The key difference is that the number of states are small relative to the size of the observable vector. At first glance, this might appear restrictive, since an equilibrium in dynamic heterogeneous-agent models with aggregate uncertainty typically involves forward looking agents that forecast prices that depend on the distribution of households. The exact solution therefore requires its inclusion into the state vector, making it infinite-dimensional. Krusell and Smith (1998) show that in their model, replacing the asset distribution with its first moment results in a ”good” approximation for equilibrium of their model. Our assumption is in a similar spirit, though not necessarily as stark. Moreover, the larger the cross-section, the larger the admissible number of states. In Appendix XXX we show that the workhorse heterogeneous-agent models with aggregate shocks satisfy this assumption.

This assumption is also easily testable, as we show in Section XXX and Appendix XXX. One simple numerical test is to simulate a long sequence $\{\mathbf{y}_t\}$ and compute principle components.

[12] Practically speaking, the researcher might not have observation at each point in $\mathcal{Z}$. Our framework is immune to this eventuality.

[13] Examples include Winberry (2018), Bayer et al. (2024), Auclert et al. (2021b)

Assumption 1 implies that the cross-section loads on only a few principle components. We further discuss this and other tests within the context of our applications, in Sections XXX and XXX. We place some additional restrictions on the Gaussian linear state-space model (1).

Assumption 2. *Linear state-space system* (1) *satisfies the following restrictions*

1. $\mathbf{G}, \mathbf{A}$ *has full column rank (i.e.* $rank(\mathbf{G}) = rank(\mathbf{A}) = r$*)*
2. $\left\|\mathbf{G}^\top \mathbf{G}\right\| = O(N)$*, where* $\|\cdot\|$ *denotes the Frobenius norm*
3. $\|\mathbf{R}\| = o(N)$

The first condition requires that the columns of $\mathbf{G}$ and $\mathbf{A}$ to be linearly independent. To gain some intuition, consider an economic model for which the rows of $\mathbf{y}_t$ contain consumption of different households, then columns of $\mathbf{G}$ represent the response of household's consumption to each of factor. Condition 2 then implies that responses of yi, t to each factor are sufficiently different across individuals. Where y_{it} is individual consumption, the assumption requires that there be enough heterogeneity in the responses to aggregate across households. A representative agent model would violate this assumption, for example. This condition highlights our identification strategy that exploits the rich heterogeneity in the cross-section. The assumption on $\mathbf{A}$ means that there is no redundant state. Importantly, the assumptions allow for higher-order lags for the states.

The second condition concerns the asymptotic property of the model when the number of observables grows and is standard in the factor analysis literature (e.g. Chamberlain and Rothschild 1982, Stock and Watson 2002, Bai and Ng 2006). To continue the examples of household consumption, 2. this implies that the cross-sectional variance of consumption is finite. The third condition requires that the variances do not grow too fast (not faster then N). As will become important in our empirical section, this assumption is weak enough to allow for heteroskedasticity and correlated measurement errors.[14]

The starting point for our theoretical analysis is the vector auto-regression representation of system (1).

[14]See Appendix XXX for further discussion.

Proposition 1. *There exists an infinite-order VAR representation of DFM (1) in $\mathbf{y}_t$, given by*

$$\mathbf{y}_t = \sum_{j=1}^{\infty} \mathbf{B}_j^{\infty}\, \mathbf{y}_{t-j} + \mathbf{a}_t \tag{2}$$

$$\mathbb{E}[\mathbf{a}_t\, \mathbf{y}_{t-j}^{\top}] = \mathbf{0} \quad \textit{for all } j \geq 1$$

$$\mathbb{E}[\mathbf{a}_t\, \mathbf{a}_t^{T}] =: \mathbf{\Omega}$$

$$\mathbf{B}_j^{\infty} = \mathbf{G}(\mathbf{A} - \mathbf{K}\,\mathbf{G})^{j-1}\,\mathbf{K} \quad \forall j \geq 1 \tag{3}$$

$$rank(\mathbf{B}_j^{\infty}) = r \quad \forall j \geq 1$$

where $\mathbf{K} = \mathbf{A}\,\mathbf{\Sigma}_{\infty}\,\mathbf{G}^{\top}\,\mathbf{\Omega}^{-1}$ *and* $\mathbf{\Sigma}_{\infty} = \mathbf{C}\,\mathbf{C}^{\top} + \mathbf{K}\,\mathbf{R}\,\mathbf{K}^{\top} + (\mathbf{A} - \mathbf{K}\,\mathbf{G})\mathbf{\Sigma}_{\infty}(\mathbf{A} - \mathbf{K}\,\mathbf{G})^{\top}$

Proposition 1 states the population formula for computing the the expectations of $\mathbf{y}_{t+1}$, given the information contained in its infinite history. Since $\mathbf{B}_j^{\infty} \neq \mathbf{0} \quad \forall j \geq 1$, it implies that a truncation induces a non-trivial loss in prediction accuracy. Our theoretical analysis studies what happens to this prediction accuracy as $N \to \infty$.

Lemma 1. *Under Assumption 2, as the number of observables grows ($N \to \infty$), the matrix* $\mathbf{A} - \mathbf{K}\,\mathbf{G} \to \mathbf{0}$.

Corollary 1. *When* $\mathbf{A} - \mathbf{K}\,\mathbf{G} = \mathbf{0}$, $\mathbb{E}[\mathbf{f}_t\,|\,\mathbf{y}^t] = \mathbf{L}\,\mathbf{y}_t$ *and* $\mathbb{E}[\mathbf{y}_{t+1}\,|\,\mathbf{y}^t] = \mathbf{G}\,\mathbf{K}\,\mathbf{y}_t$ *where* $\mathbf{L} = \mathbf{\Sigma}_{\infty}\,\mathbf{G}^{\top}\,\mathbf{\Omega}^{-1}$.

Lemma 1 and Corollary 1 taken together show that as $N \to \infty$, the estimate of the hidden state $\mathbf{f}_t$ conditional only the contemporaneous data $\mathbf{y}_t$ and that conditional on the full history of data $\mathbf{y}^t$ converge, i.e. when $\mathbf{A} - \mathbf{K}\mathbf{G}$, $\mathbb{E}[\mathbf{f}_t\,|\,\mathbf{y}^t] = \mathbb{E}[\mathbf{f}_t\,|\,\mathbf{y}_t]$. Furthermore, the Markovian property of the system implies that next period's forecasts computed conditional on $\mathbf{y}_t$ and $\mathbf{y}^t$ also converge.

Theorem 1. *Suppose Assumption 2 holds. Then as $N \to \infty$,*

1. $\mathbf{B}_j^{\infty} \to \mathbf{0}\ \forall j \geq 2$. *Furthermore,* $\limsup(N^{j-1}\,\|\mathbf{B}_j^{\infty}\|) < \infty\ \forall j \geq 2$

2. *The infinite-order VAR representation of (1) collapses to a first-order VAR representation where*

$$\mathbf{y}_t = \mathbf{B}_1^{\infty}\,\mathbf{y}_{t-1} + \mathbf{a}_t \tag{4}$$

$$\mathbb{E}[\mathbf{a}_t\,\mathbf{y}_{t-1}^{\top}] = \mathbf{0}$$

$$\mathbb{E}[\mathbf{a}_t\,\mathbf{a}_t^{T}] =: \mathbf{\Omega}$$

$$\mathbf{B}_1^{\infty} = \mathbf{G}\,\mathbf{K} \quad \forall j \geq 1$$

$$rank(\mathbf{B}_1^{\infty}) = r$$

Theorem 1 states that as N increases, all the auto-regressive coefficient matrices in the VAR(∞) representation other than the first lag converge to the zero matrix. Thus, the rank-r VAR(1) coincides with the VAR(∞) representation.

Theorem 2. *Let $\mathcal{P}_\rho(\mathbf{y}_t)$ be the reduced-rank vector auto-regressions (4) where $\rho \in P$ is the population maximum likelihood estimator. Suppose Assumption 1 and 2 holds. As $N \to \infty$, $\mathcal{P}_\rho(\mathbf{y}_t)$ approximates $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$ in the sense that the Kullback-Leibler divergence converges to zero*

$$\lim_{N\to\infty} \int \log \frac{\mathcal{M}_{\theta_0}(\mathbf{y}_t)}{\mathcal{P}_\rho(\mathbf{y}_t)} \mathcal{M}_{\theta_0}(\mathbf{y}_t) d\,\mathbf{y}_t = 0 \tag{5}$$

where we suppress the dependence on N for each object for notational clarity.[15] Theorem 1 states that as N becomes large, the structural heterogeneous-agent equilibrium model is well-approximated by a reduced-rank VAR(1). Since computing (4) is fast, this theorem opens the door to rapid computation of the likelihood $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$, intermediated through the reduced-rank VAR(1). Though the logic justifying our algorithm involves the state-space systems with hidden factors, our algorithm conveniently does not require estimating them, only requiring the reduced-rank VAR(1).

3.3 Reduced-rank first-order VAR

Our theoretical results in the previous subsection justify a fast algorithm for approximating the $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$. In this section, we explore two approaches to compute the first-order reduced-rank VAR

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}\,\mathbf{y}_{t-1} + \mathbf{a}_t \\ rank(\mathbf{B}) &= r \\ \mathbb{E}[\mathbf{a}_t\,\mathbf{a}_t^\top] &= \mathbf{\Omega} \end{aligned} \tag{6}$$

The first approach uses canonical correlations analysis of Anderson and Rubin (1949) and Anderson (1951), which we adopt in our main algorithm and describe here. Its main benefit is that it obtains analytic expressions for population $\mathbf{B}$ and $\mathbf{\Omega}$, thus improving on accuracy and speed.[16]

[15] As is made clear in the proof in Appendix XXX, each object in this integral is itself dependent on N, i.e. $\mathcal{M}^N_{\theta_0^N}(\mathbf{y}_t^N)$, and $\mathcal{P}^N_{rho^N}(\mathbf{y}_t^N)$. The theorem therefore states that both the "goal" and the "approximation" are moving as N increases, and that in the limit, they converge to the same object.NEED TO IMPROVE.

[16] The second uses a computationally efficient algorithm that may be used when the structural model is able to be simulated from, but does not admit analytic expressions of population covariances. The approach builds on the Dynamic Mode Decomposition, which we describe in Appendix B.1.

Let the population covariance matrices be defined by

$$\mathbf{\Sigma}_0 = \mathbf{E}[\mathbf{y}_t \, \mathbf{y}_t^\top] \quad \mathbf{\Sigma}_1 = \mathbf{E}[\mathbf{y}_t \, \mathbf{y}_{t-1}^\top] \tag{7}$$

The canonical correlation and variates in the population are defined by

$$\begin{pmatrix} -\rho\mathbf{\Sigma}_0 & \mathbf{\Sigma}_1 \\ \mathbf{\Sigma}_1^\top & -\rho\mathbf{\Sigma}_0 \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = 0 \tag{8}$$

where ρ, α, γ satisfy

$$\begin{vmatrix} -\rho\mathbf{\Sigma}_0 & \mathbf{\Sigma}_1 \\ \mathbf{\Sigma}_1^\top & -\rho\mathbf{\Sigma}_0 \end{vmatrix} = 0, \quad \alpha'\mathbf{\Sigma}_0\alpha = 1, \quad \gamma'\mathbf{\Sigma}_0\gamma = 1 \tag{9}$$

Then, the reduced-rank VAR matrix is given by

$$\mathbf{B} = \mathbf{\Sigma}_1\mathbf{\Gamma}\mathbf{\Gamma}^\top \tag{10}$$

where $\mathbf{\Gamma} = [\gamma_1, \ldots, \gamma_r]^\top$. Furthermore, the covariance matrix of the residuals is given by

$$\mathbf{\Omega} = \mathbf{\Sigma}_0 - \mathbf{B}\,\mathbf{\Sigma}_0\,\mathbf{B}^\top \tag{11}$$

In both applications below, we employ the Sequence-Space Jacobian (SSJ) solution method of Auclert et al. (2021b). That the solution provides a moving average representation of the observables means that we can easily conveniently compute population covariances and therefore reduced-rank population projection coefficients using the algorithm above. See Section 5.4 for details.

For other solution methods that do not admit analytic population covariance, simulation is still available. To do so, simulate a long sequence of $\{\mathbf{y}_{t+j}\}_{j=1}^{J}$ from $\mathcal{M}_\theta(\mathbf{y}_t)$, to approximate $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ and proceed.

After computing $\mathbf{B}$ and $\mathbf{\Omega}$, the likelihood computation is trivial, given by

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_{T+1}) = \sum_{t=1}^{T} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\det(\mathbf{\Omega}) - \frac{1}{2}(\mathbf{y}_t - \mathbf{B}\,\mathbf{y}_{t-1})^\top\mathbf{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\,\mathbf{y}_{t-1}) \tag{12}$$

3.4 The number of factors

A natural question in our strategy is the choice of the number of factors in the linear state-space model representation. In this section, we suggest a multitude of heuristic and quantitative procedures that offers insights into the appropriate number of factors. In our experience, we have found that our algorithm scales well with r, so one might choose to err on the side of caution and choose a large r rather than a too small one.

Check 1. From the principle components literature, Bai and Ng (2002) show that consistent estimation of the number of factors in the data can be attained by minimizing the information criterion[17]

$$IC(n) = V(n) + n\left(\frac{M+T}{MT}\right)\log\left(\frac{MT}{M+T}\right) \tag{13}$$

where $V(n) = (MT)^{-1}\sum_{i=1}^{M}\sum_{t=1}^{T}(a_{it}^{n})^2$.

Check 2. Another approach is to simulate the cross-sectional data from $\mathcal{M}(\theta)$ and study the decay rate of the associated singular values. This approach is a common heuristic used by Dynamic Mode Decomposition (DMD) practitioners[18,19]. An appropriate choice of r is the number after which the decay rate of the singular values is close to zero.

Check 3. An additional implication of Proposition 1 is that the VAR residuals $\tilde{\mathbf{a}}_t$ are serially uncorrelated. Since there is nothing in the first-order VAR that imposes such a restriction, it follows from the innovations representation of the DFM (1). Given a simulated data, compute the reduced-rank VAR residuals

$$\tilde{\mathbf{a}}_t = \widetilde{\mathbf{y}}_t - \mathbf{B}_r\,\widetilde{\mathbf{y}}_{t-1} \tag{14}$$

We construct the sample covariance matrix (15) and choose the r that minimizes the the sample covariance matrix of the residuals.

[17] Though the theoretical analysis in Bai and Ng (2002) is done for principle components estimation of factor models, the same theory applies to any other consistent estimation procedure, as $M, T \to \infty$.

[18] Appendix B.1 describes the DMD algorithm to compute efficiently compute reduced-rank VAR matrices from data

[19] See, for example, Brunton and Kutz (2022, sec. 7.2))

$$\widehat{E}[\mathbf{a}_{t+1}\,\mathbf{a}_t^\top] = \frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{a}}_{t+1}\widetilde{\mathbf{a}}_t^\top \tag{15}$$

Check 4. The final check is to compute a metrics for the "fit" of the VAR to the data, akin to a modified R^2. Given simulated data $\widetilde{\mathbf{Y}}$, and a fixed r, compute the VAR residuals in equation (14)

Denote $R^2_{m,r}$ as the individual-level R^2 for the VAR regression for $m = 1, \ldots, M$ (i.e. for each row of $\mathbf{y}_t$), given by

$$R^2_{m,r} = 1 - \frac{\sum_{t=2}^{T}\tilde{a}^2_{m,t}}{\sum_{t=2}^{T}\widetilde{\mathbf{y}}_{m,t} - \frac{1}{T}\sum_{t=2}^{T}\widetilde{\mathbf{y}}_{m,t-1}}$$

where $\widetilde{a}_{m,t}$ is the m-th element of $\widetilde{\mathbf{a}}_t$. Then, calculate the aggregate R^2_r of the approximating model by a weighted sum of the cross-sectional R^2_m for $m = 1, \ldots, M$.

$$R^2_r = \tfrac{1}{M}\sum_{m=1}^{M} w(m) R^2_{m,r} \tag{16}$$

where $w(m)$ is some weighting function.[20] The appropriate r is the smallest value that maximizes R^2_r.

3.5 Algorithm

The above sections set out the theoretical and computational arguments for our estimation strategy. To recap succinctly, the logic is as follows: Assumption 1 means that $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$ has a linear state-space representation with $r \ll N$ factors. Using Assumption 1, we show that a rank-r first-order VAR approximates the linear state-space model (and therefore $\mathcal{M}_{\theta_0}(\mathbf{y}_t)$) when N is large, and that the approximation error vanishes at rate N. One might compute the likelihood using either the canonical correlations method, or the DMD method, depending on the application.

Algorithm 1 presents pseudo-code for approximating the likelihood of observable data $\{\mathbf{y}_t\}$ implied by $\mathcal{M}$.

[20] In our example below, we fix an equal weighting scheme, and sample individuals from the stationary distribution of $\mathcal{M}$.

Algorithm 1 Likelihood approximation

1. Fix some structural parameters θ
2. Choose the rank, r, as discussed in section 3.4
3. If population covariances are computable:
   - Compute $\mathbf{B}_r(\theta)$ and $\mathbf{\Omega}_r(\theta)$ using the equations (10) and (11)
4. If not:
   - Simulate time-series $\tilde{\mathbf{y}}_1(\theta), \dots, \tilde{\mathbf{y}}_{J+1}(\theta)$ from $\mathcal{M}$ for a large J and create data matrices $\widetilde{\mathbf{Y}}(\theta)$ and $\widetilde{\mathbf{Y}}'(\theta)$[21]
   - Calculate $\mathbf{B}_r(\theta)$ and $\mathbf{\Omega}_r(\theta)$ in (B.2) and (B.3)
5. Approximate the log- likelihood $f(\mathbf{y}_1, \dots, \mathbf{y}_{T+1} \,|\theta)$ implied by $\mathcal{M}$ by computing (12)

3.6 Bayesian estimation

Our likelihood approximation algorithm can be easily paired with any monte-carlo sampling scheme to approximate the posterior distribution of the parameters. Under Assumptions 1 and 2, our theorems imply that our approximation error of posterior also vanishes as the cross-section becomes large ($N \to \infty$). In the empirical example in Section 5, we pair our algorithm with a Random Walk Metropolis Hastings algorithm. Algorithm 2 provides a pseudo-code for one iteration of the RWMH sampler.

Algorithm 2 One iteration of Random Walk Metropolis Hastings

For iteration n with structural parameter θ^{n-1}:

1. Draw $\theta^* \sim q(\cdot|\theta^{n-1})$
2. Approximate likelihood $f(\mathbf{Y}\,|\theta^*)$ using Algorithm 1
3. Compute $r = \min\left\{1, \frac{f(\mathbf{y}_1,\dots,\mathbf{y}_{T+1}\,|\theta^*)p(\theta^*)}{f(\mathbf{Y}\,|\theta^{n-1})p(\theta^{n-1})}\right\}$
4. Accept θ^* with probability r
5. **if** accept, $\theta^n = \theta^*$, **else** $\theta^n = \theta^{n-1}$

3.7 Difficulties with other possible approaches

Kalman filter An obvious question at this point is why not simply evaluate the likelihood of the linear-state space represenation (1) with the Kalman filter? The answer is that evaluating the likelihood via the Kalman filter requires knowing the matrices $\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}$, which itself must be estimated from the simulated data. To see why this might be a problem, consider an example where $N = 300$ and $\mathcal{M}$ is approximately rank $r = 2$. Then one needs to estimate the 606 parameters of $\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}$ and then use them to compute the likelihood of the data.[22] Since the matrices also depend on structural parameters, one would need to insert an additional loop for each iteration of the estimation procedure, making it computationally inefficient.

Moving average representation

State-space representation

Frequency domain

Plagborg-Møller (2019)

4 Krusell and Smith laboratory

4.1 Model description

The economy is populated by heterogeneous households of mass 1 that maximize their infinite-limetime utility $\mathbb{E}[\sum_{t=0}^{\infty} \beta^t u(c_t)]$ where utility function $u = \frac{c^{1-\phi}}{1-\phi}$. Each household faces two idiosyncratic states: productivity ε and assets a. We assume that productivity follows an exogenous markovian process, where the probability of moving from ε to ε' is given by $\pi(\varepsilon, \varepsilon')$. All households work an exogeneous amount of hours n, normalized to equal 1. Households earnings are $w\varepsilon_t$ where w denotes the hourly efficiency wage, common to all households. Agents can choose to save in assets a, which have a rate of return R. Households also face a borrowing constraint $a \geq \bar{a}$, which we will set to zero.

Output in this economy is produced by a representative firm that takes uses aggregate capital K_t and labor L_t as inputs using a Cobb-Douglas production function $Y_t = Z_t K_t^{\alpha} L_t^{1-\alpha}$, where Z_t is aggregate total factor productivity. One unit of labor costs the firm w_t, and one unit of capital costs the firm R_t. We will assume that TFP follows an AR(1) process.

$$Z_{t+1} = \rho_z Z_t + \sigma_z \nu_{t+1} \tag{17}$$

[22] $\mathbf{G}$ has 300×2 parameters, $\mathbf{R}$ has one parameter, $\mathbf{A}$ has 2 parameters and $\mathbf{C}$ has 3 parameters.

where $\nu_{t+1} \sim \mathcal{N}(0,1)$. Let $\Lambda_t(a,\varepsilon)$ denote the mass of households with assets a and idiosyncratic productivity ε. Households need to forecast prices R and w that depend on both aggregate productivity and the distribution of assets in the economy. Thus, the household has two idiosyncratic (a,ε) and two aggregate states (Z_t,Λ_t). Let $V(a,\varepsilon;Z_t,\Lambda_t)$ be the value function of household with assets a, productivity ε and that faces aggregate productivity Z_t and the aggregate distribution of wealth Λ_t. The household's problem is given by

$$
\begin{aligned}
V(a,\varepsilon;Z_t,\Lambda_t) &= \max_{\{c,a'\}} \frac{c^{1-\phi}}{1-\phi} + \beta \int V(a',\varepsilon';Z_{t+1},\Lambda_{t+1})\pi(\varepsilon'|\varepsilon)d\varepsilon' dF(Z_{t+1}|Z_t) \\
c(a,\varepsilon;Z_t,\Lambda_t) + a'(a,\varepsilon;Z_t,\Lambda_t) &= y(\varepsilon_t;Z_t,\Lambda_t) + R(Z_t,\Lambda_t)a \\
y(\varepsilon;Z_t,\Lambda_t) &= w(Z_t,\Lambda_t)\varepsilon \\
\Lambda_{t+1} &= \Psi(Z_t,\Lambda_t)
\end{aligned}
$$

where Ψ denotes the law of motion of the endogenous distribution of households, and where we make explicit the dependence on aggregate states.

Market clearing implies that

1. the goods market clears

$$\int c(a,\varepsilon;Z_t,\Lambda_t)\Lambda_t(a,\varepsilon)dad\varepsilon + \int a'(a,\varepsilon;Z_t,\Lambda_t)\Lambda_t(a,\varepsilon)dad\varepsilon = Y_t + (1-\delta)K_t \tag{18}$$

2. the labor market clears

$$L_t = \int \varepsilon\Lambda_t(a,\varepsilon)dad\varepsilon \tag{19}$$

3. the asset market clears

$$K_t = \int a\Lambda_t(a,\varepsilon)dad\varepsilon \tag{20}$$

Our goal will be to estimate the shock parameters of this model, $(\rho_z,\sigma_z,\rho_\xi,\sigma_\xi)$ using both aggregate and cross-section data. In the next two sections, we outline how we organize the thousands of micro datapoints, as well as our theoretical results that underpin our algorithm to perform such an estimation.

4.2 Organizing the cross-section data

The household's consumption policy function, similar to the value function, depends on the two idiosyncratic states and the two aggregate states. Since the aggregate states is common for all households, let us suppress its notation by defining the consumption policy function at time t by $c_t(a, \varepsilon) \equiv c(a, \varepsilon; Z_t, \Lambda_t)$. This notation makes clear that through the lens of the model, individuals respond differently to aggregate states only through their differences in idiosyncratic states. We use this insight to organize the micro data by binning individuals according to their idiosyncratic states in each period. Within this Krusell and Smith model example, this requires discretizing the state-space into N points of the $a \times \varepsilon$ state space; and grouping individuals into associated bins. We will define the consumption of the bin (a, z) as the mean consumption of individuals inside that bin. Repeating this for each N bins, we obtain our cross-section vector $(c_t(\mathbf{z}_1), c_t(\mathbf{z}_2), \ldots, c_t(\mathbf{z}_N))^\top$. Section XXX discusses how to accommodate unobserved idiosyncratic states.

4.3 Connecting our theory

The main assumption underpinning our algorithm is that the equilibrium model has a Gaussian linear state-space representation with relatively few states. Notwithstanding the rank condition, this statement is uncontroversial in the standard heterogeneous-agent models. Figure 1 validates this assumption in the Krusell and Smith model, by plotting simulated paths of consumption for 1000 households over 120 quarters.[23] Although there are idiosyncratic divergences, the evolving cross-section appears to exhibit a common factor. This observation is confirmed by computing principle components of the data. The red line in the left chart plots the first principle component of the data, and the right chart plots the contribution to total variance of the 10 leading principle components. It again suggests the presence of one dominant common factor, $r = 1$. That $N = 1000$ in this setting implies that our $N >> r$ assumption is satisfied.[24]

Under Assumptions 1 and 2, Corollary 1 states that as the size of the cross-section goes large, $\mathbb{E}[\mathbf{f}_t | \, \mathbf{y}^t] \approx \mathbb{E}[\mathbf{f}_t | \, \mathbf{y}_t]$. This implies that the hidden factor can be approximately well inferred from contemporaneous data compared to its infinite past. To see the theorem at work in our laboratory, conduct the following experiment. We simulate an economy with 1000 individuals from the

[23]See Appendix XXX for the calibration underlying this exercise.

[24]Appendix XXX performs the battery of checks outlined in Section 3.4, providing further evidence that $r = 1$.

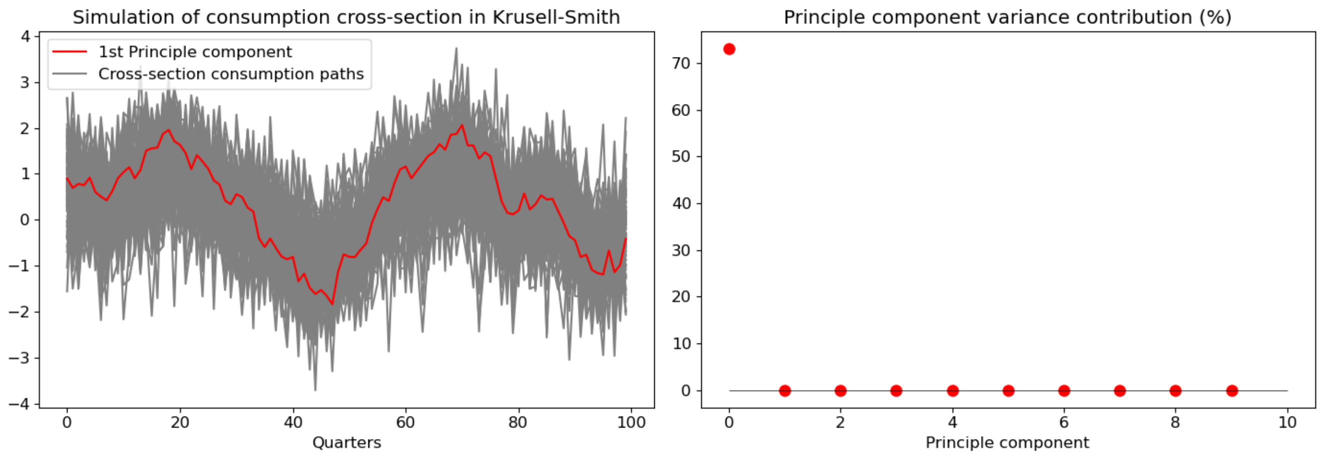

Figure 1: Evolution of cross-section of consumption in a simulated Krusell and Smith economy

structural model above. We imagine an econometrician inferring hidden states by only observating data pertaining to N individuals. The econometrician then computes principle components of the observed data.[25] Figure 2 shows the estimates hidden factors in for $N = 5$ (gray) and $N = 500$ (blue), and the red line is the ”true” TFP series generating the data All series are normalized to unit variance. Indeed the econometrician is better able to estimate the contemporaneous factor as the cross-section increase in size. Moreover, the better estimate also allows a better estimate of the stochastic process of TFP, the estimated AR(1) coefficient of the factor associated with $N = 5$ is 0.90, while $N = 500$ is 0.96 and the true parameter value is 0.95.

4.4 Monte-carlo simulation

We use the Krusell and Smith model as a laboratory to test-drive our theory in a controlled environment. Calibrating the model as in Appendix XXX, we simulate consumption paths of 300 households drawn from the steady-state distribution, as well as aggregate output, nominal interest rate and aggregate consumption for 120 quarters. We choose these dimensions to closely replicate what researchers might face in practice. We employ our algorithm to compute maximum likelihood estimates of the parameters of the shock processes in an effort to recover the values that generated the data. We repeat this exercise 500 times to approximate the finite-sample distribution of the maximum likelihood estimator. Table 1 shows the mean and standard deviation of the finite-sample distributions, and compares them against two alternatives common in the literature: 1) using only aggregate (output, consumption, interest rate), called `Agg`) and 2) aggregate data *and* the second

[25]Importantly for the validity of our exercise, principle components does not assume any particular dynamic process for the data, and so only uses concurrent data to estimate factors.

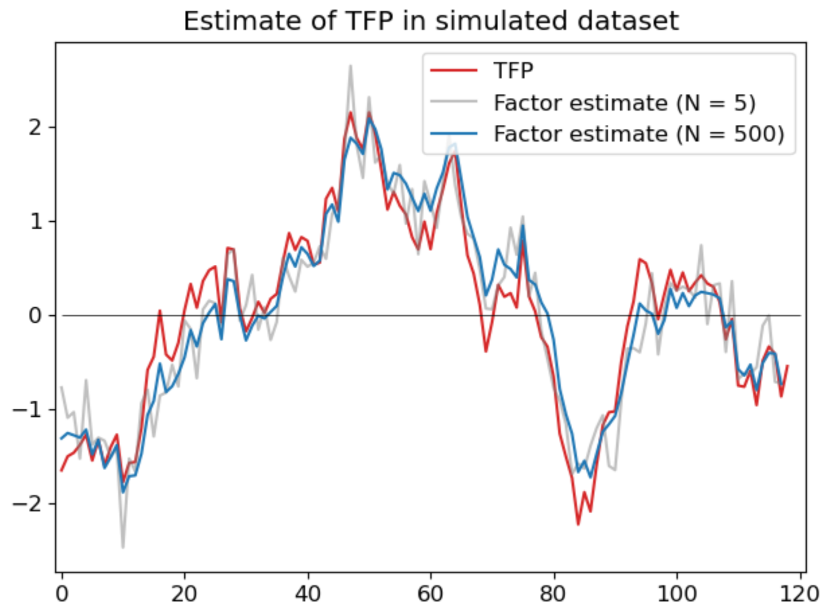

Figure 2: Estimate of hidden factor in simulated dataset

moment of the consumption distribution, called `Agg+`. We term the estimation with the micro data `Micro`. The values colored in red depict mean of the estimates that are closest to the true value (bias), and the values in blue denote estimates that have the smallest standard deviation (efficiency). In all cases bar one, the `Micro` estimates have means closest to the true (least biased), and the smallest dispersion (most efficient).

Figure 3 plots the histogram of the distributions with `Agg` in gray, `Agg+` in blue and `Micro` in red. In the top panel is those for ρ_z and σ_z. The finite-sample distributions look similar across all three estimations, suggesting that there isn't much to be learned from the additional cross-section data in identifying the stochastic process for TFP. The story is much different for the parameters of the redistribution shock (bottom panel). Estimating the model with aggregate data on its own leads to severely biased and dispersed parameter estimates. For example, the estimates for ρ_z ranges from 0.5 to 1.0, with very little mass around the true estimate.[26] Adding the variance in consumption (`Agg+`) improves the bias and dispersion of the estimates. Introducing the cross-section (`Micro`) dramatically reduces the bias and the dispersion in the estimates, leading to more efficient and accurate inference.[27]

[26]The (gray) histogram bunches at 0.5 and 1.0 for ρ_z, which are the bounds used for the optimizer. The estimates would therefore be even more dispersed were we to allow for looser bounds.

[27]That marginal propensities to consume (MPCs) differ across households in the economy implies that ξ_t is formally identified in this model – a redistribution of income from high to low productivity households itself generates an aggregate response. In that sense, estimating the model on only aggregate data is not a "strawman".

			Agg		Agg+		Micro	
Shock	Parameter	True value	Mean	Std	Mean	Std	Mean	Std
TFP	ρ_z	0.95	0.95	0.005	0.95	0.004	0.95	0.005
	σ_z	0.5	0.52	0.04	0.51	0.04	0.50	0.03
Redistribution	ρ_ξ	0.8	0.72	0.21	0.77	0.1	0.80	0.02
	σ_ξ	0.3	0.07	0.08	0.31	0.05	0.30	0.03

Table 1: Finite-sample parameter distribution with 500 monte-carlo samples

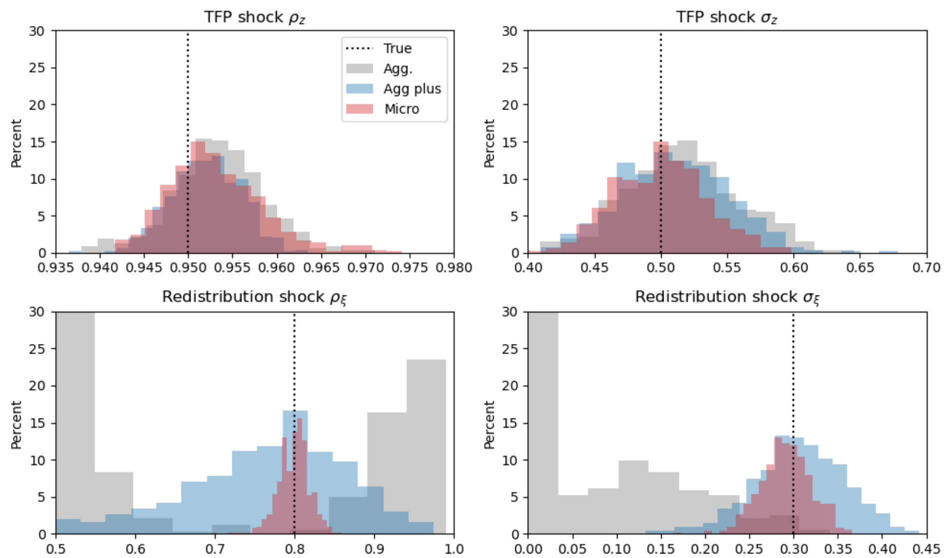

Figure 3: Finite sample parameter distribution across 500 monte-carlo samples

While the qualitative result – that cross-section data is more informative of parameters that affect the cross-section – its magnitude, even in this very simple model is noteworthy. It is possible that in models with much richer interactions between the cross-sections and aggregates, the addition of the cross-section may also help in more accurately estimating other aggregate parameters than would otherwise be the case. In Appendix XXX we repeat this exercise with a much σ_ξ relative to σ_z. We show that in that economy, the cross-section data also aids in pinning down TFP, further highlighting the benefits of incorporating cross-section data in formal parameter estimation.

4.5 Results

5 Application: Medium-scale HANK with earnings heterogeneity

We consider a medium-scale heterogeneous agent model, with aggregate shocks that reminiscent of the representative-agent literature (e.g. Christiano et al. 2005, Smets and Wouters 2007). As has been documented in the previous literature, the unequal incidence of aggregate shocks and the consequences thereof are a major source of difference between HANK models and RANK model. As such, the model features two important types, in a similar spirit to Alvez et al. (2020), which will subsequently be estimated with micro data.

5.1 Model

Time is discrete and runs forever, $t = 0, 1, \dots$.

5.1.1 Households

There is a unit measure of infinitely-lived households in the economy. At time t, a household consumes c_t and supplies hours of labor h_t, chosen by a labor union (see Section XXX). They receive per period utility

$$U(c_t, h_t) = \frac{c^{1-\sigma}}{1-\sigma} - \varphi \frac{h_t^{1+\phi}}{1+\phi} \tag{21}$$

where σ is the inverse of the EIS and ϕ is the inverse of the Frisch elasticity of labor supply. A household earns three kinds of income: labor, dividend and asset income. Labor earnings $y_t(z)$ and dividend income $D_t(z)$ depend on the household's idiosyncratic stochastic labor productivity z and are given by

$$y_t(z) = w_t h_t \Gamma_t(z) \tag{22}$$

$$D_t(z) = D_t z \tag{23}$$

where w_t is the real wage and D_t is the aggregate dividend in the economy, defined below. We normalize productivity to integrate to one, $\int_0^1 z_{it} di = 1$.

The function $\Gamma_t(z)$ is commonly called the "incidence function" and will play a central role in our analysis.[28] We consider Γ_t as a generalization of the standard setup in which $\Gamma_t(z) = z$ and therefore $y_t(z) = w_t h_t z$, where $h_t z$ is interpreted as *efficiency hours*. In such settings, z plays a dual role. On the one hand, it encodes that more productive households have higher earnings (rationalizing the interpretation of efficiency hours); on the other, it implies that they are more sensitive to changes in w_t and therefore experience more volatile income over the business cycle. Previous papers have estimated regressions at the micro level, inferring that low income households have more volatile income that other groups.[29] We follow Alves et al. (2020) by replacing z with

[28]See Werning (2015),Auclert and Rognlie (2018)

[29]See for example Guvenen et al. (2017)

$\Gamma_t(z)$ in order to be consistent the empirical findings. Unlike Alves et al. (2020) who estimate the parameters of Γ_t using external regressions on micro-data, we will internally estimate the parameters of Γ_t jointly with the other structural parameters of the model using cross-sectional data, which is made feasible by our likelihood approximation algorithm. See Section 5.2 for parameteriztion of Γ_t.

Post-tax earnings of the household take the Heathcote et al. (2017) form of $(1-\tau^y)y_t(z)^{1-\xi}$ where τ^y determines the average level of the marginal tax rate and ξ determines its slope. Dividends are also taxed at rate τ_t^D. Finally, households can save and borrow through a risk-free asset, subject to an ad-hoc borrowing constraint $a_t \geq \underline{a}$, which earns the real interest rate r_t. Let $V_t(a, z)$ denote the value function of a household with assets a and productivity z. Given the above setup, its Bellman equation is given by

$$V_t(a, z) = \max_{c,a'} \left\{ \frac{c^{1-\sigma}}{1-\sigma} - \varphi \frac{h_t^{1+\phi}}{1+\phi} + \beta \mathbb{E}_t \left[V_{t+1}(a', z') | z \right] \right\} \tag{24}$$

$$c + a' = (1 - \tau_t^y) y_t(z)^{1-\xi} + (1 - \tau_t^D) D_t z + (1 + r_t) a \tag{25}$$

$$y_t(z) = w_t h_t \Gamma_t^y(z) \tag{26}$$

$$a' \geq \underline{a}$$

5.1.2 Firms

Labor unions We closely follow Auclert et al. (2024) in what has become a standard setup in the New Keynesian sticky wage literature. We assume that labor hours h_{it} are determined by the labor demand of unions, a continuum of which operate in a monopolistically competitive market. Each union k aggregates the efficient hours of its workers to a union-specific task

$$n_{kt} = \int h_{ikt} \Gamma_t^h(z_i) di \tag{27}$$

A competitive labor packer then aggregates labor hours across the unions with constant elasticity of substitution ϵ_w

$$N_t = \left(\int n_{kt}^{\frac{\epsilon_w - 1}{\epsilon_w}} dj \right)^{\frac{\epsilon_w}{\epsilon_w - 1}} \tag{28}$$

and then sells it to intermediate goods firms for real wage w_t. Each union k sets a common real wage w_{kt} amongst all its members subject to a quadratic adjustment cost à la Rotemberg (1982) on nominal wages. We restrict the union to a uniform labor allocation rule, i.e. $n_{ikt} = n_{kt}$. This implies that the union sets the real wage w_{kt} to maximize the average utility of its members. The union's problem is therefore given by

$$
\begin{aligned}
\max_{w_{kt+l}, n_{kt+l}} \quad & \mathbb{E}_t \sum_{k=0}^{\infty} \beta^l \left\{ \left[(C_{t+l})^{-\sigma} (1-\tau_t^y) w_{kt+l} - \varphi N_{t+l}^{\phi} \right] n_{kt+l} - \frac{\epsilon_w}{2\kappa_w} \log \left(\frac{w_{kt+l}}{w_{kt+l-1}} \pi_{t+l} \right)^2 \right\} \\
s.t. \quad n_{kt+l} &= \left(\frac{w_{kt+l}}{w_{t+l}} \right)^{-\epsilon} N_{t+l} \\
n_{kt+l} &= \int h_{kt+l,i} \Gamma_t^y(z_i) \, di \\
N_{t+k} &= \left(\int n_{kt+l}^{\frac{\epsilon_w - 1}{\epsilon_w}} \, dk \right)^{\frac{\epsilon_w}{\epsilon_w - 1}}
\end{aligned}
$$

where the union takes as given the packer demand as a function of its relative real wage $\frac{w_{kt}}{w_t}$. This setup implies that all unions set the same real wage, i.e. $w_{kt} = w_t$ and all households are demanded the same efficient hours, i.e. $n_{kt} = N_t$, and so the first order condition of the labour union leads to the wage Phillips curve

$$
\pi_t^w = \kappa_w \left[\varphi N_t^{\phi} - \frac{\epsilon_w - 1}{\epsilon_w} (1-\tau_t^y)(C_t)^{-\sigma} w_t \right] N_t + \beta \, \mathbb{E}_t \, \pi_{t+1}^w + v_t^w \tag{29}
$$

where v_t^w is a AR(1) wage-markup shock.

Final-good producers A perfectly competitive representative final-good producer aggregates the continuum of retail firms $j \in (0,1)$ with constant elasticity of substitution ϵ

$$
Y_t = \left(\int_0^1 y_{jt}^{\frac{\epsilon-1}{\epsilon}} \, dj \right)^{\frac{\epsilon}{\epsilon-1}}
$$

Taking prices as given, the demand for intermediate good j is given by

$$
y_{jt} = \left(\frac{p_{jt}}{P_t} \right) Y_t, \quad \text{where} \quad P_t = \left(\int_0^1 p_{jt}^{1-\epsilon} dj \right)^{\frac{1}{1-\epsilon}} \tag{30}
$$

Intermediate-good firms We follow closely a standard specification of intermediate-good firms. Intermediate-good firms operate in a monopolistically competitive market. Each differentiated good j is produced with a Cobb-Douglas production function

$$y_{jt} = \Theta_t k_{jt}^{\alpha} l_{jt}^{1-\alpha} \tag{31}$$

Firms hire capital and labor at prices r_t^k and w_t and pay a fixed cost Ξ. In choosing prices p_{jt}, they take into account their demand by the final good firm and Rotemberg (1982) quadratic price adjustment costs Ψ_{jt}^p on their price inflation relative to a fraction ι_p of the price changes last period.

The Bellman equation for the retail firms is consequently

$$J_t^R(p_{jt-1}, p_{jt-2}) = \max_{k_{jt}, l_{jt}, y_{jt}, p_{jt}} \left\{ \frac{p_{jt}}{P_t} y_{jt} - w_t n_{jt} - r_t^k k_{jt} - \Psi_{jt}^p - \Xi + \mathbb{E}_t \left[\frac{J_{t+1}^R(p_{jt}, p_{jt-1})}{1+r_t} \right] \right\}$$

subject to

$$\begin{aligned}
y_{jt} &= \Theta_t k_{jt}^{\alpha} l_{jt}^{1-\alpha} \\
y_{jt} &= \left(\frac{p_{jt}}{P_t} \right)^{-\epsilon} Y_t \\
\Psi_{jt}^p &= \frac{\psi_p}{2} \left[\log \left(\frac{p_{jt}}{p_{jt-1}} \right) - \iota_p \log \left(\frac{p_{jt-1}}{p_{jt-2}} \right) \right]^2 Y_t
\end{aligned}$$

The first order condition of the firms imply constant capital-labor ratios

$$\frac{(1-\alpha)}{\alpha} \frac{r_t^k}{w_t} = \frac{N_t}{K_t}$$

In a symmetric equilibrium where all firms choose $y_{jt} = y_t, k_{jt} = k_t = K_t$ and $l_{jt} = l_t = N_t$ implies identical marginal costs given by

$$mc_t = \frac{1}{\Theta_t} \left(\frac{r_t^k}{\alpha} \right)^{\alpha} \left(\frac{w_t}{1-\alpha} \right)^{1-\alpha}$$

giving rise to the aggregate production function

$$Y_t = \Theta_t K_t^{\alpha} N_t^{1-\alpha} \tag{32}$$

and the aggregate price Phillips curve with associated price markup shock v_t^p that follows an AR(1) process.

$$\pi_t - \iota_p \pi_{t-1} = \underbrace{\frac{\psi_p}{\epsilon}}_{\kappa_p} \left(mc_t - \frac{\epsilon - 1}{\epsilon} \right) + \frac{1}{1 + r_t} \mathbb{E}_t \left[\frac{Y_{t+1}}{Y_t} (\pi_{t+1} - \iota_p \pi_t) \right] + v_t^p \tag{33}$$

Furthermore, aggregate resources spent of adjustments costs are

$$\Psi_t^p = \frac{\psi_p}{2} \left[\log(\pi_t) - \iota_p \log(\pi_{t-1}) \right]^2 Y_t$$

and profits from the intermediate firms, distributed as dividends are given by:

$$D_t^R = Y_t - w_t N_t - r_t^k K_t - \Psi_t^p$$

Capital good firms A representative firms owns the capital stock and rents it to the intermediate good producers at rate r_t^k. The firm chooses investment I_t and capital tomorrow K_{t+1} subject to the capital law of motion. The Bellman equation is given by

$$J_t^K(K_t, I_{t-1}) = \max_{K_t, I_t} \left\{ r_t^k K_t - I_t + E_t \left[\frac{J_{t+1}^K(K_{t+1}, I_t)}{1 + r_t} \right] \right\} \tag{34}$$

subject to

$$K_{t+1} = (1 - \delta) K_t + \mu_t e^{v_t^\mu} \left[1 - S \left(\frac{I_t}{I_{t-1}} \right) \right] I_t \tag{35}$$

where $\delta \in (0, 1)$ is the depreciation rate of capital, and μ_t is the marginal efficient of investment as in Justiniano et al. (2011), and $S(\cdot)$ is a convex function that satisfies $S(1) = S'(1) = 0$ and $S''(1) = \psi_i$. We will set $S(x) = \frac{\psi_i}{2}(x-1)^2$. Defining Tobin's Q as the marginal value of capital at time t, $\frac{\partial_K J_{t+1}(K_t, I_t)}{1+r_t}$, the dynamics of investment are characterized by

$$Q_t = \frac{r_{t+1}^k + E_t[Q_{t+1}(1 - \delta)]}{1 + r_{t+1}} \tag{36}$$

Finally, the capital good firm makes profits, distributed as dividends

$$D_t^K = r_t^K K_{t-1} - I_t \tag{37}$$

5.1.3 Government

The fiscal authority issues one-period real bonds B_t, collects taxes T_t and conducts government spending G_t, bound by the government budget constraint

$$B_{t+1} + T_t = (1 + r_t)B_t + G_t \tag{38}$$

Given a particular income tax rate τ_t^y, tax revenues are the sum of labor and dividend taxes

$$T_t = \tau_t^y w_t N_t + \tau_{ss}^D D_t \tag{39}$$

The tax rate τ_t^y is chosen according to a rule that prevents large swings in the tax rate but ensures that the real government debt remains stationary.

$$T_t = \tau_{ss}^y w_t N_t + \tau_{ss}^D D_t + (1 - \rho^B)(B_t - B_{ss}) \tag{40}$$

Finally, we assume monetary policy follows a smoothed Taylor rule

$$i_t = \rho^r i_{t-1} + (1 - \rho^r)(\phi_\pi \pi_t + r_{ss}) + v_t^r \tag{41}$$

with monetary policy shock v_t^r, while the Fisher equation links the nominal interest rate to the real interest rate

$$(1 + r_t) = \frac{1 + i_{t-1}}{1 + \pi_t} \tag{42}$$

Aggregate shocks There are seven aggregate shocks: TFP shock, monetary policy shock, government spending shock, price markup shock, wage markup shock, labor incidence shock, dividend incidence shock.Each follow an independent AR(1) process.

$$\Theta_t = \rho_\Theta \Theta_{t-1} + \sigma_\Theta \epsilon_t^\Theta$$
$$v_t^r = \rho_r v_{t-1}^r + \sigma_r \epsilon_t^r$$
$$v_t^g = \rho_g v_{t-1}^g + \sigma_g \epsilon_t^g$$
$$v_t^p = \rho_p v_{t-1}^p + \sigma_p \epsilon_t^p$$
$$v_t^w = \rho_w v_{t-1}^w + \sigma_w \epsilon_t^w$$
$$v_t^\mu = \rho_\mu v_{t-1}^\mu + \sigma_\mu \epsilon_t^\mu$$

Equilibrium The aggregate resource constraint is

$$Y_t - \Psi_t^p - \Xi = C_t + G_t + I_t \tag{43}$$

The asset market clearing is

$$A_t = B_t \tag{44}$$

Labour market clearing requires

$$h_t = N_t \tag{45}$$

A *rational expectation equilibrium* consists of a sequence of policy functions $\{c_t, a_t, h_t\}$, a sequence of value functions $\{V_t\}$, a sequence of prices $\{w_t, r_t^k, \pi_t, \pi_t^w, \tau^y, r_t, i_t\}$, a sequence of aggregate objects $\{Y_t, C_t, K_t, N_t, \Psi_t^p, D_t, I_t, Q_t, B_t, T_t\}$, a sequence of distribution $\{F_t\}$, a sequence of exogenous states $\{\Theta_t, v_t^r, v_t^g, v_t^p, v_t^w, v_t^\mu\}$, and a sequence of beliefs over prices such that

1. Given the sequence of value functions, prices, and policy functions, the household Bellman equation holds.

2. Given the sequence of beliefs over prices, all agents optimize.

3. The evolution of the distribution is consistent with the policy.

4. The sequence of beliefs over prices is rational.

5. All markets clear.

5.2 Incidence functions

The crucial block of this model is given by labor incidence function $\Gamma_t^y(z)$, which encodes differences in sensitivity to aggregate income for different productivity levels.

In logs, we specify the incidence function by

$$\log \Gamma_t^y(z) \equiv \Gamma^y(z, Y_t, v_t^r) = \log z + Beta(F_z(z); \alpha^y, 1) \log \frac{w_t N_t}{w_{ss} N_{ss}} + Beta(F_z(z); \alpha^r, 1) v_t^r \tag{46}$$

We normalize the function such that both $zBeta(F_z(z); \alpha^y, 1)z$ and $\Gamma_t^y(z)$ integrate to one. The shape of the incidence is determined by two parameters, α^y and α^r. For example, $\alpha^y < 1$ determines a downward sloping $Beta(F_z(z); \alpha^y, 1)$ as z increases; and conversely for $\alpha^y > 1$. The case where $\alpha^y = 1$ implies equal incidence across all households. Figure 4 plots the labor incidence associated with different values of α. In steady state, $\Gamma_t(z) = z$, thus collapsing to the standard specification. When $w_t N_t$ is above its steady state value and $\alpha^y < 1$, earnings of low productivity households increase more than high productivity relative to the steady state level. To see this, differentiate (46) with respect to z to obtain equation (47). In steady state, the second term is equal to zero, implying that outside of steady state, earnings of low productivity household increase relatively more.

$$\frac{\partial \log \Gamma_t(z)}{\partial z} = \frac{1}{z} + \underbrace{\frac{Beta(F_z(z); \alpha^y, 1)}{\partial z}}_{<0} \log \left(\frac{w_t N_t}{w_{ss} N_{ss}} \right) \tag{47}$$

The same intuition carry over to the third term in (46), which describe the additional income incidence across the distribution related solely to a monetary policy shock. For $\alpha^r \neq 1$ this term implies differential sensitivities of aggregate income conditional on a monetary policy shock. One interpretation is that it encodes that idea that sectoral employment varies by productivity level and that these sectors have differential exposures to monetary policy shocks. XXX and XXX provide some empirical evidence for the presence of this channel.

In calibration these incidence functions, the previous literature has estimated micro regressions of the form

$$\log y_{it} = \alpha + \beta_i \log Y_t + e_{it} \qquad \forall i \tag{48}$$

In our model, this estimate is invalid owing to the contemporaneous relationship between wages, hours and monetary policy shocks. More generally, the conventional regression limits the possible specifications of the model, while the present approach, since it jointly estimates the incidence parameters, is allows for a much more general class of specifications. For example, might include additional shocks, such as TFP in this incidence function, representing the fact that skilled and unskilled labor earnings respond differently to TFP shocks.[30]

figures_v2/incidence_functions.png

Figure 4: Labor incidence functions for different parameterizations of $Beta(z; \alpha, \beta)$

5.3 Solution in sequence-space

We employ the sequence-space Jacobian (SSJ) method of Auclert et al. (2021b) to obtain a linearized solution of the model in Section 5.1. Although the original SSJ method is developed for computing the aggregate dynamics, it can be easily extended to obtain a solution for the *micro* consumption dynamics. Let $\mathbf{y}_t = (c_{1,t}, \dots, c_{M,t})^\top$ be the vector of cross-sectional consumption and $\boldsymbol{\epsilon}_t \in \mathbb{R}^S$ be the vector of fundamental shocks at time t. In our model, $\boldsymbol{\epsilon}_t$ consists of three shocks: TFP shock, monetary policy shock, and income dispersion shock. We have the following proposition.

Proposition 2. *In the linearized equilibrium, $\mathbf{c}_t$ has a moving average (MA) representation*

$$\mathbf{y}_t = \mathbf{y}_{ss} + \sum_{j=0}^{\infty} \Theta_j^y \boldsymbol{\epsilon}_{t-j} \tag{49}$$

Furthermore, the MA coefficient matrix Θ_j^y is given by

$$\Theta_j^y = \sum_{p \in \mathcal{P}} \mathcal{J}_p^y F^j \mathcal{I}_e^p$$

where $\mathcal{P}$ denotes the set of aggregate inputs that enter the household's problem and

- *$\mathcal{J}_p^y \in \mathbb{R}^{M \times \infty}$ is the cross-section of gradients of consumption wrt. the future path of aggregate input p*
- *$\mathcal{I}_e^p \in \mathbb{R}^{\infty \times S}$ is the impulse response functions of aggregate input p*
- *F is the shift-forward operator*

[30] From a more practical perspective, Alves et al. (2020) note that β_i estimates are sensitive to use of log versus other transformations. Our approach is again immune to this external regression considerations.

In light of Proposition 2, simulation of the micro consumption dynamics is straightforward, and computation of the MA coefficient matrices is trivial because $\mathcal{J}^c_p$ and $\mathcal{I}^p_e$ are products of the SSJ method.[31] To operationalize the algorithm, we truncate the horizon, to T. This implies that $\mathcal{J}^c_p \in \mathbb{R}^{M\times T}$, $\mathcal{I}^p_e \in \mathbb{R}^{T\times S}$. In practice, we truncate the horizon at $T = 300$.

5.4 Population VAR(1) coefficients

Because the SSJ solution method implies a (truncated) moving-average representation of the observables (both aggregate and cross-section), its open the door to a more precise implementation of our algorithm. The Consider the truncated (practical) version of the MA representation (49) where $T = 300$.

Given a vector of structural parameters, the population covariances are thus available in closed-form. To obtain the VAR(1) covariance matrix, we need the coincident and lagged covariance matrix. This is given by

$$\mathbb{E}[\mathbf{y}_t\,\mathbf{y}_t^\top] = \sum_{j=0}^{J} \Theta^y_j \Sigma \Theta^{y\top}_j \tag{50}$$

$$\mathbb{E}[\mathbf{y}_t\,\mathbf{y}_{t-1}^\top] = \sum_{j=0}^{J} \Theta^y_j \Sigma_{-1} \Theta^{y\top}_j \tag{51}$$

$$\tag{52}$$

where $\Sigma = diag(\sigma_1^2, \ldots, \sigma_S^2)$ and Σ_{-1} is the same matrix but shifted one below the off diagonal

$$\Sigma_{-1} = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ \sigma_1^2 & 0 & \ldots & 0 & 0 \\ 0 & \sigma_2^2 & \ldots & 0 & \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & \ldots & \sigma_S^2 & 0 \end{bmatrix} \tag{53}$$

Having these population covariances in closed form rids the need for simulating the model, allowing for a much faster and accurate likelihood computation. Similarly, we could use the same arguments to compute the two-lag autocovariance to compute the population reduced-rank VAR(2) as described in Section B.2.

[31]The gradients $\mathcal{J}^c_p$ are computed by backward iteration in the first step of the ”Fake news algorithm”.

5.5 Calibration

The model is in quarterly frequency. Table 2 reports the parameters that we calibrate. In addition, we internally calibrate β to clear the asset market, which has value 0.967. We estimate other 11 model parameters: $[\kappa_w, \kappa_p, \psi_p, \psi_i, \rho_i, \phi_\pi, \alpha^y, \alpha^{mp}]$, and the AR(1) persistence and innovation variance parameters of the 6 aggregate shocks above. In all, this gives us 23 parameters to estimate.

Parameter	Interpretation	Value	Justification
σ	CRRA	0.5	
ϕ^{-1}	Frisch elasticity	2.0	
α	Labor share	0.33	
τ_d^{ss}	Dividend tax rate	0.2	
Ξ	Interm. goods fixed cost	0.0	
δ	Capital depreciation rate	0.02	
ρ_b	Response of tax rate to debt	0.4	
ι_p	Inflation indexation	0.25	

Table 2: Calibrated parameters

5.6 Data

Our cross-sectional data is sourced from the Interview Survey section of the Consumer Expenditure Survey (CEX). The CEX provides detailed information about household consumption expenditures in the U.S. The Interview Survey is a rotating panel of households chosen to be represented of the U.S. population. Each household is interviewed for a maximum of four quarters. Expenditure is reported each quarter, and income is reported the first and last quarters, and wealth is reported only in the last quarter. We use CEX data between 1990 Q1 and 2023 Q1.

We construct our income data by defining labor income as wage/salary income plus self-employed income. We define assets in the model as closely as possible to liquid assets in the data. For each period, we normalize the asset holdings by average quarterly labor income. Prior to 2013, this is equal to the value in checking/brokerage accounts plus saving accounts in banks, credit unions etc. and loads and securities held in mutual funds.[32]. After 2013, the concept is conveniently aggregated into the total market value in checking, savings, money market accounts, CDs, etc. and other similar accounts[33]. Before progressing, we do some preliminary cleaning of the data: we drop households where the principal earner is below 25 or above 65 years old; those where incomes negative positive; those whose quarterly consumption are above two times their annual income.

[32]The CEX codes are respectively `CKBKACTX`, `SAVACCTX` and `SECESTX`

[33]The CEX code is `LIQUIDX`

Individual productivity. Households in our model have two idiosyncratic state-variables, productivity z_{it} and assets a_{it}. By the organization principle, we need to estimate the unobserved productivity of each individual at time t. Our approach follows from Floden and Lindé (2001), who estimate AR(1) processes for household productivity from U.S data. The identifying assumption is that permanent income differences can be captures by individuals' specific characteristics. The first step is to run a pooled Mincer-style regression. We regress log labor income on age, age-squared, plus dummies for gender, occupation and education level with time-fixed effects in the follow specification

$$\log y_{it} = \varphi_t + \varphi_1 + \varphi_2 AGE_{it} + \varphi_3 AGE_{it}^2 + \tag{54}$$

$$\boldsymbol{\varphi}_M MALE_{it} + \boldsymbol{\varphi}_E \mathbf{EDU}_{it} + \varphi_O \mathbf{OCC}_{it} + z_{it} \tag{55}$$

where $\mathbf{EDU} = [EDU_1, EDU_2, \ldots, EDU_8]^\top$ is a vector of education dummies and $\mathbf{OCC} = [OCC_1, OCC_2, \ldots, OCC_15]^\top$ is a vector of occupation dummies. Table 3 shows the results of the regression. The adjusted R^2 is around 0.2, inline with existing results for the U.S. from Floden and Lindé (2001).

Variable	Estimate	p-value
const	8.038	0.00
AGE	0.081	0.00
AGE^2(/100)	-0.089	0.00
MALE	0.124	0.00
EDUCA	0.024	0.00
R^2	0.24	
TN	341,785	

Table 3: Labor income regression results on U.S data

The persistent component of log-productivity analogous to the model object z_{it} is defined as the residuals of the above regression.[34]

$$z_{it} = \log y_{it} - \widehat{\log y}_{it} \tag{56}$$

The above algorithm allows us to infer the individual-level productivity for each quarter. For each period, we then place households into 9 productivity bins, representing the quantiles of

[34] Appendix Figure E.2 compares the productivity distribution in the data with the unconditional distribution in the model implied by our calibration.

productivity.[35]

We similarly split the asset space into 9 equally spaced bins, ranging from 0 to $200,000. Taking the cartesian product with the productivity bins gives us 81 bins for which we have observables. Our choices imply that the bins are kept fixed across time periods.

We then place individuals into bins depending on their associated state variables and compute the mean consumption in each bin. Naturally in some periods, there are no individuals in some areas of the state space, so we interpolate between the missing data. Importantly, note that we interpolate *within* a time-period, not *across* time-periods. Figure E.3 in the Appendix visualizes the extent of the missing data. The result of these operations is a quarterly series of consumption for individuals in each point of the discretized state-space implied by the model in Section 5.1.

One interesting representation of our data is to notice that each column is an estimate of the consumption policy function in a given period. Figure 5 plots the time-average (or "steady-state") consumption policy functions as a ratio to average labor income. Interestingly, consumption appears to share some familiar characteristics to those in our economic models. They are increasing in productivity and assets, and are highly non-linear at low asset levels and linear at high asset levels.[36]

! figures_v2/average_consumption_policy_function.pn

Figure 5: Consumption policy function inferred from CEX data

Going further, Figure 6 plots the policy function for low, middle and high productivity individuals at different points in time. The green is the average policy function in 2004, considered as the peak of the "Great Moderation period" with low inflation and high aggregate growth. The red lines show the average policy function in 2021, which covers the COVID-19 recession. There are discernible differences in the policy function at all productivity levels. In all cases, the level of consumption was lower in 2021 relative to 2004. During 2021 the consumption consumption falls for high productivity individuals with large liquid asset holdings. Explain more.

They key message here is that the policy function change as the aggregate states in the economy change which forms the bedrock of our identification strategy.

[35] Connect to calibration of productivity process

[36] Arellano et al. (2017) infer more general consumption and income functions from PSID data that are non-parametric functions of many covariates. We do not take that approach here since our goal is estimation of the structural model in Section 5.1

figures_v2/state_dependent_consumption_policy_func

Figure 6: State-dependent consumption functions inferred from CEX

5.7 Estimation

We estimate the remaining structural parameters using an off-the-shelf Bayesian approach as described in An and Schorfheide (2007) and Fernández-Villaverde et al. (2016). First, we perform an extensive mode-finding algorithm, by initializing a bounded `L-BFGS-B` algorithm at 10 random points in the permissible state space. We then take the parameter vector associated with the highest posterior value and perform an additional optimization using the `Nelder-Mead` method. For the `Micro` estimation, we employ our method to efficiently approximate the likelihood; and for the `Agg` estimation, we use the standard likelihood representation associated with the moving-average representation as described by Auclert et al. (2021a).

We then approximate the posterior distribution of the parameters using a random-walk Metropolis Hastings algorithm with tuned proposal covariance matrix and adaptive step size (Atchadé and Rosenthal (2005)). We generate 300,000 draws and discard the first 50,000. Though we employ a simple MCMC algorithm here, our method doesn't preclude more sophisticated techniques, such as the Hamiltonian Monte Carlo (Hoffman et al. (2014)) or Sequential Monte Carlo (Herbst and Schorfheide (2014))

5.8 Results

Table 4 presents the posterior mean and standard deviation of the structural parameters from both `Micro` and `Agg` estimations. `Micro` estimates imply a flatter slope of the price Phillips curve and the wage Phillips curve than `Agg`. implying a higher degree of monetary neutrality. Moreover, estimates also imply a more aggressive and longer lasting response of monetary policy via larger Taylor coefficients ϕ_π and ρ_i.

`Micro` estimates of the incidence parameters are both less than one, implying that low income households are significantly more sensitive to fluctuations in aggregate income relative to high income households. `Agg` estimates of the cross-section parameters imply the similarly qualitative exposures, although they are much less precisely estimated than their `Micro` counterparts; and the mean magnitudes are much smaller.

In almost all cases, the `Micro` estimates of standard deviation of the shocks are larger in magnitude than `Agg`. For example the wage markup shock is almost 10 times larger, and the efficiency of investment shock is almost 3 times as large in the `Micro` estimates compared to `Agg`.

		Prior			Posterior (Micro)		Posterior (Agg)	
Parameter	Interpretation	Distribution	Mean	Std	Mean	Std	Mean	Std
κ_p	Phillips curve slope	Gamma	0.05	0.03	0.121	0.018	0.063	0.006
κ_w	Wage Phillips curve slope	Gamma	0.05	0.03	0.165	0.042	0.006	0.004
ϕ_π	Taylor Coefficient	Normal	1.5	0.3	2.228	0.056	1.248	0.023
ϕ_T	Tax smoothing	Beta	0.5	0.2	0.06	0.011	0.777	0.016
ρ_i	Taylor persistence	Beta	0.5	0.2	0.875	0.017	0.235	0.011
α_y	Aggregate income incidence	LogNormal	1.0	0.5	0.078	0.012	1.557	0.007
α_{mp}	Monetary policy incidence	LogNormal	1.0	0.5	0.418	0.083	0.813	0.018
ρ_Z	Persistence TFP shock	Beta	0.5	0.1	0.473	0.046	0.792	0.005
ρ_{ε^r}	Persistence Monetary policy shock	Beta	0.5	0.1	0.373	0.063	0.385	0.038
ρ_{ε^G}	Persistence Government spending shock	Beta	0.5	0.1	0.429	0.03	0.57	0.019
ρ_{ε^w}	Persistence Wage markup shock	Beta	0.5	0.1	0.197	0.047	0.398	0.009
ρ_{ε^π}	Persistence Price markup shock	Beta	0.5	0.1	0.205	0.038	0.353	0.021
ρ_{ε^μ}	Persistence MEI shock	Beta	0.5	0.1	0.995	0.002	0.616	0.025
σ_Z	Std TFP shock	Inverse Gamma	1.0	0.2	0.509	0.066	0.206	0.056
σ_{ε^r}	Std Monetary policy shock	Inverse Gamma	1.0	0.2	0.076	0.011	0.186	0.024
σ_{ε^G}	Std Government spending shock	Inverse Gamma	1.0	0.2	0.259	0.028	0.206	0.032
σ_{ε^w}	Std wage markup shock	Inverse Gamma	1.0	0.2	1.095	0.096	0.195	0.034
σ_{ε^π}	Std price markup shock	Inverse Gamma	1.0	0.2	0.201	0.019	0.089	0.014
σ_{ε^μ}	Std MEI shock	Inverse Gamma	1.0	0.2	0.419	0.085	0.16	0.029
σ_Y	Output meas. error	Inverse Gamma	1.0	0.2	0.38	5.457	0.419	0.028
σ_C	Consumption meas. error	Inverse Gamma	1.0	0.2	0.007	0.649	0.432	0.039
σ_i	Interest Rate meas. error	Inverse Gamma	1.0	0.2	19.244	3.111	0.544	0.049
σ_π	Inflation meas. error	Inverse Gamma	1.0	0.2	0.85	0.703	0.138	0.019
σ_w	Wage inflation meas. error	Inverse Gamma	1.0	0.2	0.858	0.6	0.834	0.054
σ_I	Investment Growth meas. error	Inverse Gamma	1.0	0.2	0.786	0.616	0.332	0.096
σ_c	Micro consumption meas. error	Inverse Gamma	3.0	0.4	4.283	0.037	-	-
σ_y	Micro income meas. error	Inverse Gamma	3.0	0.4	0.93	0.021	-	-

Table 4: Prior and posterior parameter estimates. NOTE: All standard deviations are presented $\times 100$

Figure 7 plots the percentage change in idiosyncratic income associated with a one percent increase in aggregate income (left) and negative 25bps (expansionary) monetary policy shock (right) *relative to the equal exposure benchmark*. The swathes denote the 90% credible set, while the solid lines denote the mean.

In the left panel, `Micro` imply that incomes for the lowest decile increase by 2% more than the equal exposure benchmark. This is compared to only 0.25% implied by `Agg` estimates. Moreover, the slope is much steeper in the `Micro` estimates, suggesting a much larger degree of redistribution compared to `Agg` estimates. As we will see below in Section XXX, these differences have important implications for the amplification of aggregate shocks, since the low percentile households are the ones that also have a high MPC.[37] The larger parameter uncertainty of `Agg` estimates in Table 4 translate into much wider credible sets compared to `Micro`.

The right panel shows the percentage change in individual income following a 1% monetary policy shock, relative to the equal exposure benchmark. The mean estimates of `Agg` and `Micro` also imply opposite signs for the relative elasticities. The `Micro` estimate of $\alpha_{mp} > 1$ implies that high productivity households are more sensitive to monetary policy shocks than low productivity households. This is seen here by a fall in income following a contractionary monetary policy shock and a relative rise for low productivity individuals. The mean estimates of `Agg` imply the opposite: a contractionary monetary policy shock *increases* inequality response of inequality following a monetary policy shock, although the parameter is so imprecisely estimated that the credible set crosses the zero.

That high income households are more sensitive to monetary policy shocks has some empirical standing. By uses individual and sectoral data from the Current Population Survey (CPS), Dolado et al. (2021) document that a monetary expansion is associated with an increase in the skill premium by 0.4pp and relative employment of skilled versus unskilled by 0.35pp. This result implies that incomes of the highly skilled are more sensitive to changes in monetary policy than those of the low skilled. They rationalize these findings through a structural model in which high-skilled labor and capital are complements, while low-skilled labor and capital are substitutes. Thus, the authors conclude that a monetary policy shock raises labor income inequality, consistent with the `Micro` estimates. Contrary to our estimates Coibion et al. (2017) use identified monetary policy shocks of Romer and Romer (2004) and find that contractionary monetary policy shock increase labor income inequality.

[37] See Patterson (2023) for an empirical investigation into this covariance.

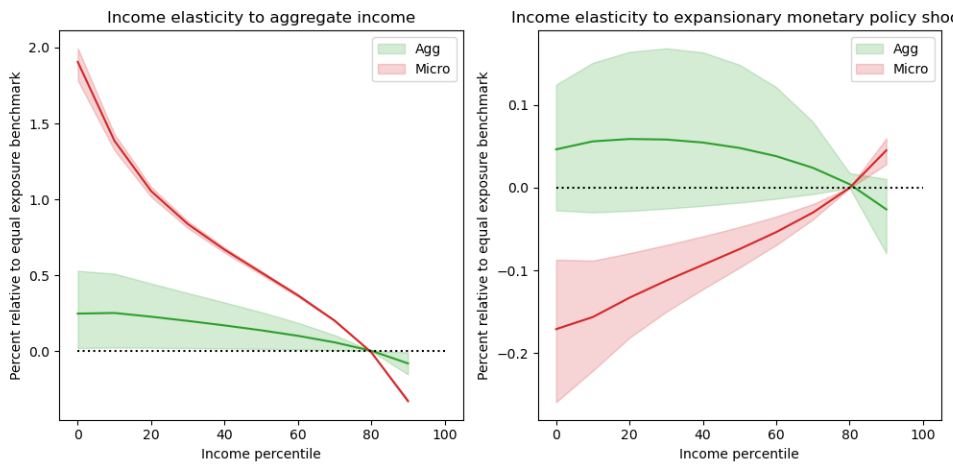

Figure 7: Estimated income exposures by percentile

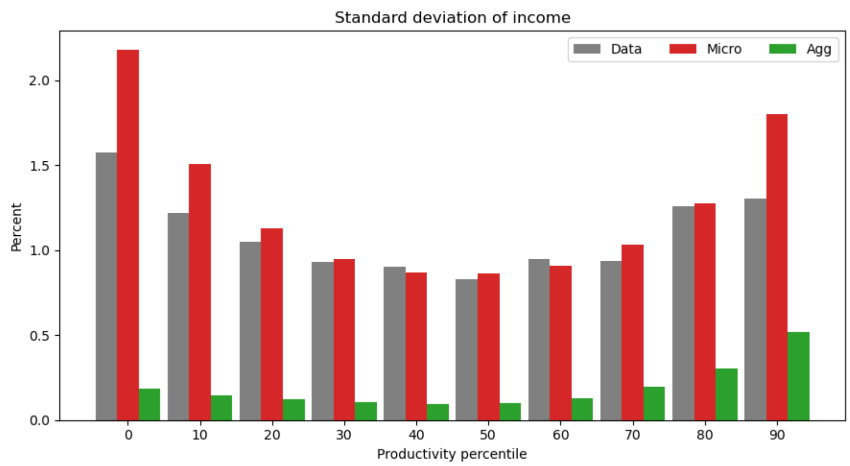

Figure 8: Idiosyncratic income volatility

We compare our parameter estimates by studying their ability to match second moments of the data. Figure 8 plots the standard deviation of idiosyncratic income for each decile of the productivity distribution associated with the posterior modes of both estimates. We are able compute the idiosyncratic income volatilities in the model analytically, since the sequence-space jacobian method obtains an MA representation of the solution. The income volatilities are U-shaped in the data, with low and high productivity households

Figure 9: Impulse response to 25bps monetary policy shock

	Standard deviation (%)					
	Output	Consumption	Investment	Price inflation	Wage inflation	interest rate
Data	0.60	0.57	1.87	0.22	0.89	0.69
Agg						
Micro	0.59	0.72	3.43	0.36	1.00	0.36

Table 5: Standard deviation of aggregate variables

Figure 10: Output impulse response decomposition

Figure 11: Forecast Error Variance Decomposition

Table 6: Caption

6 Concluding remarks

We develop an indirect inference method for estimating HA business-cycle models using micro data. The key idea is to approximate the data-generating process with a low-dimensional dynamic factor model and use the implied likelihood for inference. Employing the Dynamic Mode Decomposition algorithm, the likelihood evaluation is fast and simple. Moreover, our estimation procedure can seamlessly accommodate the sequence-space solution method, while most currently available estimation methods (e.g. Liu and Plagborg-Møller 2023) are designed for state-space solution only.

Our method is based on two assumptions: 1) the HA model is well-approximated by a low-dimensional dynamic factor model, and 2) the cross-sectional dimension of the micro data is large. We show that the first assumption holds in a wide range of HA models and provide a theoretical justification for it. In our simulation study, we show that our method works well on a realistic dataset, verifying the empirical relevance of the second assumption.

Comparing with other conventional methods including time-series estimation with cross-sectional moments and frequency-domain estimation, our method delivers a better estimate both in terms of bias and standard error because of the more efficient use of cross-sectional information. As our method is based on approximated likelihood, we show that it can be easily pair with Bayesian methods to conduct Bayesian indirect inference.

We conclude with two directions for future research. First, a method for filtering the aggregate shocks using micro data remains to be developed. The sequence-space filtering method in McKay and Wieland (2021) seems promising. Second, it would be intriguing to estimate a calibrated HANK model using acutual micro data (e.g. CEX) and contrast the results with those derived from aggregate data. These disparities will offer new insights into model misspecification and aid in refining our modeling choices. We leave these to future works.

7 Conclusion

We demonstrate the power of our approach on simulated data generated by the structural model in Section 5.1. In addition to the calibrated parameters in Table 2, Table 7 lists the parameters that generate the simulated data, and that we aim to recover.

To replicate a realistic dataset, our simulated data is 120 periods, representing 30-years' worth of quarterly data. We randomly sample 300 individuals states from the stationary distribution and add i.i.d measurement error to the data. We fix the variance of the measure error such that it

Parameter	Interpretation	Value	Justification
Model parameters			
κ_w	Slope of wage Phillips Curve	0.1	
κ_p	Slope of price Phillips Curve	0.075	
ψ_p	Interm. goods adjustment cost	0.5	
ι_p	Interm. goods adjust. cost indexing	0.25	
ψ_i	Capital goods adjustment cost	1.4	
ρ_i	Taylor rule lag	0.80	
ϕ_π	Taylor rule inflation coef.	1.25	
β_a^y	Labor incidence Beta fn. parameter (α)	0.5	
β_b^y	Labor incidence Beta fn. parameter (β)	1.0	
β_a^D	Dividend incidence Beta fn. parameter (α)	1.0	
β_b^D	Dividend incidence Beta fn. parameter (β)	1.0	
Shock process parameters			
ρ_Θ	TFP shock AR(1)	0.9	
ρ_r	Monetary policy AR(1)	0.95	
ρ_p	Price markup AR(1)	0.8	
ρ_w	Wage markup AR(1)	0.75	
ρ_g	Govt. spending AR(1)	0.82	
ρ_μ	Investment marginal efficiency AR(1)	0.7	
σ_Θ	TFP shock std	0.015	
σ_r	Monetary policy shock std	0.025	
σ_p	Price markup shock std	0.02	
σ_w	Wage markup shock std	0.015	
σ_g	Govt. spending shock std	0.01	
σ_μ	Investment marginal efficiency shock std	0.025	

Table 7: Targeted parameters for simulation exercise

accounts for 20% of the total variation. We estimate the standard errors along with the parameters of interest. On the aggregate data, we simulate aggregate output, inflation, consumption, wage inflation, price inflation, and investment. We also add measurement error such that it accounts for 10% of the total variation in each series and estimate the ”relative size” parameter as well.

Number of factors To choose the number of factors in our auxiliary model, we simulate time-series of consumption for $M = 300$ households, each of length $T = 10,000$, as well as aggregate output, consumption, nominal interest rate, price inflation, wage inflation and investment. We demean the time-series for each household and stack them vertically to create the simulated data matrix $\mathbf{Y}^{sim}$. We implement the checks outlined in Section 3.4 to infer an appropriate r. Figure 12 shows the 10 largest singular values from the data matrix $\mathbf{Y}^{sim}$. There appears to be 6 dominant singular values, with the rest almost zero.

Table 8 computes the R^2 statistic in equation (16) and the information criterion of (13) for an figures_v2/singular_values_scaled_simulated.png

Figure 12: Singular values of $\mathbf{Y}^{sim}$

increasing r. The first row of the table shows that the R^2 doesn't increase in r after $r = 3$. It shows that the R^2 statistic increases until around $r = 13$ and then plateaus. The maximum value of the R^2 is around 0.5, which isn't surprising for data generated i.i.d measurement error.

The information criterion, which penalizes large r, is shown in the second row, which falls until $r = 11$ and then plateaus. Finally, we study the residuals associated with the rank-reduced first-order VAR. The third row computes the maximum absolute autocovariance of the VAR residuals $\mathbf{a}_t$, computed via (15). We find shows significant autcorrelation for $r = 1, 2, dots 11$, suggesting they are inadequate in satisfying the assumptions that define a first-order VAR. For $r \geq 12$, the maximum absolute autocovariance is close to zero, 0.13. Taking the results of the checks holistically, we fix $r = 15$.

Number of factors, N	1	2	...	10	11	12	13	14	15
$R^2(r)$	0.313	0.361	...	0.473	0.490	0.495	0.496	0.496	0.496
$IC(r)$	-4562.1	-4573.8	...	-4601.9	-4602.1	-4602.1	-4602.1	-4602.1	-4602.1
max $\lvert\hat{E}[a_{t+1}, a_t]\rvert$	37.27	16.08	...	2.97	1.29	0.05	0.13	0.13	0.13

Table 8: R^2 and IC for an increasing N

7.1 Simulation Results

Table XXX shows the

7.2 Comparison with estimation using aggregate data

Aggregate data only For comparison, we estimate the model parameters using aggregate data in two ways. The first is implemented using the MLE procedure by Auclert et al. (2021b) with only aggregate data (which we label `Agg`). The aggregate data consists of output, inflation, and nominal rate series, each of length $T = 120$. For a fair comparison with our method, we add measurement errors to the aggregate data which accounts for 10% of the total variation.[38] The result of this estimation is presented in the `Agg` columns of Table **??**. For most parameters, except the size of

[38]Since the Auclert et al. (2021b) method computes the likelihood of the aggregate data exactly, without measurement errors, their method will definitely deliver a better estimate than ours. Aruoba et al. (2016) argues that measurement error accounts for 20% of the variation in official US GDP measures. Thus, we view the 10% measurement error as a useful benchmark.

MP shock and income dispersion shock, the `Agg` performs worse than `MicroDMD` both in terms of bias and standard error. Figure **??** shows the distribution of the estimates. It shows that that the finite-sample distribution is much more dispersed and ill-behaved than our method.

Aggregate data plus cross-sectional moments Next, we compare our method against the population approach of including micro data into the estimation by constructing time-series for a few cross-sectional moments (e.g. Bayer et al. 2024 and Mongey and Williams 2017). We label this approach `Agg+`. The main advantage of this method is its simplicity and speed, since the likelihood of aggregate time-series can be efficiently computed via Kalman filter or using the full variance-covariance matrix in the same way as `Agg`. However, the *ex-ante* static aggregation of micro data may induce unnecessary information loss. In principle, `MicroDMD` makes better use of the micro data by utilizing the DMD algorithm to extract the most informative dynamic structures underlying the micro data.

To illustrate this point, we append a cross-sectional moment, the variance of log consumption, to the macro data and redo the aggregate estimation exercise. The results are reported in the `Agg+` columnds of Table **??**. The inclusion of cross-sectional moment brings the mean estimate closer to the truth and substantially lowers the standard error (compared to `Agg`), most notably for the estimates of the slope of the wage NKPC and the TFP shock process. Nevertheless, the performance of the estimator is still significantly worse than our method, suggesting that our method retains cross-sectional information beyond simple moments.

Do these these differences in the estimated parameters translate to meaningful differences in the objects that macro-economists care about? Figure **??** suggest that they do. It plots the impulse responses of aggregate output to the three shocks in the model: TFP, monetary policy and the income dispersion shock. For all three shocks the width of the confidence bands for `MicroDMD` is significantly reduced compared to both the `Agg` and `Agg+`.

Overall, the results suggest that our method which exploits the rich information contained in the micro data is better able to recover the true parameters of the model, and generally with a lower standard error in finite samples. The results emphasizes the advantage of using micro data in the estimation of heterogeneous-agent models.

7.3 Comparison with other estimation methods using micro data

Frequency-domain estimation The difficulty of exact likelihood evaluation of the micro data is the high dimensionality $(M \times T)$ of the variance-covariance matrix. One way to tackle this computational challenge is to evaluate the likelihood in the frequency-domain using the Whittle approximation, as in Hansen and Sargent (1981), Christiano and Vigfusson (2003), and Plagborg-Møller (2019). We label this method `MicroFD`. The Whittle approximation decomposes the entire $M \times T$-dimensional variance-covariance matrix into the sum of M-dimensional frequency-specific matrices. Thanks to the Fast Fourier Transform, the decomposition and associated likelihood evaluation is fast and only requires the sequence-space solution of the model. A key difference between the frequency-domain estimation method and ours is that it requires large T for accurate approximation, while our method requires large M. Since in reality the cross-sectional dimension of the micro dataset is usually larger than the time dimension, we argue that our method is more suitable in practice. We apply the frequency-domain estimation method to the simulated micro datasets and report the results in the `MicroFD` columns of Table **??**.[39] The results suggest that our method dominates the frequency-domain estimation method both in terms of bias and standard error, consistent with the asymptotic theory.

Full-information estimation Liu and Plagborg-Møller (2023) develops a full-information likelihood based approach for the estimation of heterogeneous-agent models. Our method is different in two dimensions. First, their method requires both macro and micro data. The macro data is used to infer the conditional distribution of the aggregate states which pin down the (conditional) likelihood of micro data. In contrast, our method can do inference base solely on micro data and can still be intuitively extended to incorporate macro data, a point that we further discuss in Section 8. Second, to infer the aggregate states, their method requires a state-space solution of the model computed using dimension-reduction algorithms such as Winberry (2018). On the other hand, since our indirect inference strategy is simulation-based, we only require the ability to simulate from the model and hence can accommodate sequence-space solutions as well as state-space solutions of the model. That being said, when Liu and Plagborg-Møller (2023)'s method is applicable, the full-information nature guarantees that their estimator is more efficient.

To summarize, our method provides a practical middle ground between the benchmark full-information method and conventional methods – it enjoys the substantive efficiency gain from

[39] The details of the estimation procedure can be found in Appendix C.1.

using micro data but is no harder to apply – it can be coded up in only a few lines of code – than aggregate-data-based methods.

8 Robustness and extension

In this section, we provide additional analytical and simulation results on the approximation quality of the low-dimensional dynamic factor model and discuss multiple extensions of our method.

8.1 Possibility of low-rank approximation

The validity of our method relies on the assumption that the heterogeneous-agent model generating the micro data can be approximated by a low-dimensional DFM. In Section 3, we argue that this is a reasonable assumption for a wide range of models and suggest heuristic procedures for testing this assumption using data simulated from the model. While these are good enough for practitioners, there is still no theoretical guarantee that a low-rank approximation is possible. Here we fill this gap by deriving a sufficient condition on the sequence-space solution of the model that renders a low-dimensional DFM representation.

Proposition 3. *Let $\mathcal{P}$ be the set of endogenous aggregate inputs (e.g. real wages), $\mathcal{E}$ be the set of exogenous shock processes (e.g. TFP), and $\mathcal{J} := \{\mathcal{J}^p_x : p \in \mathcal{P}, x \in \mathcal{E}\}$ be the set of general equilibrium Jacobians. Suppose all the exogenous shock processes are AR(1). If for any $x \in \mathcal{E}$ and $p \in \mathcal{P}$, we have the commutability condition,*

$$F\mathcal{J}^p_x = \mathcal{J}^p_x F,$$

where F is the shift-forward operator, then a low-dimensional DFM representation exists.

Intuitively, $F\mathcal{J}^p_x$ is the effect of the shock on the economy next period, while $\mathcal{J}^p_x F$ is the effect of a news shock on the economy today. The two effects will coincide if the HA distribution doesn't move in response to shocks, as this is the only endogenous state variable. Although the commutability condition will not hold exactly for most models, the slackness of the condition serves as a lower bound for the low-rank approximation quality.

We evaluate the normalized slackness $\|F\mathcal{J}^p_x - \mathcal{J}^p_x F\| / \|F\mathcal{J}^p_x\|$ for each shock and input in our small-scale HANK model and report the results in Table 9. There are three endogenous "prices" that the households care about – real interest rate, average real after-tax labor income, and average hours. Overall, the slackness is about 10% of the Frobenius norm of the GE Jacobian, except for average hours wrt. TFP shock which amounts to 23.8%. Redo with new model

Table 9: Slackness of the commutability condition $F\mathcal{J}_x^p = \mathcal{J}_x^p F$

	TFP	MP	Income disp
r_t	0.080	0.094	0.099
$(1-\tau_t)w_t n_t$	0.082	0.101	0.099
n_t	0.238	0.080	0.054

NOTE. The slackness is computed as $\|F\mathcal{J}_x^p - \mathcal{J}_x^p F\|/\|F\mathcal{J}_x^p\|$

Appendix A Proofs

A.1 Proof of Proposition 1

Proof. Associated with state-space system (1) is its innovations representation.[40]

$$\begin{aligned} \widehat{\mathbf{x}}_{t+1} &= \mathbf{A}\ \widehat{\mathbf{x}}_t + \mathbf{K}\,\mathbf{a}_t \\ \mathbf{y}_t &= \mathbf{G}\ \widehat{\mathbf{x}}_t + \mathbf{a}_t \end{aligned} \tag{A.1}$$

where $\widehat{\mathbf{x}}_t = \mathbb{E}[\mathbf{x}_t \,|\, \mathbf{y}^{t-1}]$, $\mathbf{a}_t = \mathbf{y}_t - \mathbb{E}[\mathbf{y}_t \,|\, \mathbf{y}^{t-1}]$, $\mathbf{a}_t \perp \mathbf{a}_s \,\forall t \neq s$ for $\mathbf{y}^t = \{\mathbf{y}_s\}_{s<t}$ and the Hilbert space $\mathcal{H}(\mathbf{a}^t) = \mathcal{H}(\mathbf{y}^t)$. Furthermore, $\mathbf{\Omega} \equiv \mathbb{E}[\mathbf{a}_t\,\mathbf{a}_t^\top] = \mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top + \mathbf{R}$, where $\mathbf{\Sigma}_\infty$ and $\mathbf{K}$ satisfy

$$\begin{aligned} \mathbf{\Sigma}_\infty &= \mathbb{E}[\mathbf{x}_t - \widehat{\mathbf{x}}_t][\mathbf{x}_t - \widehat{\mathbf{x}}_t]^\top \\ &= \mathbf{C}\,\mathbf{C}^\top + \mathbf{K}\,\mathbf{R}\,\mathbf{K}^\top + (\mathbf{A} - \mathbf{K}\,\mathbf{G})\mathbf{\Sigma}_\infty(\mathbf{A} - \mathbf{K}\,\mathbf{G})^\top \\ \mathbf{K} &= \mathbf{A}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top(\mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top + \mathbf{R})^{-1}. \end{aligned}$$

Notice that $rank(\mathbf{K}) = N$. Rearranging (A.1) gives an expression for $\mathbf{x}_{t+1}$ in terms of $\mathbf{y}_t$ and $\mathbf{x}_t$

$$\begin{aligned} \widehat{\mathbf{x}}_{t+1} &= \mathbf{A}\,\widehat{\mathbf{x}}_t + \mathbf{K}(\mathbf{y}_t - \mathbf{G}\,\widehat{\mathbf{x}}_t) \\ &= (\mathbf{A} - \mathbf{K}\,\mathbf{G})\widehat{\mathbf{x}}_t + \mathbf{K}\,\mathbf{y}_t \end{aligned}$$

Substituting into the measurement equation of (A.1) gives

$$\begin{aligned} \mathbf{y}_t &= \mathbf{G}[(\mathbf{A} - \mathbf{K}\,\mathbf{G})\widehat{\mathbf{x}}_{t-1} + \mathbf{K}\,\mathbf{y}_{t-1}] + \mathbf{a}_t \\ &= \mathbf{G}\,\mathbf{K}\,\mathbf{y}_{t-1} + \mathbf{G}(\mathbf{A} - \mathbf{K}\,\mathbf{G})\widehat{\mathbf{x}}_{t-1} + \mathbf{a}_t \end{aligned}$$

Notice that $\mathbf{B}_1^\infty := \mathbf{G}\,\mathbf{K}$ is a rank N matrix. Moreover, so it $\mathbf{A} - \mathbf{K}\,\mathbf{G}$. Iterating backward gives us

[40] A detailed derivation can be found in Lungqvist and Sargent (2018), Ch. 2

the desired result

$$\mathbf{y}_t = \sum_{j=1}^{\infty} \mathbf{B}_j^{\infty}\, \mathbf{y}_{t-j} + \mathbf{a}_t \tag{A.2}$$

$$\mathbb{E}[\mathbf{a}_t\, \mathbf{y}_{t-j}^{\top}] = \mathbf{0} \quad \text{for all } j \geq 1$$

$$\mathbb{E}[\mathbf{a}_t\, \mathbf{a}_t^{T}] = \mathbf{\Omega} = \mathbf{G}\, \mathbf{\Sigma}_{\infty}\, \mathbf{G}^{\top} + \mathbf{R}$$

$$\mathbf{B}_j^{\infty} = \mathbf{G}(\mathbf{A} - \mathbf{K}\, \mathbf{G})^{j-1}\, \mathbf{K} \quad \forall j \geq 1 \tag{A.3}$$

where $\text{rank}(\mathbf{B}_j^{\infty}) = N \quad \forall j \geq 1$. □

A.2 Proof of Lemma 1

Proof. Consider a sequence of models $\{\mathcal{M}_M\}$ indexed by the number of observables $M \in \mathbb{N}$. For each M, the model $\mathcal{M}_M$ is given by

$$\mathbf{x}_{t+1} = \mathbf{A}\, \mathbf{x}_t + \mathbf{C}\, \mathbf{w}_{t+1}$$
$$\mathbf{y}_t = \mathbf{G}_M\, \mathbf{x}_t + \mathbf{v}_t,$$

where shocks $\mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N\times N})$, measurement errors $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_M)$ and $\mathbf{w}_s \perp \mathbf{v}_\tau$ for all s, τ. Note that the matrices $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{N\times N}$ are fixed across M, meaning that the transition equation of the unobserved state is invariant to the number of observables. In the following, $\|\cdot\|$ denotes the Frobenius norm.

By Proposition 1, we have

$$\mathbf{K}_M = \mathbf{A}\, \mathbf{\Sigma}_{\infty,M}\, \mathbf{G}_M^{\top}(\mathbf{G}_M\, \mathbf{\Sigma}_{\infty,M}\, \mathbf{G}_M^{\top} + \mathbf{R}_M)^{-1}$$

where $\mathbf{\Sigma}_{\infty,M} \in \text{GL}(N, \mathbb{R})$ solves the matrix Ricatti equation

$$\mathbf{\Sigma}_{\infty,M} = \mathbf{C}\, \mathbf{C}^{\top} + \mathbf{K}_M\, \mathbf{R}_M\, \mathbf{K}_M^{\top} + (\mathbf{A} - \mathbf{K}_M\, \mathbf{G}_M)\mathbf{\Sigma}_{\infty,M}(\mathbf{A} - \mathbf{K}_M\, \mathbf{G}_M)^{\top} \tag{A.4}$$

$$= \mathbf{A}\, \mathbf{\Sigma}_{\infty,M}\, \mathbf{A}^{\top} + \mathbf{C}\, \mathbf{C}^{\top} - \mathbf{A}\, \mathbf{\Sigma}_{\infty,M}\, \mathbf{G}^{\top}(\mathbf{G}_M\, \mathbf{\Sigma}_{\infty,M}\, \mathbf{G}_M^{\top} + \mathbf{R}_M)^{-1}\, \mathbf{G}_M\, \mathbf{\Sigma}_{\infty,M}\, \mathbf{A}^{\top} \tag{A.5}$$

Such an invertible solution always exists and is unique under the maintained stability assumption. Furthermore, by construction, $\mathbf{\Sigma}_{\infty,M} = \mathbb{E}[\mathbf{x}_t - \mathbb{E}[\mathbf{x}_t \mid \mathbf{y}^{t-1}]][\mathbf{x}_t - \mathbb{E}[\mathbf{x}_t \mid \mathbf{y}^{t-1}]]^{\top}$.

We will first show that $\mathbf{\Sigma}_{\infty,M} \to \mathbf{C}\, \mathbf{C}^{\top} = \mathbb{E}[\mathbf{x}_t - \mathbb{E}[\mathbf{x}_t \mid \mathbf{x}_{t-1}]][\mathbf{x}_t - \mathbb{E}[\mathbf{x}_t \mid \mathbf{x}_{t-1}]]^{\top}$. By Assumption **??**, $rank(\mathbf{G}_M) = N$, so we have the normal equation

$$\mathbf{x}_t = (\mathbf{G}_M^{\top}\, \mathbf{G}_M)^{-1}\, \mathbf{G}_M^{\top}\, \mathbf{y}_t - (\mathbf{G}_M^{\top}\, \mathbf{G}_M)^{-1}\, \mathbf{G}_M^{\top}\, \mathbf{v}_t$$

Combine with the state transition equation and we obtain

$$\mathbf{x}_t = \mathbf{A}(\mathbf{G}_M^{\top}\, \mathbf{G}_M)^{-1}\, \mathbf{G}_M^{\top}\, \mathbf{y}_{t-1} - \mathbf{A}(\mathbf{G}_M^{\top}\, \mathbf{G}_M)^{-1}\, \mathbf{G}_M^{\top}\, \mathbf{v}_{t-1} + \mathbf{C}\, \mathbf{w}_t$$

Put $F[\mathbf{x}_t \mid \mathbf{y}_{t-1}] := A(\mathbf{G}_M^\top \mathbf{G}_M)^{-1} \mathbf{G}_M^\top \mathbf{y}_{t-1}$. The forecast variance of this linear predictor is

$$\begin{aligned}\mathbb{E}[\mathbf{x}_t - F[\mathbf{x}_t \mid \mathbf{y}_{t-1}]][\mathbf{x}_t - F[\mathbf{x}_t \mid \mathbf{y}_{t-1}]]^\top &= \mathbf{A}(\mathbf{G}_M^\top \mathbf{G}_M)^{-1} \mathbf{G}_M^\top \mathbf{R}_M \mathbf{G}_M (\mathbf{G}_M^\top \mathbf{G}_M)^{-1} \mathbf{A}^\top + \mathbf{C}\,\mathbf{C}^\top \\ &= \sigma_v^2 \mathbf{A}(\mathbf{G}_M^\top \mathbf{G}_M)^{-1} \mathbf{A}^\top + \mathbf{C}\,\mathbf{C}^\top\end{aligned}$$

where we have use the assumption that $\mathbf{R}_M = \sigma_v^2 \mathbf{I}_M$. Now, by Assumption **??**, as $M \to \infty$, $\left\|(\mathbf{G}_M^\top \mathbf{G}_M)^{-1}\right\| \to 0$. Therefore, we have

$$\begin{aligned}\left\|\mathbb{E}[\mathbf{x}_t - F[\mathbf{x}_t \mid \mathbf{y}_{t-1}]][\mathbf{x}_t - F[\mathbf{x}_t \mid \mathbf{y}_{t-1}]]^\top - \mathbf{C}\,\mathbf{C}^\top\right\| &= \sigma_v^2 \left\|\mathbf{A}(\mathbf{G}_M^\top \mathbf{G}_M)^{-1} \mathbf{A}^\top\right\| \\ &\leq \sigma_v^2 \left\|\mathbf{A}\right\|^2 \left\|(\mathbf{G}_M^\top \mathbf{G}_M)^{-1}\right\| \to 0\end{aligned}$$

We conclude that $\mathbb{E}[\mathbf{x}_t - F[\mathbf{x}_t \mid \mathbf{y}_{t-1}]][\mathbf{x}_t - F[\mathbf{x}_t \mid \mathbf{y}_{t-1}]]^\top \to \mathbf{C}\,\mathbf{C}^\top$ pointwise as $M \to \infty$. Finally, since conditional expectation minimizes mean-square errors, we have

$$\begin{aligned}\mathbf{\Sigma}_{\infty,M} &= \mathbb{E}[\mathbf{x}_t - \mathbb{E}[\mathbf{x}_t \mid \mathbf{y}^{t-1}]][\mathbf{x}_t - \mathbb{E}[\mathbf{x}_t \mid \mathbf{y}^{t-1}]]^\top \\ &\preceq \mathbb{E}[\mathbf{x}_t - F[\mathbf{x}_t \mid \mathbf{y}_{t-1}]][\mathbf{x}_t - F[\mathbf{x}_t \mid \mathbf{y}_{t-1}]]^\top\end{aligned}$$

where $\preceq$ represents the Loewner order.[41] Clearly, we must have $\mathbf{C}\,\mathbf{C}^\top \preceq \mathbf{\Sigma}_{\infty,M}$ because $\mathbb{E}[\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{y}^{t-1}] = \mathbb{E}[\mathbf{x}_t \mid \mathbf{x}_{t-1}]$. Then by the continuity and anti-symmetry of the Loewner order, we conclude that $\mathbf{\Sigma}_{\infty,M} \to \mathbf{C}\,\mathbf{C}^\top$ pointwise as $M \to \infty$.

We are ready to prove that $\|\mathbf{A} - \mathbf{K}_M \mathbf{G}_M\| \to 0$. By the matrix Ricatti equation (A.5), we have

$$\left\|(\mathbf{A}\,\mathbf{\Sigma}_{\infty,M})^{-1}(\mathbf{\Sigma}_{\infty,M} - \mathbf{C}\,\mathbf{C}^\top)(\mathbf{A}^\top)^{-1}\right\| = \left\|\mathbf{I}_N - \mathbf{G}_M^\top(\mathbf{G}_M \mathbf{\Sigma}_{\infty,M} \mathbf{G}_M^\top + \mathbf{R}_M)^{-1} \mathbf{G}_M \mathbf{\Sigma}_{\infty,M}\right\|$$

Let $M \to \infty$ and we have the limit

$$\left\|\mathbf{I}_N - \mathbf{G}_M^\top(\mathbf{G}_M \mathbf{\Sigma}_{\infty,M} \mathbf{G}_M^\top + \mathbf{R}_M)^{-1} \mathbf{G}_M \mathbf{\Sigma}_{\infty,M}\right\| \to 0$$

Note that

$$\begin{aligned}\|\mathbf{A} - \mathbf{K}_M \mathbf{G}_M\| &= \left\|\mathbf{A} - \mathbf{A}\,\mathbf{\Sigma}_{\infty,M} \mathbf{G}_M^\top(\mathbf{G}_M \mathbf{\Sigma}_{\infty,M} \mathbf{G}_M^\top + \mathbf{R}_M)^{-1} \mathbf{G}_M\right\| \\ &\leq \|\mathbf{A}\| \left\|\mathbf{I}_N - \mathbf{\Sigma}_{\infty,M} \mathbf{G}^\top(\mathbf{G}_M \mathbf{\Sigma}_{\infty,M} \mathbf{G}_M^\top + \mathbf{R}_M)^{-1} \mathbf{G}_M\right\| \\ &= \|\mathbf{A}\| \left\|\mathbf{I}_N - \mathbf{G}_M^\top(\mathbf{G}_M \mathbf{\Sigma}_{\infty,M} \mathbf{G}_M^\top + \mathbf{R}_M)^{-1} \mathbf{G}_M \mathbf{\Sigma}_{\infty,M}\right\|\end{aligned}$$

where the last equality follows from taking transpose and the symmetry of $\mathbf{R}_M$ and $\mathbf{\Sigma}_{\infty,M}$. Let $M \to \infty$ and we have $\|\mathbf{A} - \mathbf{K}_M \mathbf{G}_M\| \to 0$, as desired.

[41] For any pair of positive semidefinite matrices $A, B \in \mathbb{R}^{N \times N}$, $A \preceq B$ iff $B - A$ is positive semidefinite.

We can further compute the convergence rate. Note that

$$
\begin{aligned}
M\left\|\mathbf{A}-\mathbf{K}_M\,\mathbf{G}_M\right\| &\leq M\left\|\mathbf{A}\right\|\left\|(\mathbf{A}\,\boldsymbol{\Sigma}_{\infty,M})^{-1}(\boldsymbol{\Sigma}_{\infty,M}-\mathbf{C}\,\mathbf{C}^\top)(\mathbf{A}^\top)^{-1}\right\| \\
&\leq \left\|\mathbf{A}\right\|\left\|(\mathbf{A}\,\boldsymbol{\Sigma}_{\infty,M})^{-1}\right\|\left\|M(\boldsymbol{\Sigma}_{\infty,M}-\mathbf{C}\,\mathbf{C}^\top)\right\|\left\|(\mathbf{A}^\top)^{-1}\right\| \\
&\leq \sigma_v^2\left\|\mathbf{A}\right\|^3\left\|(\mathbf{A}\,\boldsymbol{\Sigma}_{\infty,M})^{-1}\right\|\left\|(\mathbf{A}^\top)^{-1}\right\|\left\|\left(\frac{1}{M}\,\mathbf{G}_M^\top\,\mathbf{G}_M\right)^{-1}\right\|
\end{aligned}
$$

By Assumption **??** and our result that $\boldsymbol{\Sigma}_{\infty,M}\to\mathbf{C}\,\mathbf{C}^\top$, the RHS converges to some positive number as $M\to\infty$. Thus, we conclude that $\limsup_{M\to\infty} M\left\|\mathbf{A}-\mathbf{K}_M\,\mathbf{G}_M\right\|<\infty$.

□

A.3 Proof of Corollary ??

Proof. Manipulating the innovations representation from the proof of Proposition 1 gives

$$
\begin{aligned}
\widehat{\mathbf{x}}_{t+1} &= \mathbf{A}\,\mathbf{x}_t+\mathbf{K}(\mathbf{y}_t-\mathbf{G}\,\widehat{\mathbf{x}}_t) && \text{(A.6)}\\
&= (\mathbf{A}-\mathbf{K}\,\mathbf{G})\widehat{\mathbf{x}}_t+\mathbf{K}\,\mathbf{y}_t && \text{(A.7)}
\end{aligned}
$$

Define $\widetilde{\mathbf{x}}_t:=\mathbb{E}[\mathbf{x}_t\,|\,\mathbf{y}^t]$. So, $\widehat{\mathbf{x}}_{t+1}=\mathbf{A}\,\widetilde{\mathbf{x}}_t$.Next, suppose $\mathbf{A}-\mathbf{K}\,\mathbf{G}=\mathbf{0}$. Then,

$$
\begin{aligned}
\widehat{\mathbf{x}}_{t+1} &= \mathbf{K}\,\mathbf{y}_t && \text{(A.8)}\\
\mathbf{A}\,\widetilde{\mathbf{x}}_t &= \mathbf{A}\,\mathbf{L}\,\mathbf{y}_t && \text{(A.9)}
\end{aligned}
$$

for $\mathbf{L}=\boldsymbol{\Sigma}_\infty\,\mathbf{G}\,\boldsymbol{\Omega}^{-1}$ such that $\mathbf{K}=\mathbf{A}\,\mathbf{L}$. $\mathbf{A}\,\mathbb{E}[\mathbf{x}_t\,|\,\mathbf{y}^t]=\mathbf{A}\,\mathbf{L}\,\mathbf{y}_t$. Assuming an invertible $\mathbf{A}$, we have that

$$
\mathbb{E}[\mathbf{x}_t\,|\,\mathbf{y}^t]=\mathbf{L}\,\mathbf{y}_t
$$

i.e. that a forecast of $\mathbf{x}_t$ using all past observables $\mathbf{y}^t$ is equivalent to just using the current observables vector $\mathbf{y}_t$. □

A.4 Proof of Theorem 1

Proof. Consider the case $j=2$. By the definition of Frobenius norm, we have

$$
\begin{aligned}
\|\mathbf{B}_2^\infty\| &= \|\mathbf{G}(\mathbf{A}-\mathbf{K}\,\mathbf{G})\,\mathbf{K}\| \\
&= \sqrt{\mathrm{tr}\{\mathbf{K}^\top(\mathbf{A}-\mathbf{K}\,\mathbf{G})^\top\,\mathbf{G}^\top\,\mathbf{G}(\mathbf{A}-\mathbf{K}\,\mathbf{G})\,\mathbf{K}\}} \\
&= \sqrt{\mathrm{tr}\{(\mathbf{A}-\mathbf{K}\,\mathbf{G})^\top(\mathbf{G}^\top\,\mathbf{G})(\mathbf{A}-\mathbf{K}\,\mathbf{G})(\mathbf{K}\,\mathbf{K}^\top)\}} \\
&= \sqrt{\mathrm{tr}\left\{(\mathbf{A}-\mathbf{K}\,\mathbf{G})^\top\left(\frac{1}{M}\,\mathbf{G}^\top\,\mathbf{G}\right)[M(\mathbf{A}-\mathbf{K}\,\mathbf{G})](\mathbf{K}\,\mathbf{K}^\top)\right\}} && \text{(A.10)}
\end{aligned}
$$

By the matrix Ricatti equation (A.4), we have

$$\sigma_v^2\, \mathbf{K}\, \mathbf{K}^\top = \mathbf{\Sigma}_\infty - \mathbf{C}\, \mathbf{C}^\top - (\mathbf{A} - \mathbf{K}\, \mathbf{G})\mathbf{\Sigma}_\infty(\mathbf{A} - \mathbf{K}\, \mathbf{G})^\top$$

By Lemma 1, as $M \to \infty$, the RHS goes to $\mathbf{0}$. Thus, we have $\mathbf{K}\, \mathbf{K}^\top \to \mathbf{0}$ as $M \to \infty$.

Take $\limsup$ of equation (A.10) and use the continuity of tr and multiplication:

$$\limsup_{M\to\infty} \|\mathbf{B}_2^\infty\| = \sqrt{\operatorname{tr}\left\{\mathbf{0} \cdot \left(\lim_{M\to\infty} \frac{1}{M}\, \mathbf{G}^\top\, \mathbf{G}\right) \cdot \left[\limsup_{M\to\infty} M(\mathbf{A} - \mathbf{K}\, \mathbf{G})\right] \cdot \mathbf{0}\right\}}$$

By Assumption **??**, $\lim_{M\to\infty} \frac{1}{M}\, \mathbf{G}^\top\, \mathbf{G}$ exists and is finite. By Lemma 1, $\limsup_{M\to\infty} M(\mathbf{A} - \mathbf{K}\, \mathbf{G})$ is finite. We conclude that $\limsup_{M\to\infty} \|\mathbf{B}_2^\infty\| = 0$ and hence $\|\mathbf{B}_2^\infty\| \to 0$. Clearly, the case $j > 2$ can be proved in the same way, as $(\mathbf{A} - \mathbf{K}\, \mathbf{G})^j \to \mathbf{0}$. Inspecting equation (A.10), we can further conclude that for all $j \geq 1$

$$\limsup_{M\to\infty} M^{j-1} \left\|\mathbf{B}_j^\infty\right\| < \infty$$

Given that $\left\|\mathbf{B}_j^\infty\right\| \to 0$ for all $j \geq 2$, the infinite-order VAR (2) collapses to the first-order VAR (4), as claimed. □

A.5 Proof of Theorem 2

Proof. Using the infinite-order VAR (2), we can write the DFM likelihood as

$$\begin{aligned}
\ell^{DFM}(\mathbf{Y}; \mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}) &= \sum_{t=2}^{T} \ell(\mathbf{y}_t \mid \mathbf{y}^{t-1}; \mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}) \\
&= -\frac{1}{2} \sum_{t=2}^{T} \left\{\log|\mathbf{\Omega}| + \left(\mathbf{y}_t - \sum_{j=1}^{t-1} \mathbf{B}_j^\infty\, \mathbf{y}_{t-j}\right)^\top \mathbf{\Omega}^{-1} \left(\mathbf{y}_t - \sum_{j=1}^{t-1} \mathbf{B}_j^\infty\, \mathbf{y}_{t-j}\right)\right\}
\end{aligned}$$

where $\mathbf{\Omega} = \mathbf{G}\, \mathbf{\Sigma}_\infty\, \mathbf{G}^\top + \mathbf{R}$ is the variance-covariance matrix of the innovation. Similarly, using the first-order VAR (4), we can write the likelihood as

$$\begin{aligned}
\ell^{1}(\mathbf{Y}; \mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}) &= \sum_{t=2}^{T} \ell(\mathbf{y}_t \mid \mathbf{y}_{t-1}; \mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}) \\
&= -\frac{1}{2} \sum_{t=2}^{T} \left\{\log|\mathbf{\Omega}| + \left(\mathbf{y}_t - \mathbf{B}_1^\infty\, \mathbf{y}_{t-1}\right)^\top \mathbf{\Omega}^{-1} \left(\mathbf{y}_t - \mathbf{B}_1^\infty\, \mathbf{y}_{t-1}\right)\right\}
\end{aligned}$$

Subtract the two expressions and we obtain

$$
\begin{aligned}
&|\ell^{DFM}(\mathbf{Y};\mathbf{A},\mathbf{C},\mathbf{G},\mathbf{R}) - \ell^{1}(\mathbf{Y};\mathbf{A},\mathbf{C},\mathbf{G},\mathbf{R})| \\
&\quad = \frac{1}{2}\left|\sum_{t=2}^{T}\left\{\left(\sum_{j=2}^{t-1}\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j}\right)^{\top}\boldsymbol{\Omega}^{-1}\left(\sum_{j=2}^{t-1}\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j}\right) + 2\left(\sum_{j=2}^{t-1}\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j}\right)^{\top}\boldsymbol{\Omega}^{-1}(\mathbf{a}_t)\right\}\right| \\
&\quad \le \frac{1}{2}\sum_{t=2}^{T}\left\{\lambda_{\max}(\boldsymbol{\Omega}^{-1})\left\|\sum_{j=2}^{t-1}\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j}\right\|^2 + 2\left(\sum_{j=2}^{t-1}\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j}\right)^{\top}\boldsymbol{\Omega}^{-1}(\mathbf{a}_t)\right\}
\end{aligned}
$$

where $\lambda_{\max}(\boldsymbol{\Omega}^{-1})$ denotes the largest eigenvalue of $\boldsymbol{\Omega}^{-1}$ and $\mathbf{a}_t = \mathbf{y}_t - \sum_{j=1}^{t-1}\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j}$.

It suffices to show that

1. $\mathbb{E}\left\|\sum_{j=2}^{t-1}\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j}\right\|^2 \to 0$ for all $t = 1,\ldots,T$
2. $\limsup_{M\to\infty}\lambda_{\max}(\boldsymbol{\Omega}^{-1}) < \infty$
3. $\mathbb{E}(\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j})^{\top}\boldsymbol{\Omega}^{-1}\,\mathbf{a}_t = 0$ for all $j \ge 2$ and $t = 1,\ldots,T$

Claim 1. Using the measurement equation for $\mathbf{y}_t$, we have

$$
\left\|\frac{1}{\sqrt{M}}\,\mathbf{y}_t\right\| \le \left\|\frac{1}{\sqrt{M}}\,\mathbf{G}\right\|\|\mathbf{x}_t\| + \left\|\frac{1}{\sqrt{M}}\,\mathbf{v}_t\right\| = \sqrt{\operatorname{tr}\left\{\frac{1}{M}\,\mathbf{G}^{\top}\,\mathbf{G}\right\}}\cdot\|\mathbf{x}_t\| + \sqrt{\frac{1}{M}\sum_{i=1}^{M}v_{i,t}^2}
$$

Note that the distribution of $\|\mathbf{x}_t\|$ is invariant to M and has finite mean. By Assumption **??**, we have

$$
\sqrt{\operatorname{tr}\left\{\frac{1}{M}\,\mathbf{G}^{\top}\,\mathbf{G}\right\}}\cdot\|\mathbf{x}_t\| \overset{a.s.}{\to} \lambda\,\|x_t\|
$$

for some $\lambda > 0$. By SLLN, we have

$$
\sqrt{\frac{1}{M}\sum_{i=1}^{M}v_{i,t}^2} \overset{a.s.}{\to} \sigma_v
$$

It follows that $\limsup_{M\to\infty}\left\|\frac{1}{\sqrt{M}}\,\mathbf{y}_t\right\| < \infty$ almost surely. For any $j \ge 2$, by Theorem 1, we have

$$
\begin{aligned}
\limsup_{M\to\infty}\|\mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j}\| &\le \limsup_{M\to\infty}(\sqrt{M}\,\|\mathbf{B}_j^{\infty}\|)\left\|\frac{1}{\sqrt{M}}\,\mathbf{y}_{t-j}\right\| \\
&= \underbrace{\limsup_{M\to\infty}\sqrt{M}\,\|\mathbf{B}_j^{\infty}\|)}_{=0}\cdot\underbrace{\limsup_{M\to\infty}\left\|\frac{1}{\sqrt{M}}\,\mathbf{y}_{t-j}\right\|}_{<\infty\ a.s.}
\end{aligned}
$$

Thus, $\left\|\mathbf{B}_j^\infty \mathbf{y}_{t-j}\right\| \overset{a.s.}{\to} 0$ for all $j \geq 2$ and $t = 1, \ldots, T$. It follows that when M sufficiently large, $\left\|\sum_{j=2}^{t-1} \mathbf{B}_j^\infty \mathbf{y}_{t-j}\right\|^2$ is uniformly bounded above almost surely. Then by Dominated Convergence Theorem, we have $\mathbb{E}\left\|\sum_{j=2}^{t-1} \mathbf{B}_j^\infty \mathbf{y}_{t-j}\right\|^2 \to 0$ for all $t = 1, \ldots, T$.

Claim 2. Fix M. By Spectral Theorem, there exists $\mathbf{P}, \mathbf{D} \in \mathbb{R}^{M\times M}$ such that $\mathbf{P}^\top \mathbf{P} = \mathbf{I}_M$, $\mathbf{D}$ is diagonal, and $\mathbf{G}\,\boldsymbol{\Sigma}_\infty\,\mathbf{G}^\top = \mathbf{P}^\top \mathbf{D}\mathbf{P}$. Then

$$\boldsymbol{\Omega} = \mathbf{G}\,\boldsymbol{\Sigma}_\infty\,\mathbf{G}^\top + \mathbf{R} = \mathbf{P}^\top \mathbf{D}\mathbf{P} + \sigma_v^2 \mathbf{P}^\top \mathbf{P} = \mathbf{P}^\top(\mathbf{D} + \sigma_v^2 \mathbf{I}_M)\mathbf{P}$$

It follows that $\boldsymbol{\Omega}^{-1} = \mathbf{P}^\top(\mathbf{D} + \sigma_v^2 \mathbf{I}_M)^{-1}\mathbf{P}$. Since all the entry of $\mathbf{D}$ is non-negative, the largest eigenvalue of $\boldsymbol{\Omega}^{-1}$ is smaller than $1/\sigma_v^2$. Then clearly $\limsup_{M\to\infty} \lambda_{\max}(\boldsymbol{\Omega}^{-1}) < \infty$

Claim 3. Fix $j \geq 2$. As shown in Claim 2, we can write $\boldsymbol{\Omega}^{-1} = \mathbf{P}^\top(\mathbf{D} + \sigma_v^2 \mathbf{I}_M)^{-1}\mathbf{P}$. Then

$$\begin{aligned}(\mathbf{B}_j^\infty \mathbf{y}_{t-j})^\top \boldsymbol{\Omega}^{-1} \mathbf{a}_t &= (\mathbf{P}\,\mathbf{B}_j^\infty \mathbf{y}_{t-j})^\top(\mathbf{D} + \sigma_v^2 \mathbf{I}_M)^{-1}(\mathbf{P}\,\mathbf{a}_t) \\ &= \sum_{i=1}^{M} \frac{1}{d_i + \sigma^2 - v}(\mathbf{P}\,\mathbf{B}_j^\infty \mathbf{y}_{t-j})_i \cdot (\mathbf{P}\,\mathbf{a}_t)_i\end{aligned}$$

where $(\cdot)_i$ denote the i entry of the vector. Clearly, for any i, $(\mathbf{P}\,\mathbf{B}_j^\infty \mathbf{y}_{t-j})_i \in \mathcal{H}(\mathbf{y}^{t-1})$ and $(\mathbf{P}\,\mathbf{a}_t)_i \in \mathcal{H}(\mathbf{a}_t)$. Then by the orthogonality condition $\mathbf{a}_t \perp \mathcal{H}(\mathbf{y}^{t-1})$, we have

$$\mathbb{E}[(\mathbf{P}\,\mathbf{B}_j^\infty \mathbf{y}_{t-j})_i \cdot (\mathbf{P}\,\mathbf{a}_t)_i] = 0 \quad \forall i$$

It follows that $\mathbb{E}(\mathbf{B}_j^\infty \mathbf{y}_{t-j})^\top \boldsymbol{\Omega}^{-1} \mathbf{a}_t = 0$, as desired.

By the three claims, the proof is complete and we conclude that

$$\lim_{M\to\infty} \mathbb{E}|\ell^{DFM}(\mathbf{Y}; \mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}) - \ell^1(\mathbf{Y}; \mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R})| = 0$$

□

A.6 Proof of Proposition 2

Proof. WLOG, let $\mathbf{c}_{ss} = \mathbf{0}$. As shown in Auclert et al. (2021b), up to first-order, the household's policy can be written as

$$\mathbf{c}_t = \sum_{j=0}^{\infty} \sum_{p\in\mathcal{P}} \frac{\partial\,\mathbf{c}}{\partial p_j}\,\mathbb{E}_t[\tilde{p}_{t+j}]$$

where $\frac{\partial\,\mathbf{c}}{\partial p_j} \in \mathbb{R}^M$ is the derivative of individual policy wrt. the j-period ahead aggregate input $p \in \mathcal{P}$ and $\tilde{p}_{t+j}$ denotes the deviation of p from its steady-state value.

Put $\tilde{p}_{t:} := (\tilde{p}_t, \tilde{p}_{t+1}, \dots)^\top$. Using the impulse response functions, we can write

$$\begin{aligned}\mathbb{E}_t[\tilde{p}_{t:}] &= \mathbb{E}_{t-1}[\tilde{p}_{t:}] + \mathcal{I}_e^p \boldsymbol{\epsilon}_t \\ &= F\, \mathbb{E}_{t-1}[\tilde{p}_{t-1:}] + \mathcal{I}_e^p \boldsymbol{\epsilon}_t\end{aligned}$$

where F is the shift forward operator. Iterate backward and we obtain the MA representation

$$\mathbb{E}_t[\tilde{p}_{t:}] = \sum_{j=0}^{\infty} F^j \mathcal{I}_e^p \boldsymbol{\epsilon}_{t-j}$$

Let $\mathcal{J}_p^c$ be the infinite-dimensional matrix of which the j column is $\frac{\partial \mathbf{c}}{\partial p_j}$. Substitute back into the policy function:

$$\mathbf{c}_t = \sum_{p \in \mathcal{P}} \mathcal{J}_p^c\, \mathbb{E}_t[\tilde{p}_{t:}] = \sum_{p \in \mathcal{P}} \mathcal{J}_p^c \sum_{j=0}^{\infty} F^j \mathcal{I}_e^p \boldsymbol{\epsilon}_{t-j} = \sum_{j=0}^{\infty} \underbrace{\sum_{p \in \mathcal{P}} \mathcal{J}_p^c F^j \mathcal{I}_e^p}_{\Psi_j^c} \boldsymbol{\epsilon}_{t-j}$$

□

A.7 Proof of Proposition 3

Proof. Let $\mathcal{I}_{e,x}^p$ denote the impulse response functions of p wrt. an exogenous shock to x. Then

$$\mathcal{I}_{e,x}^p = \mathcal{J}_x^p \mathcal{I}_e^x$$

where $\mathcal{I}_e^x = (1, \rho_x, \rho_x^2, \dots)^\top$ is the impulse response function of x wrt. the shock and $\rho_x \in (0,1)$ is the associated AR(1) coefficient. Recall that by Proposition 2, we have

$$\begin{aligned}\mathbf{c}_t &= \sum_{j=0}^{\infty} \sum_{p \in \mathcal{P}} \mathcal{J}_p^c F^j \mathcal{I}_e^p \boldsymbol{\epsilon}_{t-j} \\ &= \sum_{j=0}^{\infty} \sum_{p \in \mathcal{P}} \sum_{x \in \mathcal{E}} \mathcal{J}_p^c F^j \mathcal{I}_{e,x}^p \boldsymbol{\epsilon}_{t-j}^x \\ &= \sum_{j=0}^{\infty} \sum_{x \in \mathcal{E}} \sum_{p \in \mathcal{P}} \mathcal{J}_p^c F^j \mathcal{J}_x^p \mathcal{I}_e^x \boldsymbol{\epsilon}_{t-j}^x\end{aligned}$$

Clearly, if $F\mathcal{J}_x^p = \mathcal{J}_x^p F$, then $F^j\mathcal{J}_x^p = \mathcal{J}_x^p F^j\ \forall j \in \mathbb{N}$. Using this condition, we have

$$\begin{aligned}\mathbf{c}_t &= \sum_{j=0}^{\infty}\sum_{x\in\mathcal{E}}\sum_{p\in\mathcal{P}} \mathcal{J}_p^c F^j \mathcal{J}_x^p \mathcal{I}_e^x \boldsymbol{\epsilon}_{t-j}^x \\ &= \sum_{j=0}^{\infty}\sum_{x\in\mathcal{E}}\sum_{p\in\mathcal{P}} \mathcal{J}_p^c \mathcal{J}_x^p F^j \mathcal{I}_e^x \boldsymbol{\epsilon}_{t-j}^x \\ &= \sum_{x\in\mathcal{E}}\sum_{p\in\mathcal{P}} \mathcal{J}_p^c \mathcal{J}_x^p \left(\sum_{j=0}^{\infty} F^j \mathcal{I}_e^x \boldsymbol{\epsilon}_{t-j}^x\right)\end{aligned}$$

Note that

$$\sum_{j=0}^{\infty} F^j \mathcal{I}_e^x \boldsymbol{\epsilon}_{t-j}^x = \sum_{j=0}^{\infty} \rho_x^j \mathcal{I}_e^x \boldsymbol{\epsilon}_{t-j}^x = \mathcal{I}_e^x \sum_{j=0}^{\infty} \rho_x^j \boldsymbol{\epsilon}_{t-j}^x = \mathcal{I}_e^x x_t$$

Thus, the policy function becomes

$$\mathbf{c}_t = \sum_{x\in\mathcal{E}}\sum_{p\in\mathcal{P}} \mathcal{J}_p^c \mathcal{J}_x^p \mathcal{I}_e^x x_t = \sum_{x\in\mathcal{E}} \mathbf{G}_x\, x_t$$

where $\mathbf{G}_x := \sum_{p\in\mathcal{P}} \mathcal{J}_p^c \mathcal{J}_x^p \mathcal{I}_e^x$ is the impulse response function of $\mathbf{c}_t$ wrt. shock to x. Now, let $\mathbf{x}_t$ be the vector of the shock process x. Then we have the low-dimensional DFM representation

$$\begin{aligned}\mathbf{x}_{t+1} &= \mathbf{A}\,\mathbf{x}_t + \boldsymbol{\epsilon}_{t+1} \\ \mathbf{c}_t &= \mathbf{G}\,\mathbf{x}_t\end{aligned}$$

where $\mathbf{A}$ is the diagonal matrix of the AR(1) coefficients and $\mathbf{G}$ is the matrix from stacking $\mathbf{G}_x$. $\square$

Appendix B Algorithms

B.1 Dynamic Mode Decomposition

Another option for computing the reduced-rank VAR(1) matrix $\mathbf{B}_r$ is to employ the Dynamic Mode Decomposition (DMD) algorithm. The DMD has become a workhorse tool in the fluid dynamics literature, originally introduced by Schmidt and Sesterhenn (2010) and later developed by Tu et al. (2014). Existing applications of the DMD also include epidemiology, neuroscience and video processing (see Brunton and Kutz (2022)).

Given simulated data matrices $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}'$, the DMD estimates the reduced-rank VAR associated with the simulated data by solving

$$\mathbf{B}_r = \underset{\text{rank}(B)=r}{\arg\min} \left\| \widetilde{\mathbf{Y}}' - B\widetilde{\mathbf{Y}} \right\| \tag{B.1}$$

where $\|\cdot\|$ denotes the Frobenius norm. To compute $\mathbf{B}_r$, represent $\mathbf{Y}$ with a reduced Singular Value Decomposition (SVD)

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^\top$$

where $\widetilde{\mathbf{U}}$ is $N \times N$, $\widetilde{\mathbf{\Sigma}}$ is $N \times N$ and $\widetilde{\mathbf{V}}$ is $T \times N$. We compress $\widetilde{\mathbf{Y}}$ by using its r largest singular values:

$$\widetilde{\mathbf{Y}} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top,$$

where $\mathbf{U} = \widetilde{\mathbf{U}}[:, : r]$, $\mathbf{\Sigma} = \widetilde{\mathbf{\Sigma}}[: r, : r]$ has r singular values as its only non-zero entries, and $\mathbf{V}^\top = \widetilde{\mathbf{V}}^\top[: r, :]$. Here $\mathbf{U}$ is $N \times T$, $\mathbf{V}$ is $T \times r$, $\mathbf{\Sigma}$ is $r \times r$, and $\mathbf{V}^\top$ is $r \times T$.[42]

We use this reduced-order SVD approximation of $\widetilde{\mathbf{Y}}$ to compute

$$\mathbf{B}_r = \widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}^+, \tag{B.2}$$

where by construction $\mathbf{B}_r$ is rank r.[43] The covariance matrix of the residuals, $\widetilde{\mathbf{a}}_t = \widetilde{\mathbf{y}}_t - \mathbf{B}_r\widetilde{\mathbf{y}}_{t-1}$, is computed via

$$\mathbf{\Omega}_r = \tfrac{1}{T-1}\sum_{t=1}^{T} \widetilde{\mathbf{a}}_t\widetilde{\mathbf{a}}_t^\top \tag{B.3}$$

B.2 Second-order reduced-rank VAR

Ib some instances, we might want to estimate a reduced-rank VAR(2). MORE CONTEXT HERE

[42] Note that all we need here is a truncated SVD, which can be very efficiently computed using existing machine-learning packages (e.g. scikit-learn).

[43] The DMD also provides estimates of the underlying factors $\mathbf{x}_t$ as well as estimates of $\mathbf{G}$ and $\mathbf{A}$. Though we do not use them here, see **?**, sec. 2.1 for full details of the DMD algorithm and connections with linear state-space models.

$$\mathbf{y}_t = \mathbf{B}_{1,r+1}\,\mathbf{y}_{t-1} + \mathbf{B}_{2,r}\,\mathbf{y}_{t-2} + \mathbf{a}_t \tag{B.4}$$

$$\text{where} \quad rank(\mathbf{B}_{1,r+1}) = r+1 \tag{B.5}$$

$$rank(\mathbf{B}_{2,r}) = r \tag{B.6}$$

Computing the reduced-rank VAR matrices becomes slightly more involved. At first glance it may seem a possible route to create

$$\mathbf{y}_t = [\mathbf{B}^{1,r+1}\mathbf{B}_{2,r}] \begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \end{bmatrix} \tag{B.7}$$

and proceed as before. But neither approaches above enable the imposition of a partition of reduced-rank matrices in the required way. One solution is to the invoke Frisch-Waugh-Lovell Theorem. First, we will compute the regression coefficient

$$\mathbf{y}_{t-1} = \mathbf{D}_1\,\mathbf{y}_{t-2} + \mathbf{v}_{1t} \tag{B.8}$$

$$\mathbf{D}_1 = \mathbb{E}[\mathbf{y}_{t-1}\,\mathbf{y}_{t-2}^\top]\mathbb{E}[\mathbf{y}_{t-2}\,\mathbf{y}_{t-2}^\top]^{-1} \tag{B.9}$$

$$\mathbb{E}[\mathbf{v}_{1t}\mathbf{v}_{1t}^\top] = \mathbf{\Sigma}_0 - \mathbf{D}_1\mathbf{\Sigma}_0\mathbf{D}_1^\top \tag{B.10}$$

By Frisch-Waugh-Lovell, $\mathbf{B}_1$ is the regression coefficient in

$$\mathbf{y}_t = \mathbf{B}_1\,\mathbf{v}_{1t} + \boldsymbol{\varepsilon}_{\mathbf{1t}} \tag{B.11}$$

To then estimate the reduced-rank $\mathbf{B}_{1r}$, we implement the canonical correlations formula in equaton (B.11). Similarly for $\mathbf{B}_{2,r}$,

$$\mathbf{y}_{t-2} = \mathbf{D}_2\,\mathbf{y}_{t-1} + \mathbf{v}_{2t} \tag{B.12}$$

$$\mathbf{D}_2 = \mathbb{E}[\mathbf{y}_{t-2}\,\mathbf{y}_{t-1}^\top]\mathbb{E}[\mathbf{y}_{t-1}\,\mathbf{y}_{t-1}^\top]^{-1} \tag{B.13}$$

$$\mathbb{E}[\mathbf{v}_{2t}\mathbf{v}_{2t}^\top] = \mathbf{\Sigma}_0 - \mathbf{D}_2\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^\top\mathbf{D}_2^\top + \mathbf{D}_2\mathbf{\Sigma}_0\mathbf{D}_2^\top \tag{B.14}$$

Then by Frisch-Waugh-Lovell, $\mathbf{B}_2$ is the regression coefficient is

$$\mathbf{y}_t = \mathbf{B}_2\,\mathbf{v}_{2t} + \boldsymbol{\varepsilon}_{\mathbf{2t}} \tag{B.15}$$

Similarly to estimate the reduced-rank $\mathbf{B}_{2r}$, we implement the Anderson (1951) formula on (B.15).

Given $\mathbf{B}_{1r}$ and $\mathbf{B}_{2r}$, we can compute the covariance matrix

$$\mathbf{\Omega} = \mathbb{E}[\mathbf{a}_t\,\mathbf{a}_t^\top] = \mathbf{\Sigma}_0 - \mathbf{B}_{1r}\,\mathbf{\Sigma}_0\,\mathbf{B}_{1r}^\top - \mathbf{B}_{2r}\,\mathbf{\Sigma}_0\,\mathbf{B}_{2r}^\top - \mathbf{B}_{1r}\,\mathbf{\Sigma}_1\,\mathbf{B}_{2r}^\top - \mathbf{B}_{2r}\,\mathbf{\Sigma}_1^\top\,\mathbf{B}_{1r}^\top \tag{B.16}$$

Appendix C Estimation details

C.1 Frequency-domain estimation

We use the same 500 simulated micro datasets $(M \times T)$ as in the main estimation exercise. Recall that by Proposition 2, the de-meaned data has a MA representation

$$\mathbf{c}_t = \sum_{j=0}^{\infty} \Psi_j^c \boldsymbol{\epsilon}_{t-j} + \mathbf{v}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \Sigma_e)$$

where $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{R})$ is measurement error. For a given set of parameters θ, we can efficiently compute the MA coefficient matrices Ψ_j^c using the SSJ method.

Following Hansen and Sargent (1981), we approximate the likelihood using Whittle approximation:

$$L(\mathbf{c}; \theta) = -\frac{1}{2} \sum_{j=0}^{T-1} \left[\log 2\pi + \log(\det S(\omega_j; \theta)) + \mathbf{tr}(S(\omega_j; \theta)^{-1} I(\mathbf{c}; \omega_j)) \right] \tag{C.1}$$

where $\omega_j := \frac{2\pi j}{T}$, $S(\omega_j; \theta)$ is the spectral density of $\mathbf{c}$ at frequency ω_j, and $I(\mathbf{c}; \omega_j)$ is the periodogram of the data at frequency ω_j. By definition, the periodogram is given by

$$I(\mathbf{c}; \omega_j) := \frac{1}{T} \left(\sum_{t=1}^{T} \mathbf{c}_t \exp(-i\omega_j t) \right) \left(\sum_{t=1}^{T} \mathbf{c}_t \exp(i\omega_j t) \right)'$$

By the MA representation, the spectral density is given by

$$S(\omega_j; \theta) = \left(\sum_{j=0}^{\infty} \Psi_j^c(\theta) \exp(-i\omega_j j) \right) \Sigma_e \left(\sum_{j=0}^{\infty} \Psi_j^c(\theta) \exp(i\omega_j j) \right)' + \mathbf{R}$$

Note that both $I(\mathbf{c}; \omega_j)$ and $S(\omega_j; \theta)$ are M-dimensional matrices and can be computed by applying the Discrete Fourier Transform to the data matrix $\mathbf{c}$ and MA coefficient array $\{\Psi_j^c : j = 0, \ldots, T\}$. Also, the symmetry of Fourier transform implies that we only need to evaluate the summands in (C.1) for $j = 0, \ldots, \lfloor \frac{T-1}{2} \rfloor$

Given a dataset $\mathbf{c}$, we construct the likelihood using the formula above and find the parameter that maximizes the likelihood using standard optimization algorithm. The distribution of the estimates is plotted in Figure E.1.

C.2 Random Walk Metropolis Hastings estimation

We apply a simple Random Walk Metropolis Hastings Algorithm to sample from the posterior distribution of the parameters. We also use a tuned proposal covariance matrix and adaptive step size proposed by Atchadé and Rosenthal (2005) and Haario et al. (2001).

We set the prior of the parameters to be flat. We initialize the MCMC sampler at the mode of the posterior distribution and generate 50,000 draws, discarding the first 10,000.

Appendix D Simulation details

Below, we provide a short summary of the models used to generate Figure **??**. In all cases, our simulated dataset has 300 units in the cross-section, of length $T = 10,000$.

Krusell-Smith: We use a standard Krusell-Smith model, the code of which is available on the SHADE-econ github page. We simulate the cross-section of consumption, $M = 300, T = 10,000$ with one aggregate shock, TFP.

One-Asset HANK: This is the same model in Section 5, with two aggregate shocks, TFP and Monetary policy; and one cross-sectional shock to income dispersion.

Two-Asset HANK: We implement the two-asset HANK model, the code of which is available on the SHADE-econ github page. There are three aggregate shocks – TFP, government spending and r^*.

Hetero. Firms: We implement a version of the investment model with heterogeneous-firms by Khan and Thomas (2008), using the code provided by Liu and Plagborg-Møller (2023). There is only one aggregate shock – TFP.

Appendix E Supplementary tables and figures

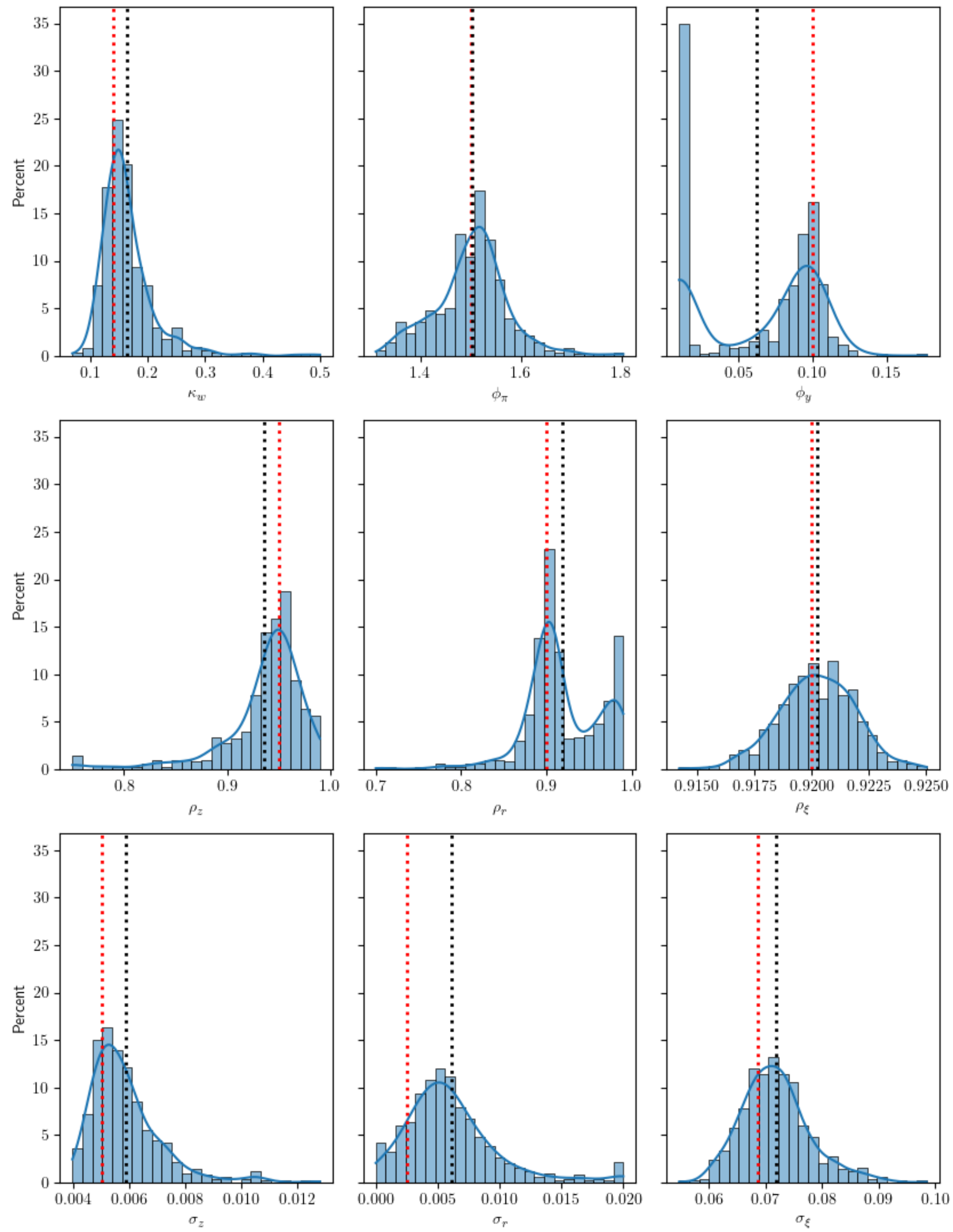

Figure E.1: Micro data (FD estimation): Finite-sample parameter distribution

NOTE. The plots are generated from 500 Monte Carlo draws. Red line is the true value and black line is the mean of the estimates.

figures_v2/productivity_distribution_compare.png

Figure E.2: Productivity distirbution comparison: model and data

figures_v2/missing_data.png

Figure E.3: Consumption data-availability matrix for CEX data

References

ACHARYA, S., W. CHEN, M. DEL NEGRO, K. DOGRA, E. MATLIN, AND R. SARFATI (2020): "Estimating HANK: macro time series and micro moments," *Manuscript, March*.

ALVES, F., G. KAPLAN, B. MOLL, AND G. L. VIOLANTE (2020): "A further look at the propagation of monetary policy shocks in HANK," *Journal of Money, Credit and Banking*, 52, 521–559.

ALVEZ, F., G. KAPLAN, B. MOLL, AND G. L. VIOLANTE (2020): "A Further Look at the Propagation of Monetary Policy Shocks in HANK," *Journal of Money, Credit and Banking*, 52, 521–559.

AN, S. AND F. SCHORFHEIDE (2007): "Bayesian analysis of DSGE models," *Econometric reviews*, 26, 113–172.

ANDERSON, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *The Annals of Mathematical Statistics*, 22, 327–351.

ANDERSON, T. W. AND H. RUBIN (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics*, 20, 46–63.

ARELLANO, M., R. BLUNDELL, AND S. BONHOMME (2017): "Earnings and consumption dynamics: a nonlinear panel data framework," *Econometrica*, 85, 693–734.

ARUOBA, S. B., F. X. DIEBOLD, J. NALEWAIK, F. SCHORFHEIDE, AND D. SONG (2016): "Improving GDP measurement: A measurement-error perspective," *Journal of Econometrics*, 191, 384–397.

ATCHADÉ, Y. F. AND J. S. ROSENTHAL (2005): "On adaptive markov chain monte carlo algorithms," *Bernoulli*, 11, 815–828.

AUCLERT, A., B. BARDÓCZY, M. ROGNLIE, AND L. STRAUB (2021a): "Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models," Tech. rep., Stanford.

——— (2021b): "Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models," *Econometrica*, 89, 2375–2408.

AUCLERT, A. AND M. ROGNLIE (2018): "Inequality and aggregate demand," Tech. rep., National Bureau of Economic Research.

AUCLERT, A., M. ROGNLIE, AND S. LUDWIG (2024): "The Intertemporal Keynesian Cross," *Journal of Political Economy, forthcoming*.

BAI, J. AND S. NG (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

——— (2006): "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions," *Econometrica*, 74, 1133–1150.

BAYER, C., B. BORN, AND R. LUETTICKE (2024): "Shocks, Frictions, and Inequality in U.S. Business Cycles," *American Economic Review*.

BAYER, C. AND R. LUETTICKE (2020): "Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods," *Quantitative Economics*, 11, 1253–1288.

BRUNTON, S. L. AND J. N. KUTZ (2022): *Data-Driven Science and Engineering: Machine Learnings, Dynamical Systems, and Control, second edition*, Cambridge University Press.

CHALLE, E., J. MATHERON, X. RAGOT, AND J. F. RUBIO-RAMIREZ (2017): "Precautionary saving and aggregate demand," *Quantitative Economics*, 8, 435–478.

CHAMBERLAIN, G. AND M. ROTHSCHILD (1982): "Arbitrage, factor structure, and mean-variance analysis on large asset markets," National Bureau of Economic Research Cambridge, Mass., USA.

——— (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

CHANG, M., X. CHEN, AND F. SCHORFHEIDE (2021): "Heterogeneity and aggregate fluctuations," Tech. rep., National Bureau of Economic Research.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of political Economy*, 113, 1–45.

CHRISTIANO, L. J. AND R. J. VIGFUSSON (2003): "Maximum likelihood in the frequency domain: the importance of time-to-plan," *Journal of Monetary Economics*, 50, 789–815.

COIBION, O., Y. GORODNICHENKO, L. KUENG, AND J. SILVIA (2017): "Innocent Bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 88, 70–89.

DOLADO, J. J., G. MOTYOVSZKI, AND E. PAPPA (2021): "Monetary policy and inequality under labor market frictions and capital-skill complementarity," *American economic journal: macroeconomics*, 13, 292–332.

FERNÁNDEZ-VILLAVERDE, J., S. HURTADO, AND G. NUNO (2023): "Financial frictions and the wealth distribution," *Econometrica*, 91, 869–901.

FERNÁNDEZ-VILLAVERDE, J., J. F. RUBIO-RAMÍREZ, AND F. SCHORFHEIDE (2016): "Solution and estimation methods for DSGE models," in *Handbook of macroeconomics*, Elsevier, vol. 2, 527–724.

FLODEN, M. AND J. LINDÉ (2001): "Idiosyncratic risk in the United States and Sweden: Is there a role for government insurance?" *Review of Economic dynamics*, 4, 406–437.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The Generalized Dynamic-Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82, 540–554.

FORNI, M. AND M. LIPPI (2001): "The Generalized Dynamic-Factor Model: Representation Theory," *Econometric Theory*, 17, 1113–1141.

GUVENEN, F., S. SCHULHOFER-WOHL, J. SONG, AND M. YOGO (2017): "Worker betas: Five facts about systematic earnings risk," *American Economic Review*, 107, 398–403.

HAARIO, H., E. SAKSMAN, AND J. TAMMINEN (2001): "An adaptive Metropolis algorithm," *Bernoulli*, 223–242.

HANSEN, L. AND T. SARGENT (1981): "Exact linear rational expectations models: specification and estimation," Tech. rep., Federal Reserve Bank of Minneapolis.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2017): "Optimal tax progressivity: An analytical framework," *The Quarterly Journal of Economics*, 132, 1693–1754.

HERBST, E. AND F. SCHORFHEIDE (2014): "Sequential Monte Carlo sampling for DSGE models," *Journal of Applied Econometrics*, 29, 1073–1098.

HOFFMAN, M. D., A. GELMAN, ET AL. (2014): "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.*, 15, 1593–1623.

JUSTINIANO, A., G. E. PRIMICERI, AND A. TAMBALOTTI (2011): "Investment shocks and the relative price of investment," *Review of Economic Dynamics*, 14, 102–121.

KHAN, A. AND J. K. THOMAS (2008): "Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics," *Econometrica*, 76, 395–436.

KRUSELL, P. AND A. A. SMITH, JR (1998): "Income and wealth heterogeneity in the macroeconomy," *Journal of political Economy*, 106, 867–896.

LIU, L. AND M. PLAGBORG-MØLLER (2023): "Full-information estimation of heterogeneous agent models using macro and micro data," *Quantitative Economics*, 14, 1–35.

LUNGQVIST, L. AND T. J. SARGENT (2018): *Recursive Macroeconomic Theory*, The MIT Press.

MCKAY, A. AND J. F. WIELAND (2021): "Lumpy durable consumption demand and the limited ammunition of monetary policy," *Econometrica*, 89, 2717–2749.

MONGEY, S. AND J. WILLIAMS (2017): "Firm dispersion and business cycles: Estimating aggregate shocks using panel data," *Manuscript, New York University*.

MORALES-JIMENEZ, C. AND L. STEVENS (2024): "Price Rigidities in U.S. Business Cycles," Tech. rep., University of Maryland, working paper.

PATTERSON, C. (2023): "The matching multiplier and the amplification of recessions," *American Economic Review*, 113, 982–1012.

PLAGBORG-MØLLER, M. (2019): “Bayesian inference on structural impulse response functions,” *Quantitative Economics*, 10, 145–184.

REITER, M. (2009): “Solving heterogeneous-agent models by projection and perturbation,” *Journal of Economic Dynamics & Control*, 33, 649–665.

ROMER, C. D. AND D. H. ROMER (2004): “A new measure of monetary shocks: Derivation and implications,” *American economic review*, 94, 1055–1084.

ROTEMBERG, J. J. (1982): “Sticky prices in the United States,” *Journal of Political Economy*, 90, 1187–1211.

ROUWENHORST, K. G. (1995): “Asset pricing implications of equilibrium business cycle models,” *Frontiers of business cycle research*, 1, 294–330.

SARGENT, T. J. AND Y. J. SELVAKUMAR (2024): “VAR and DMD,” Tech. rep., New York University, NYU Economics.

SCHMIDT, P. AND J. SESTERHENN (2010): “Dynamic Mode Decomposition of numerical and experimental data,” *Journal of Fluid Dynamics*, 656, 5–28.

SMETS, F. AND R. WOUTERS (2007): “Shocks and frictions in US business cycles: A Bayesian DSGE approach,” *American economic review*, 97, 586–606.

STOCK, J. H. AND M. W. WATSON (2002): “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 1167–1179.

TAUCHEN, G. (1986): “Finite state markov-chain approximations to univariate and vector autoregressions,” *Economics letters*, 20, 177–181.

TU, J. H., C. W. ROWLEY, D. M. LUCHTENBURG, S. L. BRUNTON, AND J. N. KUTZ (2014): “On dynamic mode decomposition: Theory and applications,” *Journal of Computational Dynamics*, 1, 391–421.

WERNING, I. (2015): “Incomplete markets and aggregate demand,” Tech. rep., National Bureau of Economic Research.

WINBERRY, T. (2018): “A method for solving and estimating heterogeneous agent macro models,” *Quantitative Economics*, 9, 1123–1151.

——— (2021): “Lumpy Investment, Business Cycles, and Stimulus Policy,” *American Economic Review*, 111, 364–396.