Estimating Linearized Heterogeneous Agent Models Using Panel Data *

Tamas Papp† **Michael Reiter**‡

This version: December 22, 2019

Abstract

We develop a method to estimate heterogeneous agent models that uses not only time series of macroeconomic aggregates, but can also incorporate micro level data (repeated cross-section or panel). The micro data may be collected at lower frequency and time-aggregated. The method is based on the linearization approach of Reiter (2009), combined with optimal state aggregation as in Reiter (2010). The model may contain decision problems with both continuous and discrete choice. Linearity of the model solution allows fast computation of second moments and likelihood. We discuss various computational devices to maximize the speed of the estimation.

JEL classification: C63, C68, E21

*We thank Christian Haefke, Leopold Soegner, participants of the 2019 JEDC conference in Gerzensee, the editors and in particular our discussant Wouter den Haan for helpful comments and discussants.

†Institute for Advanced Studies, Vienna, E-mail: tkpapp@gmail.com

‡ Institute for Advanced Studies, Vienna and NYU Abu Dhabi, E-mail: michael.reiter@ihs.ac.at

1 Introduction

With the rise of "HANK" (heterogeneous agent New Keynesian) models (Kaplan et al. 2018), heterogeneous agent models are now competing with standard DSGE models in macroeconomic policy analysis, and there is a need to estimate the parameters of this class of models. This appears to be very difficult: due to the high dimensionality of the state space, the solution of these models used to be very time-consuming, and estimation requires to solve a model many times. The early solution methods (Den Haan 1997; Krusell and Smith 1998) are nonlinear methods exploiting approximate aggregation to reduce the dimensionality. An alternative is the linearization approach of Reiter (2009), which allows for a high-dimensional state space by linearizing in aggregate variables. Linearization is convenient for estimation, because second moments and the likelihood can be computed in closed form.

It is conceptually straightforward to estimate linearized models using only aggregate data. Exploiting micro-level information in the estimation is less straightforward, in particular if these data are available at a lower frequency. Typically, aggregate data are quarterly, while household or firm data are available only annually. Time aggregation at the household level is not a simple exercise in the model, because the individual decisions can be highly nonlinear, especially when there is a discrete choice element. The existing literature uses micro data to calibrate the steady state of the model, while the properties of the fluctuations around the steady state are estimated with aggregate data, or perhaps with repeated cross-sections of data, but not exploiting the panel structure of existing data.

Our main contribution in this paper is to extend the linearization approach so as to include panel and mixed frequency data in the estimation. We handle the case where agents make both continuous and discrete choices. This is relevant both for households (labor participation, housing choice) as well as for firms (lumpy investment), but it raises specific problems in the linearization approach. The solution of this kind of models is explained in Reiter (2019), here we deal with estimation. Since the solution of heterogeneous agent models is still computationally costly, we pay special attention to the efficiency of the implementation.

The basic idea behind our approach is the following. We assume we have observations that contain information about a panel of individuals. We compute the theoretical counterpart of these observations in the model. For this purpose, we simulate a panel of individuals in the model over a given time period. The panel is initialized by the distribution of the stationary state without aggregate shocks. Each individual is hit by idiosyncratic shocks every period, but aggregate variables are always assumed to be at their steady state value. At the end of the sampling period, the observation is calculated by summing over the sampled individuals. In the stationary state, the panel observation is non-stochastic because of the law of large numbers. To compute the probability distribution of this observation in the model with aggregate shocks, which will give us the likelihood of the model parameters, we linearize the whole sampling procedure in the aggregate variables. An obstacle in this procedure is that the discrete choice of individuals and the stochastic transitions are not differentiable because of their discrete nature. We introduce smoothing operations such that the dependence on the aggregate variables is captured. We make the computation fast by using reverse-mode automatic differentiation.

This technique is useful not only for estimating a model, but more generally for the analysis of any properties of a model that are functions of individual histories. Such an analysis is made difficult by the fact that, in general, there is no straightforward way of simulating the trajectories of individual agents in a linearized model. Since individual decisions are nonlinear, but linearly approximated in aggregate variables, aggregate shocks of realistic size will cause the linearized decisions of a fraction of individuals to violate the model constraints, such as a borrowing constraint, or the requirement that probabilities are between 0 and 1. We circumvent this

problem by simulating individuals in the aggregate stationary state, and compute any variable of interest in the stationary state. We then differentiate this variable with respect to aggregate states and shocks.

We show with an example that the use of micro data can greatly improve parameter identification. Since the focus of this paper is methodological, we illustrate this with simulated data. The example model is based on Takahashi (2019), which features time-varying uncertainty, which is hard to observe directly in the data. For this exercise, we use a maximum likelihood approach, because it is the most straightforward method and very powerful when the model is correctly specified. However, our method can just as well be used in combination with limited-information estimation methods, which we think are more appropriate for real world applications. The problem of mis-specification, which probably affects any general equilibrium model, is likely to be even more severe when the whole cross-sectional distribution is modeled.

Despite the methodological advances in recent years, the full model solution procedure, including the computation of the stationary state, is indeed time consuming. To make estimation feasible for interesting models, it is useful to partition the set of parameters into those that affect the steady state (in particular the invariant cross-sectional distribution), and those that do not. In the latter category there are the parameters that govern the dynamics of the exogenous shock process, including the shock variances, as well as the specification of measurement errors. They also include policy parameters relating to stabilization policy, such as the Taylor rule. If the first set of parameters is pinned down (calibrated) through the steady state properties of the model, the estimation of the second set of parameters can be reasonably fast. If one wants to treat the general case where all parameters are estimated, it would be necessary to have some kind of hierarchical algorithm that changes more frequently those parameters that can be easily computed.

1.1 Literature on the estimation of heterogeneous agent models

Because of methodological challenges, only a small number of papers have considered estimating aspects of heterogeneous agent models, but due to methodological advances and the decreasing cost of computational power, this literature is expanding rapidly. We review related papers below, distinguishing methodologies by whether they only use "macro" data (aggregate observables), or allow the estimation to incorporate "micro" (cross-sectional) data.

Estimation using only macro data. Winberry (2016) and Winberry (2018) present a methodology for estimating heterogeneous agent models, also available as a toolbox for Dynare. In contrast to Reiter (2009), the cross-sectional distribution is approximated using a flexible parametric family following Algan, Allais, and Den Haan (2008).[1] The resulting equilibrium is solved for the steady state, then perturbed locally. The law of motion for the states, which include the parameters of the approximation, is then obtained from a first-order perturbation, and the likelihood can be formed using the Kalman filter. Conceptually, the methodology is similar to a perturbation solution of a DSGE model, now considered standard (Fernández-Villaverde et al. 2016), but the equilibrium conditions here are richer. This approach uses only aggregate data. The paper demonstrates the methodology with the model of Khan and Thomas (2008) with investment-specific shocks.

Hasumi and Iiboshi (2019) use the continuous-time solution methods of Achdou et al. (2017) and Ahn et al. (2018), a continuous-time implementation of the linear filtering approach used by the above papers. This requires Sequential Monte Carlo (SMC) methods to evaluate the likelihood. The method uses aggregate observables, particularly GDP growth, inflation, and interest rates, and no cross-sectional or panel data.

[1] Algan, Allais, Den Haan, and Rendahl (2014) also provide an introduction to the method.

Fernández-Villaverde et al. (2019) use a non-linear, continuous time-solution for heterogeneous agent models, approximating the perceived law of motion for aggregate states by a neural network. Given this law of motion, they use the Kolmogorov forward equation to compute the density of aggregate states at discrete points in time, conditional on their previous value. This allows estimation with ML or Bayesian methods, capturing inherent nonlinearity in the model. A scheme to incorporate cross-sectional likelihoods conditional on observed aggregate states is outlined in the appendix, but not implemented in the paper.

Auclert et al. (2019) show how to estimate a model using the "MIT-shock algorithm" of Boppart et al. (2018), and provide a very efficient implementation of this algorithm. If only parameters are used that do not change the steady state, and if only aggregate data are used for the estimation, a likelihood evaluation is in the range of milliseconds.

Estimation incorporating cross-sectional data. Mongey and Williams (2017) use a hybrid approach, relying on calibration, simulated method of moments, and Bayesian estimation to estimate a model which extends the fixed cost investment model of Khan and Thomas (2008) with a positive cost of raising external finance, which clusters firm investment in the state space, and an uncertainty shock like Gilchrist et al. (2014). The capital share and the depreciation rate are calibrated, various adjustment cost and idiosyncratic shock parameters are estimated using simulations without aggregate shocks to match corresponding cross-sectional data, and finally a linearized approximate solution is used for Bayesian estimation using the Kalman filter. Specifically, using the solution method of Reiter (2009), a linear approximation to the transition and observation equations is obtained, and this likelihood is used for an MCMC approach with the Metropolis-Hastings algorithm.[2] For this last part, the data consists of 4 aggregate time series: output, consumption, hours worked, and standard deviation of sales growth. Importantly, the latter is a cross-sectional moment which can be obtained as a function of the states (including the histogram approximating the distribution).

Closest to the goals of our paper is Liu and Plagborg-Møller (2019), which shows a method incorporating the likelihood of individual observed cross-sectional data, in addition to aggregate data, conditional on latent or observed states that characterize an approximation of the model. Following Andrieu et al. (2010) and Flury and Shephard (2011), they use an unbiased numerical estimate of the likelihood based draws of aggregate states from the Kalman smoother with the model estimated using the method of Reiter (2009). Importantly, this approach requires that the likelihood of observed individual data, for example consumption, is available, conditional on the state variables (which can include macro states like TFP or histogram approximations of the cross-section). Since the individual states are not observed either, this involves the computation of an approximate integral to obtain the distribution. While the methodology precludes the use of panel information, the paper demonstrates that the inclusion of cross-section data can sharpen posterior distributions.

Bayer, Born, et al. (2019) estimate a HANK model with both liquid and illiquid assets. Applying the linearization algorithm of Bayer and Luetticke (2018), they can use the Kalman filter for a Bayesian likelihood estimation. Their data include information about the cross-sectional income and wealth distribution, but no panel data.

M. Chang et al. (2018) use a functional VAR approach, modeling the aggregate state variables and a spline approximation to the log density of the cross-sections in a linear-Gaussian framework. This allows the use of the Kalman filter to estimate parameters of this reduced form model. They argue that the approach is analogous to using VAR estimation to evaluate representative agent models, and that inference about the off-diagonal terms is useful for quantifying the effect of aggregates on the cross-section and vice versa. Their model for the

[2] For a review of the application of the MH algorithm in macroeconomics, see Herbst and Schorfheide (2017)

observation equation allows cross-sectional, IID conditional on the aggregate state variables, but not panel data.

2 An Example Model

To illustrate our estimation method, we use a slight variation of the model of Takahashi (2019), which is basically the model of Y. Chang and Kim (2007) augmented by time-varying individual productivity. The key innovation of Y. Chang and Kim (2007) over standard heterogeneous agent models such as Krusell and Smith (1998), is the introduction of indivisible labor. This model is useful for our purpose for several reasons. First, it is very simple to write down, but involves an element of discrete choice (indivisible labor), which makes it harder to handle by linearization methods. We show in detail how to deal with the discrete choice element.[3] Second, it features time-varying uncertainty of idiosyncratic productivity, which is not easy to identify in the data. One of our aims is to show how the combination of micro- and macroeconomic evidence helps to identify the cyclical properties of time-varying uncertainty. Takahashi (2019), building on Y. Chang and Kim (2006), estimates the productivity process directly from the data, using PSID wage data. Since many big data sets have reliable information on wage income, but no or only noisy information on hourly wages, we think it is interesting to see whether panel information on household income can be used to recover the cyclical properties of idiosyncratic uncertainty. Of course, any such inference would be conditional on the economic model being correctly specified.

In this model, ex-ante identical and infinitely lived households are subject to shocks to their individual labor productivity. We assume that the logarithm of labor productivity is the sum of a persistent and a transitory component. The transitory part, denoted by ξ, is modeled as a uniformly distributed i.i.d. process, with constant mean and variance over time. The persistent part, denoted by x, is modeled as an AR(1) process, approximated by a finite-state Markov chain, similar to Y. Chang and Kim (2006) and Takahashi (2019). The grid of possible productivity levels is constant over time, but the transition probabilities between discrete states vary over time, governed by the aggregate variable ς_t.[4] This variable measures the conditional standard deviation of the individual Markov process of each household, and follows

$$\log(\varsigma_t - \bar{\sigma}) = \rho_\varsigma \log(\varsigma_{t-1} - \bar{\sigma}) + \varepsilon_{\varsigma,t} \tag{1}$$

We allow in general a non-zero covariance $\sigma_{z,\varsigma}$ between ε_ς and the technology innovation ϵ_z, but set $\sigma_{z,\varsigma} = 0$ in the benchmark.

Each period, the household chooses consumption and labor. Labor markets work frictionlessly, but labor is indivisible: a household can work either zero hours or a fixed number of hours $\bar{h}$. The household value function is then given by

$$V(a,x;z,\phi,\varsigma) = \int \max_{a',h\in\{0,\bar{h}\}} \left\{ u(c,h) + \beta \mathrm{E} V(a',x';z',\varsigma',\phi') \right\} \mathbf{d}F_\xi(\xi) \tag{2}$$

s.t.

$$c = w(z,\phi,\varsigma) x \xi h + (1 + r(z,\phi,\varsigma)) a) - a' \tag{3}$$

$$a' \geq \bar{a} \tag{4}$$

$$\phi' = T(z,\phi,\varsigma) \tag{5}$$

[3] How to solve heterogeneous agent models with mixed continuous and discrete choice is explained in Reiter (2019).

[4] We fix a grid $\hat{x}$ for the values of x. The transition matrix Π_t on $\hat{x}$ is then obtained as follows. We subdivide the model period (a quarter) into 16 subperiods. In a subperiod (roughly a week), the value of x is either unchanged or reaches one of the two neighboring states, such that the conditional expectation is ρx and the conditional standard deviation is given by ς_t. This gives a sub-period transition matrix $\tilde{\Pi}_t$. Then $\Pi_t = \tilde{\Pi}_t^{16}$.

where x is exogenous individual productivity, z is aggregate TFP, and ϕ is the cross-sectional distribution of agents over (a,x), and F_ξ is the cumulative probability distribution of the i.i.d. shock ξ. We assume the utility function $u(c,h)=\log(c)-Bh$.

Technology is standard. There is a representative firm with production function:

$$Y_t = F(L_t,K_t,z_t) = z_t L_t^{\alpha} K_t^{1-\alpha} \tag{6}$$

where TFP z_t follows the AR(1) process

$$\log z_t = \rho_z \log z_{t-1} + \epsilon_{z,t} \tag{7}$$

and the innovation has standard deviation σ_z. Factor prices equal marginal productivities. The model is closed by aggregating over individual labor and saving:

$$\begin{aligned} L_t &= \int\int x\xi h(a,x;z,\phi,\varsigma)\mathbf{d}F_\xi(\xi)\mathbf{d}\phi_{t-1}(a,x) \\ K_t &= \int a\mathbf{d}\phi_{t-1}(a,x) \end{aligned} \tag{8}$$

Notice that the capital stock at the beginning of period t is determined by ϕ_{t-1}, the cross-sectional distribution at the end of the previous period.

3 Model Solution and Likelihood

Estimation of the model is straightforward as long as only aggregate data are used, or snapshots of the cross-sectional distribution. Complications arise if the panel dimension comes into the picture, which is already the case if micro data are at a lower frequency, for example annual rather than quarterly, because this requires us to follow individuals over at least 4 quarters. In this section we describe the computations necessary for estimation. Section 3.1 gives a general description of the problem of the heterogeneous agent. Section 3.2 gives an outline over all the computational steps. Section 3.3 describes the state reduction technique, which was developed elsewhere but is underlying our estimation method here. Section 3.4 is the core of the paper, containing the steps of how to include panel data in the estimation. Section 3.5 gives details of the likelihood computations. In Section 3.6 we discuss the computational effort that is necessary for the different steps of the estimation. In particular, we show which elements of the computation are invariant to a change in the parameters that are estimated.

3.1 The agent problem and the cross-sectional distribution

We assume that the agent problem can be written in the following form. The agent has three state variables:

1. One endogenous continuous state variable (in the model of Section 2, end-of-period assets a').
2. An endogenous discrete state (none in the above model, but if we add labor market frictions, the employment state is an endogenous discrete state). Next period's discrete state is assumed to be a function of the discrete state and the discrete decision of the current period.
3. An exogenous discrete state (in the model of Section 2, labor productivity x).

The agent has two decision variables. First, a continuous choice variable, which we assume to be the end-of-period continuous state. Second, a discrete choice (work or not work).

Individual uncertainty comes in the form of a random change of discrete states between end of period and beginning of next period. This is modeled as a finite state Markov chain. Transition probabilities can depend on the aggregate state (time-varying uncertainty). In the computer code, we allow for idiosyncratic i.i.d. shocks on top of the Markov state, but we omit this here for notational simplicity, since an i.i.d. shock is a special case of an exogenous state.

For the value function we choose a fixed grid in the continuous state, and interpolate quadratically between grid points. For details, cf. Reiter (2019). To obtain a finite representation of the cross-sectional distribution, we partition the continuous state space into n_d histogram bins (the boundaries of the bins need not coincide with the grid for the value function). Denote by n_e the number of endogenous discrete states and by n_x the number of exogenous discrete states. A cross-sectional distribution is then characterized by a three-dimensional array of dimension $n_d \times n_e \times n_x$. Each element of the array gives the fraction of individuals with the corresponding discrete states, and with the continuous state being within the histogram bin. The array elements sum up to one.

3.2 Overview of computations

To compute the likelihood function, the following steps are necessary.

1. Computing the stationary state without aggregate shocks.
2. Differentiating the model equations at the steady state (Reiter 2009). This is done efficiently by forward-mode automatic differentiation.
3. Model reduction: reducing the dimensionality of the state vector and the value function vector, cf. Section 3.3.
4. Solving the linearized and reduced model, either by QZ-decomposition or by time iteration (Appendix B).
5. Simulating a panel of individuals (the model counterpart of the panel data), and differentiating the observables w.r.t. all aggregate variables (Section 3.4).
6. Further state reduction using principal component analysis (Section 3.4.5).
7. Computing the likelihood function (Section 3.5). This can be done either
   (a) in state space, as a function of a past state vector and the consequent shocks (Section 3.5.1)
   (b) in sequence space, as a function of the truncated history of shocks (Section 3.5.2).

The key contribution of this paper is to show how to do Step 5, for models including discrete choice, and how to do it fast.

3.3 Preliminaries: linearization and loss-less state reduction

After linearizing the model, we are left with a very large but relatively sparse system of equations. It is convenient to write the equation system in partitioned form. For this, we split the set of variables in the model into three categories:

1. the vector s of dimension n_s contains the state variables, typically all the variables that appear with time index $t-1$. In the model of Section 2, these are all the variables characterizing the cross-sectional distribution (the fraction of households within each histogram bin) as well as the two exogenous driving forces, technology and the conditional standard deviation of the idiosyncratic shock.

2. the vector v of dimension n_v contains all the variables that appear with time index $t+1$; in our model, it consists only of the value function at the grid points.

3. the vector y of dimension n_y contains all the other variables, in particular the macroeconomic aggregates such as GDP, aggregate labor etc.

Notice that we include the value function at grid points, but not the optimal choices as model variables. By differentiating the Bellman equation, the choices at time t are expressed as functions of the current variables y_t and the expected future variables $\mathrm{E}_t v_{t+1}$.[5]

Correspondingly, we split the equation systems into three types of equations:

1. The laws of motion for the state variables.

2. The forward looking equations for the variables v, in particular the linearized Bellman equations.

3. All the other equations.

The partitioned equation system can be written as

$$\begin{bmatrix} \Lambda_{ss} \\ \Lambda_{ys} \\ 0 \end{bmatrix} s_{t-1} + \begin{bmatrix} I & \Gamma_{sy} & \Gamma_{sv} \\ \Gamma_{ys} & \Gamma_{yy} & \Gamma_{yv} \\ 0 & \Gamma_{vy} & I \end{bmatrix} \begin{bmatrix} s_t \\ y_t \\ v_t \end{bmatrix} + \mathrm{E}_t \begin{bmatrix} \Phi_{sv} \\ \Phi_{yv} \\ \Phi_{vv} \end{bmatrix} v_{t+1} + \begin{bmatrix} \Psi_s \\ \Psi_y \\ 0 \end{bmatrix} \epsilon_t = 0 \tag{9}$$

Here we have used the fact that only the variables in s appear in lagged form, and only the variables in v appear in future form. For notational convenience we have assumed (as is typically the case), that the two diagonal blocks in the matrix that multiplies the time-t variables are identity matrices. It is no problem to handle more general, but sparse diagonal blocks Γ_{ss} and Γ_{vv}.

A loss-less (up to machine precision) optimal state and value function reduction (Reiter 2010, Reiter 2019) consists of two components:

- A matrix $\bar{M}$ with the following properties:
  - it has dimension $n_m \times n_s$ with $n_m < n_s$
  - there exists an $n_m \times n_m$ matrix $\hat{A}$ such that $\bar{M}\Lambda_{ss} = \hat{A}\bar{M}$, which reduces the dimension of the state vector
  - there exists an $n_y \times n_m$ matrix $\tilde{\Gamma}_{ys}$ such that $\Gamma_{ys} = \tilde{\Gamma}_{ys}\bar{M}$.

The reduced state vector m is then a linear function of the full state s:

$$m_t = \bar{M} s_t \tag{10}$$

[5] Things are simpler in a model where agents do not have a discrete choice but only solve a convex optimization problem. Then there is no need to include the value function, the equation system just includes the policy variables and differentiates the Euler equations.

- An $n_v \times n_f$-matrix $\bar{V}$ with dimension $n_f < n_v$ that spans the space in which the value function lives, so that we can write

$$v_t = \bar{V} f_t \tag{11}$$

This matrix can be obtained by forwarding the third block of the equation system many times, and performing a singular value decomposition of all the columns of the form $\Phi_{vv}^i \Gamma_{vy}$. We can choose $\bar{V}$ as orthonormal so that $\bar{V}'\bar{V} = I$.

Premultiplying the first block of equations with $\bar{M}$ and the third block with $\bar{V}'$, and substituting (11), we get the reduced linearized model:

$$\begin{bmatrix} \hat{A} \\ \tilde{\Lambda}_{ys} \\ 0 \end{bmatrix} m_{t-1} + \begin{bmatrix} I & \bar{M}\Gamma_{sy} & \bar{M}\Gamma_{sv}\bar{V} \\ \tilde{\Gamma}_{ys} & \Gamma_{yy} & \Gamma_{yv}\bar{V} \\ 0 & \bar{V}'\Gamma_{vy} & I \end{bmatrix} \begin{bmatrix} m_t \\ y_t \\ f_t \end{bmatrix} + \mathrm{E}_t \begin{bmatrix} \bar{M}\Phi_{sv}\bar{V} \\ \Phi_{yv}\bar{V} \\ \bar{V}'\Phi_{vv}\bar{V} \end{bmatrix} f_{t+1} + \begin{bmatrix} \bar{M}\Psi_s \\ \Psi_y \\ 0 \end{bmatrix} \epsilon_t = 0 \tag{12}$$

What is key for the speed of the estimation is that the matrices $\bar{M}$ and $\bar{V}$ do not depend on the estimated parameters. To achieve this, we partition the state vector into the part that describes the cross-sectional distribution (big), and the part that contains all the other variables (small). We select the matrix $\bar{M}$ such that the second set of variables are explicitly contained:

$$\bar{M} = \begin{bmatrix} \bar{M}_{1,1} & 0 \\ 0 & I \end{bmatrix}$$

The equation $\bar{M}\Lambda_{ss} = \hat{A}\bar{M}$, can then be written as

$$\begin{bmatrix} \bar{M}_{1,1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \Lambda_{ss,1,1} & \Lambda_{ss,1,2} \\ \Lambda_{ss,2,1} & \Lambda_{ss,2,2} \end{bmatrix} = \begin{bmatrix} \hat{A}_{1,1} & \hat{A}_{1,2} \\ \hat{A}_{2,1} & \hat{A}_{2,2} \end{bmatrix} \begin{bmatrix} \bar{M}_{1,1} & 0 \\ 0 & I \end{bmatrix} \tag{13}$$

We require that $\Lambda_{ss,1,1}$ and $\Lambda_{ss,2,1}$ is not affected by the parameter change (typically $\Lambda_{ss,2,1} = 0$). Then $\hat{A}_{1,1}$ and $\hat{A}_{2,1}$ are unchanged, and $\hat{A}_{1,2} = \bar{M}_{1,1}\Lambda_{ss,1,2}$ and $\hat{A}_{2,2} = \Lambda_{ss,2,2}$.

Examples of parameters that do not affect $\Lambda_{ss,1,1}$ and $\Lambda_{ss,2,1}$ are the parameters of the exogenous driving forces, or policy parameters, such as the coefficients of the Taylor rule. They do not change the way how the policy variables affect individual choices, therefore they do not change Γ_{vy} either. Parameters that affect the dynamics of the cross-sectional distribution in Λ_{ss}, or of the forward looking part of the Bellman equation in Φ_{vv}, would usually also change the steady state, so that everything needs to be recomputed.

3.4 Using panel data for estimation

3.4.1 A model for panel data

A typical situation is that the model is solved and simulated at quarterly frequency, but for individual income, consumption, etc., only annual data are available. To compute the model's implications for individual annual income, one has to follow an individual in the model for four quarters. To talk about the autocorrelation of annual income, one needs at least eight quarters. For concreteness, we take below the case of 8 quarters. Generalization to longer panels is straightforward.

Let us be clear that we do not use the full information contained in the panel data. In particular, we do not compute the likelihood of the full panel of individuals. We rather use a small number of summary statistics of the panel. Concretely, we assume that the vector of observed variables of an individual in period τ of the

panel is of the form $\mathcal{C}(k,e,x,k',d;\Omega,\tau)$. It is a function of its continuous state k, endogenous discrete state e, exogenous discrete state x, discrete choice d and the aggregate state Ω. The latter contains the cross-sectional distribution, ϕ, the exogenous driving processes and potentially some endogenous aggregate states. By making $\mathcal{C}$ a function of τ as well, we can distinguish, for example, between first year income and second year income and compute the autocorrelation. What is really observed is the sum of these variables over the panel length A (eight quarters in our example):

$$C_t = \sum_{\tau=-A+1}^{0} \mathcal{C}\left(k_{t+\tau},e_{t+\tau},x_{t+\tau},k'_{t+\tau},d_{t+\tau};\Omega_{t+\tau},\tau\right) \tag{14}$$

Notice that the aggregate state $\Omega_{t+\tau}$ can be written as a function of the initial state Ω_{t-A} and the aggregate shocks, $z_{t-A+1},\dots,z_{t+\tau}$. Similarly, given an individual policy function, and conditional on aggregate states, the individual states and decisions at time $t+\tau$ can be expressed as a function of the initial state $\omega_{t-A} \equiv (k_{t-A},e_{t-A},x_{t-A})$ and the idiosyncratic shocks $\varepsilon_{t-A+1},\dots,\varepsilon_{t+\tau}$. We therefore can express the observation of an individual starting at state ω as $C(\omega,\Omega_{t-A},\varepsilon_{t-A+1},z_{t-A+1},\dots,\varepsilon_t,z_t)$. We assume that the aggregate observation is the sample average of a function $\mathcal{S}$ of the variable C that has been accumulated over A time periods.

$$\hat{C}_t = \int \mathcal{S}(C(\omega,\Omega_{t-A},\varepsilon_{t-A+1},z_{t-A+1},\dots,\varepsilon_t,z_t))d\mu(\varepsilon_{t-A+1})\dots d\mu(\varepsilon_t)d\phi_{t-A}(\omega) \tag{15}$$

The function $\mathcal{S}$ allows us to compute higher-order moments, autocorrelations, etc., of the individual observations. We could generalize (15) by making $\mathcal{S}$ a function of more variables, but we want to keep notation concise at this point.

By integrating out the effect of the idiosyncratic shocks, the realization of $\hat{C}_t$ can be written as a function of ϕ_{t-A} and the aggregate shocks $z_{t-A+1},\dots,z_t$. Our strategy is to approximate the integrals over idiosyncratic shocks and the cross-sectional distribution in (15) by a pseudo-random sample of N households. To generate idiosyncratic shocks for all individuals, we draw random numbers $\varepsilon_{i,\tau}$ for $\tau = 1,\dots,A$ and i from 1 up to the maximum number of simulated individuals.

Why do we have to simulate? In each period, the agent faces a shock to its exogenous discrete state, which is described by a finite Markov chain.[6] If an agent can transit to n possible states in the next period, there is a maximum of n^8 different histories over the two-year sample period, conditional on the agent's initial state and the path of aggregate variables. Analytical computation of the properties of this path are in general not possible because of the nonlinear nature of the individual decision. If $n=2$, one might simulate the 2^8 possible histories. For higher n, this is clearly infeasible. We therefore have to resort to simulation techniques.

Notice that the procedure that we propose here does not require to simulate individuals outside the aggregate stationary state. As we have explained in the introduction, such a procedure would run into serious difficulties, for the following reason. The individual decision problem is solved linearly, but the dependence of the solution on aggregate variables is linearized. If a decision happens with a probability that depends on aggregate states, outside the steady state this probability is not guaranteed to be between zero and one. If agents in the simulation are given a weight, as is done in the procedure above, the weights are not guaranteed to be between 0 and 1. To avoid these problems by simulating agents in the aggregate steady state, and differentiate the variables of interest in the aggregate states.

What remains to be done is, first, to determine a concrete procedure of how to sample, and second, to approximate the observation as a linear function of the aggregate state, in a computationally efficient way.

[6] Remember from Section 3.1 that we subsume the i.i.d. shock in the Markov chain for notational simplicity.

3.4.2 Smoothing the individual transitions

A key problem that our procedure has to overcome is to capture the derivatives of discrete decisions and transitions w.r.t. aggregate variables. These derivatives are zero almost everywhere, and are therefore zero in a finite sample. To solve this, we smooth both the exogenous and endogenous transitions in a way that is differentiable in the aggregate states. In the exogenous transition phase (realization of the Markov process) we hit the agent by an interval of a continuous shock, which then might lead to a split of this agent into several successor agents with properly adjusted weights. In the endogenous transition phase, if an agent is very close to a threshold state where the discrete decision switches, we split the agent into two successor agents, capturing the two possible discrete decisions. In both cases, the effect of aggregate variables is captured in the weights of the successor agents, and this effect is differentiable.

3.4.3 Computing the panel data

We first describe the finite-sample approximation of $\hat{C}_t$ in (15) as a function of the initial distribution ϕ_{t-A}, and of individual policy functions and individual state transition functions at periods $\tau = t-A+1$ to $\tau = t$. Afterwards we will show how to linearize this function in the variables $\phi_{t-A}, z_{t-A+1}, \ldots, z_t$.

The simulation has three phases:

1. Initialization
2. Iteration
3. Aggregation

Initialization

We choose a target number N^{Target} of initial individuals (agents). Then we define a sample of individual states $\omega_{a,t+\tau}$ and weights $\hat{\phi}_a$ as follows. In each bin i of the cross-sectional distribution (consists of an exogenous and endogenous discrete state, and a histogram bin of the continuous state), place $N_i = \max\{1, \text{round}(N^T$ individuals, where $\phi_{i,t-A}$ is the fraction of individuals in bin i at the end of period $t-A$. Each individual a in bin i has the weight $\hat{\phi}_a = \phi_{i,t-A}/N_i$. The individual state $\omega_{a,t+\tau}$ is such that the continuous state is equispaced over the histogram bin (if $N_i = 1$, choose the midpoint, etc.). Set the cumulative observation to $c_{a,t+\tau} = 0$ for all a. The initial number of individuals is $N = \sum_i N_i$.

Iteration

Denote by $\mathcal{A}_\tau$ the set of all agents in the panel at the beginning of period τ. Then, for $\tau = -A+1, \ldots, 0$ and for each $a \in \mathcal{A}_\tau$, do the following.

1. Exogenous transition.

   For any agent a, denote by i its exogenous discrete state at the end of the last period. Denote by π^i the n_x-vector of conditional probabilities of reaching the n_x different exogenous discrete states, and by Π^i the cumulative sum of π^i. We can simulate the Markov process by hitting the agent with a uniformly distributed shock ε, and stipulate that she transits to state $j(\varepsilon) = \mathcal{I}(\Pi^i, \varepsilon)$, where $\mathcal{I}(x, \varepsilon)$ is defined for any vector x as the smallest integer j such that $x_j \geq \varepsilon$. This is consistent in the sense that $\int_0^1 j(\varepsilon) d\varepsilon = \pi^i_j$ for all j. However, drawing a finite sample $\varepsilon_1, \ldots, \varepsilon_n$, the derivative of the sample mean of any function

$f(j)$, $\frac{1}{n}\frac{\partial\sum_{l=1}^{n}f(j(\varepsilon_l))}{\partial\Pi_j^i}$, equals zero almost surely. In other words, the dependence of the sample on the transition probabilities cannot be obtained by differentiation. To solve this problem, we do not sample point realizations of ε, but an interval of shocks $(\varepsilon-\bar{\varepsilon},\varepsilon+\bar{\varepsilon})$ with width $2\bar{\varepsilon}$.

To allow for the possibility of $\varepsilon-\bar{\varepsilon}<0$ or $\varepsilon+\bar{\varepsilon}>1$, we define Φ^i as the cumulative sum of $[\pi^i;\pi^i;\pi^i]$). If $j=\mathcal{I}(\Phi^i,1+\varepsilon-\bar{\varepsilon})=\mathcal{I}(1+\Phi^i,\varepsilon+\bar{\varepsilon})$, we say that the agent transits to state $\widetilde{j}$ with probability one, where $\widetilde{j}$ denotes j modulus the length of Π^i, more precisely $\widetilde{j}\equiv\mathrm{mod}(j-1,n_x)+1$. The weight $\hat{\phi}_a$ of the agent is then unchanged.

If this is not the case, we will split the agent into several successors with adjusted weights. Defining $j_L\equiv\mathcal{I}(\Phi^i,1+\varepsilon-\bar{\varepsilon})$ and $j_U\equiv\mathcal{I}(\Phi^i,1+\varepsilon+\bar{\varepsilon})$, the agent is replaced by j_U-j_L+1 successors, going to states $j\in\{\widetilde{j_L},\widetilde{j_L+1},...,\widetilde{j_U-1},\widetilde{j_U}\}$ with probabilities

$$\hat{\phi}(\widetilde{j},\varepsilon)=\begin{cases}\frac{1}{2\bar{\varepsilon}}\left(\Phi^i_{j_L+1}-\varepsilon+\bar{\varepsilon}\right) & \text{if } j=j_L\\ \frac{1}{2\bar{\varepsilon}}\left(\varepsilon+\bar{\varepsilon}-\Phi^i_{j_U-1}\right) & \text{if } j=j_U\\ \frac{1}{2\bar{\varepsilon}}\left(\Phi^i_{j+1}-\Phi^i_j\right) & \text{if } j_L<j<j_U\end{cases} \tag{16}$$

A successor of agent a transiting to the discrete exogenous state j, inherits from a the continuous and discrete endogenous state, as well as the cumulate observed value $c_{a,t+\tau}$. It gets the weight $\hat{\psi}_a$ multiplied with the transition probability $\hat{\phi}(j,\varepsilon)$ defined in (16). The total number of simulated agents is therefore increased, but the sum of the weights of all agents is kept unchanged.

The following properties show that this is an unbiased sampling scheme both for the level and the derivative of the transition probabilities.

- In expectation, the weight of a successor with state j equals the transition probability, i.e., $\int_0^1\hat{\phi}(j,\varepsilon)d\varepsilon=\pi^i_j$.
- The successor weights $\hat{\phi}(l,\varepsilon)$ are differentiable with respect to the transition probabilities, and their derivatives coincide on average with the change in probabilities:

$$\frac{\partial\int_0^1\hat{\phi}(l,\varepsilon)d\varepsilon}{\partial\Pi^i_j}=\int_0^1\frac{\partial\hat{\phi}(l,\varepsilon)}{\partial\Pi^i_j}d\varepsilon=\begin{cases}1 & \text{if } l=j\\ -1 & \text{if } l=j+1\\ 0 & \text{else}\end{cases}=\frac{\partial\pi^i_l}{\partial\Pi^i_j} \tag{17}$$

To which value should $\bar{\varepsilon}$ be set? A very small $\bar{\varepsilon}$ implies that most agents transit to a single successor state with probability one, so that $\bar{\varepsilon}$ affects few individuals, increasing the sampling error. A large $\bar{\varepsilon}$ implies a fast increasing number of agents, which is computationally costly. The right choice of $\bar{\varepsilon}$ requires some experimentation.

2. Endogenous transition.

Case 1: No Threshold

Assume that agent a enters the period with continuous state k within a histogram bin (k_0,k_1) over which the discrete choice stays constant: $\mathcal{D}(k_0,e,x;\Omega)=\mathcal{D}(k_1,e,x;\Omega)$.

We then assume that the choice function $\mathcal{K}(k,e,x;\Omega)$ is continuous over (k_0,k_1), and approximate the choice by linear interpolation as

$$k'=\frac{k_1-k}{k_1-k_0}\mathcal{K}(k_0,e,x;\Omega)+\frac{k-k_0}{k_1-k_0}\mathcal{K}(k_1,e,x;\Omega)$$

This agent has then one successor with unchanged weight, new continuous state k' and the cumulative observation c_a incremented by $\mathcal{C}(k,e,x,k',\mathcal{D}(k_0,e,x;\Omega))$.

Case 2: Threshold Bin

Assume that agent a enters the period with continuous state k within a histogram bin (k_0,k_1) over which the discrete choice changes: $\mathcal{D}(k_0,e,x;\Omega) \neq \mathcal{D}(k_1,e,x;\Omega)$.

Then there is a threshold point k^{Thr} in this interval at which the agent is indifferent between discrete choices $\mathcal{D}(k_0,e,x;\Omega)$ and $\mathcal{D}(k_1,e,x;\Omega)$. The choice function $\mathcal{K}(k,e,x;\Omega)$ is in general discontinuous at k^{Thr}. Denote by $\mathcal{K}\left(k_-^{Thr},e,x;\Omega\right)$ and $\mathcal{K}\left(k_+^{Thr},e,x;\Omega\right)$ the left and right limit, respectively.

In comparison to Case 1, two changes are necessary. First, we must account for the discontinuity when interpolating the choice function $\mathcal{K}(k,e,x;\Omega)$. Second, to capture the effect of the aggregate state on the threshold point k^{Thr}, we potentially split the agent into two agents, one with a state k left of the threshold, and one with a state right of the threshold, so that the weight of the two successor agents reflects the effect of the aggregate state.

It is therefore convenient to subdivide the endogenous transition into two stages. In the first stage, we replace the agent with state k and weight $\hat{\phi}_a$ into one agent with state k_L and weight $\hat{\phi}_a p_L$ and another agent with state k_R and weight $\hat{\phi}_a(1-p_L)$ where p_L denotes the weight of the agent left of the threshold. In the second stage, we interpolate the agent's choice function and determine the transition to a new endogenous state.

(a) Case 2, Stage 1

More precisely, there are two successors if k is in the interval $(k^{Thr}-\zeta,k^{Thr}+\zeta)$, where ζ is chosen as $\zeta = \min\left\{k^{Thr}-k_0,k_1-k^{Thr}\right\}$. This is the largest interval symmetric around k^{Thr} that is contained in the bin (k_0,k_1).

This gives four subcases for the first stage.

Case 2a: $k \leq k^{Thr}-\zeta$

In this case, $k_L = k$ and $p_L = 1$. The agent has one successor with unchanged weight.

Case 2b: $k \geq k^{Thr}+\zeta$

Here $k_R = k$ and $p_L = 0$, and the agent has one successor with unchanged weight.

Case 2c: $k^{Thr}-\zeta < k \leq k^{Thr}$

This agent has two successors. The successor left of the threshold has state $k_L = k - (k^{Thr} - k)\frac{\zeta-(k^{Thr}-k)}{\zeta+k^{Thr}-k}$ and weight $\hat{\phi}_a p_L$ where $p_L = \frac{\zeta+k^{Thr}-k}{2\zeta}$. The successor right of the threshold has state k^{Thr} and and weight $\hat{\phi}_a(1-p_L)$. This case is illustrated in Figure 1.

Case 2d: $k^{Thr} < k \leq k^{Thr}+\zeta$

This agent has two successors. The successor left of the threshold has state $k_L = k^{Thr}$ and weight $\hat{\phi}_a p_L$ where $p_L = \frac{k^{Thr}+\zeta-k}{2\zeta}$. The successor right of the threshold has state $k-(k^{Thr}-k)\frac{\zeta-(k^{Thr}-k)}{\zeta+k^{Thr}-k}$ and weight $\hat{\phi}_a(1-p_L)$.

Notice the following two properties of this transition. First, aggregate capital is preserved: $k_L p_L + k_R(1-p_L) = k$. Second, on average, the probability of being below the treshhold equals that share of the histogram bin below the threshold: $\int_{k_0}^{k_1} p_L(k)dk = \frac{k^{Thr}-k_0}{k_1-k_0}$.

(b) Case 2, Stage 2

In the second stage, which is the endogenous change of the continuous state, we only have to distinguish whether the new k of the agent is left or right of the threshold. Assume that, after adjustment of k in the first stage, the agent has $k \leq k^{Thr}$, then the new continuous state k' is defined by the linear interpolation

$$k' = \frac{k^{Thr}-k}{k^{Thr}-k_0}\mathcal{K}(k_0,e,x;\Omega) + \frac{k-k_0}{k^{Thr}-k_0}\mathcal{K}\left(k_-^{Thr},e,x;\Omega\right) \tag{18}$$

and the cumulative observation c_a incremented by $\mathcal{C}(k,e,x,k',\mathcal{D}(k_0,e,x;\Omega))$.

If the agent has $k > k^{Thr}$, then the new continuous state k' is defined as

$$k' = \frac{k_1-k}{k_1-k^{Thr}}\mathcal{K}\left(k_+^{Thr},e,x;\Omega\right) + \frac{k-k^{Thr}}{k_1-k^{Thr}}\mathcal{K}(k_1,e,x;\Omega) \tag{19}$$

and the cumulative observation c_a incremented by $\mathcal{C}(k,e,x,k',\mathcal{D}(k_1,e,x;\Omega))$.

For the choice of ζ, in principle there is the same trade-off as with $\bar{\varepsilon}$, which we have discussed above. However, since it only affects agents that are in the threshold bin where the discrete choice changes, there tend to be too few agents who are split up, this is why we have set ζ to its maximal value. In a different model other choices might be better. In our specific example, it turns out that more splittings are desirable, therefore we apply the following oversampling procedure.

Oversampling

Only agents who end up in a bin where the discrete decision switches will be split in the endogenous transition. This may lead to few splits that cover the effects of the threshold, and leave a relatively large sampling error. One can generate more endogenous splittings and reduce sampling error by the following oversampling technique. In the threshold step, do the following for each agent a with weight $\hat{\phi}_a$ above some threshold $\bar{\phi}_0$:

(a) Generate k'_j for $j = 1,\ldots,n$ individual shocks.

(b) If k'_j leads to splitting, keep the split individuals, divide their weights $\phi_{a,j,t+\tau+1}$ by n. Denote the number of shocks that lead to splitting by n_a.

(c) If $n_a < n$, choose randomly one of the k'_j that do not lead to splitting, and give the agent the weight $\hat{\phi}_a \cdot \frac{n-n_a}{n}$.

For agents with weight $\hat{\phi}_a < \bar{\phi}_0$, just draw one shock.

When solving the model of Section 2, we apply a special case of this procedure. In our parameterization of the model, an agent receives an i.i.d. shock to productivity with 11 possible realizations in each period. We try all possible realizations, and keep those that lead to a splitting of the agent.

Aggregation of individual observations

After iterating forward, the finite-sample equivalent of (15) is computed as

$$\hat{C}_t = \sum_{a \in \mathcal{A}_\tau} \hat{\phi}_a \mathcal{S}(c_{a,t+\tau}) \tag{20}$$

3.4.4 Differentiating the panel data

The computation of $\hat{C}_t$ is done by a simulation of the individuals at the aggregate stationary state. In line with the linearization methodology, we will numerically compute the linearization of $\hat{C}_t$ at the stationary state w.r.t. all aggregate variables. This could be a computationally very expensive procedure. To make the computation efficient, a key assumption is that the dimension of the dependent variable $\hat{C}$ is much smaller than the number of independent variables (reduced state vector in state space representation, or truncated history of shocks in the sequence representation).[7] In such a case, the best way to differentiate is by reverse-mode automatic differentiation, exploiting the "cheap gradient principle" (Griewank and Walther 2008, Section 3.3). Traversing the whole chain of operations in Section 3.4.3 backward, we keep track of the low-dimensional gradient $\frac{\partial \hat{C}_t}{\partial x}$ for every intermediate variable x in this calculation. In contrast, the forward mode has to keep track of the gradient of $\hat{C}_t$ w.r.t. the large vector of independent variables, which includes a representation of the cross-sectional distribution at the beginning of the sample. Applying the reverse mode, the algorithm consists of a forward sweep and a return sweep. As long as the steady state does not change, all derivatives are taken at the same point, and therefore the forward sweep is unchanged, because it only records the intermediate results at the path where the derivatives are taken. (Any parameter change affecting the individual dynamics for given idiosyncratic shock sequence would effectively change the steady state.) The downside of the reverse mode is that it is harder to implement, because all the intermediate results of the forward sweep of the computation have to be stored, or recomputed in the return sweep. With very long calculations, memory limitations may arise. To economize on memory, we have hand-coded the reverse algorithm.

We next give an outline of the differentiation process. At any point in time, an agent in the sample has three differentiable characteristics: its continuous state, its weight in the sample, and the accumulated observation. The return sweep of the differentiation is initialized in the last period ($\tau = 0$) by the following formulas, obtained by differentiating (20) for all agents $a \in \mathcal{A}_\tau$:

$$\begin{aligned}
\frac{\partial \hat{C}_t}{\partial \hat{\phi}_a} &= \mathcal{S}(c_{a,t+\tau}) \\
\frac{\partial \hat{C}_t}{\partial \omega_{a,t+\tau}} &= 0 \\
\frac{\partial \hat{C}_t}{\partial c_{a,t+\tau}} &= \hat{\phi}_a \mathcal{S}_{c,t+\tau}(c_{a,t+\tau})
\end{aligned}$$

Starting from there, we iterate backwards in time, for $\tau = 0, \ldots, -A+1$. In each iteration step we do the following:

1. We compute the derivatives $\frac{\partial \hat{C}_t}{\partial \tilde{y}}$ for all aggregate variables $\tilde{y} \in Y_{t+\tau}$ and the derivatives $\frac{\partial \hat{C}_t}{\partial \tilde{p}}$ for all policy variables $\tilde{p} \in P_{t+\tau}$. Here, $P_{t+\tau}$ denotes all differentiable elements of the policy function, i.e., the continuous choice at grid points as well as the position of the switch points between different discrete decisions. $Y_{t+\tau}$ denotes the set of all current aggregate variables. Remember that the cross-sectional distribution affects agents not directly, but only through the variables contained in $Y_{t+\tau}$. These derivatives capture the following effects, summed over all agents $a \in \mathcal{A}_{t+\tau}$ in the sample:

    - the effect of the policy function on the continuous choice $\mathcal{K}(k,e,x;\Omega)$, obtained by piecewise linear interpolation;

[7] If one wants to deviate from this assumption and extract high-dimensional information from the panel of observations, one does not only have to live with a higher computational complexity, but also has to deal seriously with stochastic singularity, because the number of the aggregate shocks in the model would be much smaller than the number of observations.

- the effect on the threshold points k^{Thr};
- the effect on the exogenous transition probabilities π^i;
- a direct effect on the observed values c.

These calculations just differentiate the mechanics of the transition processes described in Section 3.4.3, using the continuation values $\frac{\partial \hat{C}_t}{\partial \phi_{a',t+\tau+1}}$, $\frac{\partial \hat{C}_t}{\partial \omega_{a',t+\tau+1}}$ and $\frac{\partial \hat{C}_t}{\partial c_{a',t+\tau+1}}$ of all successor agent a' of any agents $a \in \mathcal{A}_\tau$. In this process, we also update the derivatives $\frac{\partial \hat{C}_t}{\partial \hat{\phi}_a}$, $\frac{\partial \hat{C}_t}{\partial \omega_{a,t+\tau}}$ and $\frac{\partial \hat{C}_t}{\partial c_{a,t+\tau}}$.

The number of operations in this step is a relatively small multiple of the number of agents in the panel. It does not depend on the size of the policy grid, because the policy at a grid point only affects the agents with a state in the interval adjacent to the grid point (linear interpolation).

2. The policy variables are not part of the linear equation system (9), but each $\tilde{p} \in P_{t+\tau}$ is obtained as a linear function of $Y_{t+\tau}$ and $\mathrm{E}_t V_{t+\tau+1}$ in the process of linearizing the model equations, in particular the Bellman equation. To express the derivatives in terms of the model variables, we compute the effect of any value variable $\tilde{v} \in \mathrm{E}_t V_{t+\tau+1}$ as

$$\frac{\partial \hat{C}}{\partial \tilde{v}} = \sum_{\tilde{p} \in P_{t+\tau}} \frac{\partial \hat{C}}{\partial \tilde{p}} \frac{\partial \tilde{p}}{\partial \tilde{v}} \tag{21}$$

This computation appears to be very costly because $\mathrm{E}_t V_{t+\tau+1}$ is high-dimensional, but because of the sparsity of individual state transitions, each $\tilde{p}$ depends on few elements of $\mathrm{E}_t V_{t+\tau+1}$ only.

Then we take into account that the effect of any aggregate variable $\tilde{y}$ consists of the direct effect, computed in Step 1 above, and of the indirect effect through the policy function. In obvious notation we write this as

$$\frac{\partial \hat{C}}{\partial \tilde{y}} \mathrel{+}= \sum_{\tilde{p} \in P_{t+\tau}} \frac{\partial \hat{C}}{\partial \tilde{p}} \frac{\partial \tilde{p}}{\partial \tilde{v}} \tag{22}$$

for all $\tilde{y} \in Y_{t+\tau}$.

The number of operations in this step is dominated by the number of policy variables (size of policy grid) times the number of elements in $\mathrm{E}_t V_{t+\tau+1}$ on which each policy variables, on average, depends. The latter is not determined by the size of the grid, but by the sparsity of the transition process: if more future states are reached with positive probability in one step, more elements of $\mathrm{E}_t V_{t+\tau+1}$ are relevant today.

3. In the last step, we exploit not just the model equations, but the solution of the linearized model to compute the derivatives with respect to the state variables. From the solution of the linear model, we obtain the elements of $Y_{t+\tau}$ and $\mathrm{E}_t V_{t+\tau+1}$ as linear functions of the reduced state vector $m_{t+\tau-1}$ and the current shocks $\epsilon_{t+\tau}$: Then write

$$\frac{\partial \hat{C}}{\partial \tilde{m}} = \sum_{\tilde{y} \in Y_{t+\tau}} \frac{\partial \hat{C}}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial \tilde{m}} + \sum_{\tilde{v} \in \mathrm{E}_t V_{t+\tau+1}} \frac{\partial \hat{C}}{\partial \tilde{v}} \frac{\partial \tilde{v}}{\partial \tilde{m}} \tag{23}$$

$$\frac{\partial \hat{C}}{\partial \tilde{\epsilon}} = \sum_{\tilde{y} \in Y_{t+\tau}} \frac{\partial \hat{C}}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial \tilde{\epsilon}} + \sum_{\tilde{v} \in \mathrm{E}_t V_{t+\tau+1}} \frac{\partial \hat{C}}{\partial \tilde{v}} \frac{\partial \tilde{v}}{\partial \tilde{\epsilon}} \tag{24}$$

We collect these derivatives in the gradients $\gamma_{m,\tau}$ and $\gamma_{z,\tau}$.

We finish the return sweep by computing the impact of the initial distribution $\gamma_\phi \phi_{t-T}$. From the initialization step in Section 3.4.3, we get

$$\frac{\partial \hat{C}_t}{\partial \phi_{i,t-A}} = \frac{1}{N_i} \sum_{j=1}^{N_j} \frac{\partial \hat{C}_t}{\partial \hat{\phi}_{a_{i,j}}} \tag{25}$$

where $a_{i,j}$ runs over the N_i initial agents in bin $\phi_{i,t-A}$. We collect these derivatives in the gradient γ_ϕ. From the above steps, we obtain the following linearized expression for $\hat{C}_t$:

$$\hat{C}_t = \gamma_\phi \phi_{t-T} + \sum_{\tau=-T+1}^{0} \left[\gamma_{m,\tau} m_{t+\tau-1} + \gamma_{z,\tau} \epsilon_{t+\tau}\right] \tag{26}$$

Repeatedly applying the solution of the reduced model (12), which can be written as

$$m_t = A m_{t-1} + B \epsilon_t \tag{27}$$

we can bring (26) into the form

$$\hat{C}_t = \gamma_\phi \phi_{t-T} + \hat{\gamma}_m m_{t-T} + \sum_{\tau=-T+1}^{0} \hat{\gamma}_{z,\tau} \epsilon_{t+\tau} \tag{28}$$

At this stage, we have expressed all calculations in terms of the shocks and the reduced state vector m, except for the dependence on the cross-sectional distribution at the beginning of the panel, $\gamma_\phi \phi_{t-T}$. One can simply project γ_ϕ onto the rows of $\bar{M}$, so we replace $\gamma_\phi \phi_{t-T}$ by $\tilde{\gamma_\phi} m_{t-T}$, where $\tilde{\gamma_\phi}$ are the coefficients of the projection. However, there is in general no reason to believe that γ_ϕ is spanned by $\bar{M}$ with sufficient accuracy.[8] A more precise, even if computationally more involved, procedure is to reduce the state space even further. This step is also necessary if we want to compute the likelihood by means of the Kalman filter.

3.4.5 Further reduction of state space by principal component analysis

Define Φ as an orthonormal matrix with columns given by the first d principal components of the space in which the solution of the model lives.[9] This matrix can be found by a singular value decomposition of the impulse responses of the cross-sectional distribution to each shock. We then express the state vector as the linear superposition of the first d principal components: $\phi_t = \sum_{i=1}^{d} \Phi_i \lambda_i$. We need to find the VAR representation of the model dynamics in terms of the PCA representation:

$$\lambda_t = A_{PCA} \lambda_{t-1} + B_{PCA} z_t \tag{29}$$

To get A_{PCA}, denote by Φ_{+1} the matrix where column i is obtained by simulating the original model, starting from the state vector given by column i of Φ, with $z = 0$.[10] Since Φ spans the solution space, there must be an $d \times d$ coefficient matrix A_{PCA} such that $\Phi_{+1} = \Phi A_{PCA}$ with high numerical precision. Since Φ is orthonormal, $\Phi' \Phi_{+1} = A_{PCA}$. Since A_{PCA} gives the simulation result starting from Φ, it is the matrix that we need in

[8] One could redo the model reduction and recompute $\bar{M}$ so that it does span γ_ϕ, but this would be the computationally most costly solution.

[9] More precisely, we perform the PCA of all aggregate states except the exogenous driving processes. Φ then contains the principal components of these states, as well as the vectors that select each of the exogenous driving forces. This is to make sure that each of these exogenous variables is exactly represented in the finite representation.

[10] The method to obtain a precise simulation from the original model is described in Reiter (2019).

(29). Similarly, denote by $\hat{\Psi}_{+1}$ the matrix where column i is obtained by simulating the original model, starting from a zero state vector, and a shock vector z where the i-th component is one, and the rest is zero. Then $B_{PCA} = \Phi' \hat{\Psi}_{+1}$.

Using this representation, we can replace the term $\gamma_\phi \phi_{t-T}$ in (26) by $\gamma_\phi \Phi \lambda_{t-T}$ to get

$$\hat{C}_t = \gamma_\phi \Phi \lambda_{t-T} + \hat{\gamma}_m m_{t-T} + \sum_{\tau=-T+1}^{0} \hat{\gamma}_{z,\tau} \epsilon_{t+\tau} \tag{30}$$

The next steps depend on whether we want to compute the likelihood via the Kalman filter or via the sequence form.

3.5 Computing the likelihood

3.5.1 Computing the likelihood in the state space form

The standard way to compute the likelihood in a linear model with unobserved state vector is to write the model in state space form and apply the Kalman filter. In our context, there are two obstacles. First, the state vector of the model, even the reduced vector derived above, is large compared to number of observables. This makes the Kalman filter rather expensive to compute. Second, the covariance matrix of the state vector and also of the reduced state vector m is usually singular, which again makes application of the Kalman filter more cumbersome. We can overcome the second problem by switching to the lower-dimensional representation described in Section 3.4.5, choosing the number of principal components such that the covariance matrix is not close to singularity.

Whether it is worthwhile to use the Kalman filter depends on the number of available data periods and on the dimension of the reduced state. In case we decide for the Kalman filter, we replace in (26) the terms in m by terms in λ, using $m_t = \bar{M} \phi_t = \bar{M}\ \Phi \lambda_t$. We can then initialize the Kalman filter by the unconditional covariance matrix Σ_λ of λ_t, which solves the Lyapunov equation

$$\Sigma_\lambda = A_{PCA} \Sigma_\lambda A'_{PCA} + B_{PCA} \Sigma_\epsilon B'_{PCA} \tag{31}$$

3.5.2 Computing the likelihood via the moving average representation

If the vector of all data (observations), denoted by D, has a dimension in the range of 1000 or below, and if the state space is high-dimensional, it is computationally more efficient to compute the log likelihood by "brute force", i.e., to compute the covariance matrix Σ_D of all the observations, and apply the formula for the multivariate normal:

$$-0.5 \text{logdet} \Sigma_D - 0.5 (D - \bar{D})' \Sigma_D^{-1} (D - \bar{D}) \tag{32}$$

As observed by Auclert et al. (2019), both the log-determinant and the term $\Sigma_D^{-1}(D - \bar{D})$ can be easily recovered from the Cholesky decomposition of Σ_D.

For this, we need to approximate $\hat{C}_t$ as a linear function of the truncated history of shocks:

$$\hat{C}_t \approx \hat{C}^* + \sum_{j=1}^{T} \hat{C}_j z_{t-A+j} \tag{33}$$

We get there from (30) applying (29) repeatedly on the term λ_{t-T} and (27) repeatedly on the term m_{t-T}.

3.6 Computational complexity

The computation of the steady state and the automatic differentiation of the equation system (steps 1 and 2 in Section 3.2) are very time-consuming. If these steps have to be repeated for every new parameter vector, estimation is slowed down considerably. We therefore assume here that those parameters that do affect the steady state are calibrated, and that we only estimate the remaining parameters. Those parameters will only appear in a few places in the whole equation system, so that the linearization can be very easily updated.

Model reduction (step 3) is also a complex procedure. It can be avoided by using the "MIT-shock algorithm" of Boppart et al. (2018) with the implementation of Auclert et al. (2019). We have not yet investigated whether all the steps in the panel data estimation can be done efficiently with their approach. For our computations, we use the fact that the matrices $\bar{M}$ and $\bar{V}$ in the model reduction of Section 3.3 do not change if the steady state does not change.

Even after model reduction, the model solution (Step 4) is still slow with a QZ decomposition. To speed the solution up, we adapt the iterative algorithm of Rendahl (2017). Because we can use the earlier solution as a starting point, we can greatly reduce the number of iterations required.[11]

Linearizing the observations (Step 5) is made as efficient as possible by reverse-mode automatic differentiation, which is the topic of Section 3.4.4. Identifying the state space in which the solution lives is time-consuming, but experience so far suggests (although we cannot prove it) that this space almost does not change between different parameter values.

The final step, computation of the likelihood (step 7), is very fast in comparison to the other steps. This means, in particular, that estimation of variance parameters only, which do not affect the model solution, is very cheap compared to the estimation of model parameters.

4 Numerical Examples

4.1 Parameter values

For the data generating process, we take all parameter values that are not related to individual productivity directly from Table 1 of Takahashi (2019). These values are the discount factor $\beta = 0.9829$. a disutility of labor $B = 3.061$, working hours $\bar{h} = 1/3$, the borrowing limit $\bar{a} = -2.0$, the labor share $\alpha = 0.64$, the capital depreciation rate $\delta = 0.025$, the persistence in aggregate TFP $\rho_z = 0.95$, and the standard deviation of aggregate TFP shocks $\sigma_z = 0.007$.

We assume the i.i.d. component of log individual productivity to be uniformly distributed between -0.2 and 0.2, and approximate this with 11 potential realizations of the shock. The main point of introducing i.i.d. shocks is to smooth the cross-sectional distribution, which appears very jagged if the only idiosyncratic shock to households is a finite Markov chain of low dimension. The quarterly autocorrelation and shock variance of the Markov component was chosen such that the log of annual individual productivity has an autocorrelation of 0.98 and variance of 0.4. This is approximately the variance of the log of male hourly earnings 1990, according to Figure 16 in Heathcote et al. (2010). These targets were met by choosing $\rho_\varsigma = 0.992675$ and $\bar{\sigma} = 0.076377$. As a benchmark, we assume that aggregate TFP and idiosyncratic uncertainty are uncorrelated, $\sigma_{z,\varsigma} = 0$. The numerical exercise will test how precisely this covariance is estimated. To match the targets in Table 2 of Takahashi

[11] If only the parameters of the exogenous shock process change, the forecast function w.r.t. the cross-sectional distribution is unaffected. This means that almost all columns of the forecast matrix are known ex ante, thereby greatly reducing the dimension of the problem. We have not exploited this below, because it will not apply in more general cases.

discount factor	β	0.9829
disutility of labor	B	3.061
working hours	$\bar{h}$	$1/3$
borrowing limit	$\bar{a}$	-2.0
labor share	α	0.64
capital depreciation rate	δ	0.025
persistence HH prodtty	ρ_z	0.992675
stdev HH prodtty	$\bar{\sigma}$	0.076377
persistence in aggregate TFP	ρ_z	0.95
stdev aggregate TFP	σ_z	0.007
persistence volatility	ρ_ς	$0.2^{1/4}$
stdev volatility	σ_ς	0.022
stdev i.i.d. prodtty	σ_{ϵ_z}	0.005
stdev i.i.d. deprec.	σ_{ϵ_δ}	0.002

Table 1: Parameter values

(2019), we set the autocorrelation and innovation variability to $\rho_\varsigma = 0.2^{1/4}$ and $\sigma_\varsigma = 0.022$.

For the numerical exercise, we have enlarged the model by two aggregate i.i.d. shocks, one to the depreciation rate of capital, with a standard deviation of 0.002, and the other one an i.i.d. shock to TFP, on top of the AR component of TFP, with a standard deviation of 0.005. This was done to avoid stochastic singularity when we estimate the model with more than two data points per quarter.

4.2 Data

From this model, with the parameter values described above, we generate 100 years of data by simulation. These data consist of

- 4×100 quarterly data of
  - aggregate output
  - aggregate hours
- 100 annual data of
  - the cross-sectional variance of annual household income
  - the auto-covariance of annual household income

Maximum likelihood is very efficient in a correctly specified model with few shocks. We use a relatively long data series, but assume that only few data series are observed. Notice that we do not observe aggregate capital and aggregate labor efficiency units, so that TFP cannot be directly inferred from the data.

4.3 Monte-Carlo results

We estimate seven parameters that do not affect the steady state of the model, namely the two autocorrelation coefficients ρ_z and ρ_ς, the four standard deviations of the shocks, σ_z, σ_ς, σ_{ϵ_δ} and σ_{ϵ_z}, as well as the covariance

	True value	PanelData		AggrData		SamplingErr	
ρ_z	0.950	0.950	(0.002)	0.949	(0.005)	0.949	(0.008)
ρ_ς	0.669	0.663	(0.057)	0.597	(0.349)	0.650	(0.183)
σ_z	0.700	0.699	(0.029)	0.699	(0.033)	0.697	(0.039)
σ_ς	2.200	2.215	(0.236)	2.481	(2.053)	2.302	(0.964)
σ_{ϵ_δ}	0.200	0.198	(0.016)	0.158	(0.135)	0.190	(0.056)
σ_{ϵ_z}	0.500	0.498	(0.018)	0.497	(0.025)	0.507	(0.021)
$\sigma_{z,\varsigma}$	0.000	-0.000	(0.001)	0.001	(0.004)	-0.000	(0.003)

Notes: averages of estimated parameters, 500 repetitions;
standard deviations in parentheses

Table 2: Monte-Carlo estimation results

between the TFP and the volatility shock, $\sigma_{z,\varsigma}$. The latter is zero in the model, but the question is how well the estimation can identify this. The following table reports means and standard deviations of the estimator, obtained through a Monte Carlo study with 500 samples. The column "PanelData" shows the results when the two aggregate variables are used, namely GDP and hours worked, plus T years of the two panel statistics. This makes a total of 1000 observations. The column "AggrData" shows the results when only the aggregate variables are used. The last column, "SamplingErr", again uses all observations, but now accounts for the sampling error in the panel observations. The first column was, unrealistically, assuming that the households in the observed panel were drawing the same idiosyncratic shocks that were used for the computation of the likelihood. In the last column, in each experiment we draw not only a new set of aggregate shocks, but do the sampling procedure of Section 3.4 with different realizations of the idiosyncratic shocks. From the table it is clear that the panel estimation delivers much more precise estimates than an estimation only based on the aggregate data. The most significant difference is that the panel estimates have a much lower standard deviation for the variances of idiosyncratic volatility and of the depreciation shock. Part of this advantage is lost when we account for the sampling error in the estimation. This is because, in our simulation exercise, the sampling error introduces a slight misspecification into an otherwise correctly specified model. To mitigate the effect of the misspecification, we consider the sampling error as a special type of measurement error. One can compute the implications of the model for the size of this measurement error by repeating the sampling procedure with different idiosyncratic shocks, and analyze the statistical properties of the sampling error. We have performed the estimation assuming a measurement error variance of the panel data that conforms approximately to the estimated sampling error. A more sophisticated implementation would also account for the autocorrelation of the errors.

We consider the numerical results here mostly as a proof of concept, showing that the proposed methodology works with simulated data. Of course, estimation with aggregate data can be made more precise by observing more aggregate data series. Whether panel data are useful in an application with real data or not will always depend on the specifics of the application, including the reliability of the data and the extent of model misspecification.

4.4 Running times

All the computations were done in the programming language Julia. Computation times are reported for a laptop with an Intel Core i7-8750H CPU at 2.20GHz. For transparency, the computations were done using a single core, and the linear algebra libraries were also restricted to use only one core. Optimization of all the

computations involved, including the parallelization, is still an ongoing project.

Computation times depend on the size of the grid involved, the length of the data series, the size of the simulated panel, and also the parameters, especially the autocorrelation of shocks. For the example model, the grid for the value function has size 400×17, the grid for the cross-sectional distribution has size 1000×17, and the i.i.d. shock has 11 possible realizations. The likelihood was computed after bringing all variables into sequence form (Section 3.5.2), with MA-length of 800 quarters. Computation times also depend on the accuracy of the state aggregation. For our estimates, we use a high-precision solution where the dimension of the reduced state vector m is 693. One can considerably speed up all the computations by a slight reduction in precision. Finding the right trade-off between speed and accuracy would have to be done in every specific application. We expect to significantly reduce computation times by parallelization and further optimizations.

The preparatory steps (steady state, differentiation, state reduction), that are done only once, take in the range of 2–3 minutes, depending on how good the starting point for the steady state is. Of this, differentiation of the model equations takes about one minute, and the same for model reduction. These times would be considerable lower without i.i.d. shocks.

The likelihood evaluation is very fast if only the covariances of the shocks are changed, about 0.05 seconds if 100 years of data are used. More time is required if the model parameters (correlation coefficients) are changed. Solving the model again, after changing a parameter, but reusing the model reduction, takes roughly 5 seconds (step 4 in Section 3.2). The reverse sweep of the differentiation takes currently about 3 seconds (step 2), which depends on the sample size of the panel of agents. In the reported results, it is around 67,000 at the start, and rises to around 289,000 agents at the end of the two years, because of splitting. The computation of the invariant subspace of the model takes about 5 seconds. As we said above, this time can be saved if the invariant subspace does not change.

5 Conclusions

Big data sets are becoming more and more important in macroeconomics. They are mostly used to calibrate quantitative heterogeneous agent models, but not yet widely used for model estimation. We have developed a method that can make use of the business cycle aspects in panel data also for estimation. This is potentially very time-consuming. We have discussed the algorithmic details that allow to do the estimation reasonably fast. A single likelihood evaluation takes a few seconds if model parameters are estimated, and is in the range of milliseconds if only variance parameters are estimated.

The method we develop is useful not only for estimation, but more generally for analysing the implications of the model for any variables that include panel information. We have combined the panel method with the state reduction approach of Reiter (2010), but it could also be applied together with the "MIT-shock algorithm" in Boppart et al. (2018) and Auclert et al. (2019).

References

Achdou, Yves et al. (2017). *Income and wealth distribution in macroeconomics: A continuous-time approach*. Tech. rep. National Bureau of Economic Research.

Ahn, SeHyoun et al. (2018). "When inequality matters for macro and macro matters for inequality". In: *NBER Macroeconomics Annual* 32.1, pp. 1–75.

Algan, Yann, Olivier Allais, and Wouter J Den Haan (2008). "Solving heterogeneous-agent models with parameterized cross-sectional distributions". In: *Journal of Economic Dynamics and Control* 32.3, pp. 875–908.

Algan, Yann, Olivier Allais, Wouter J Den Haan, and Pontus Rendahl (2014). "Solving and simulating models with heterogeneous agents and aggregate uncertainty". In: *Handbook of Computational Economics*. Vol. 3. Elsevier, pp. 277–324.

Andrieu, Christophe et al. (2010). "Particle Markov chain Monte Carlo methods". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3, pp. 269–342.

Auclert, Adrien et al. (2019). "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models". Manuscript.

Bayer, Christian, Benjamin Born, et al. (2019). *Shocks, Frictions, and Inequality in US Business Cycles*. Tech. rep.

Bayer, Christian and Ralph Luetticke (July 2018). *Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods*. CEPR Discussion Papers 13071. C.E.P.R. Discussion Papers. URL: https://ideas.repec.org/p/cpr/ceprdp/13071.html.

Boppart, Timo et al. (2018). "Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative". In: *Journal of Economic Dynamics and Control* 89.C, pp. 68–92.

Chang, Minsu et al. (2018). "Heterogeneity and Aggregate Fluctuations". In: *Manuscript, University of Pennsylvania*.

Chang, Yongsung and Sun-Bin Kim (2006). "From Individual To Aggregate Labor Supply: A Quantitative Analysis Based On A Heterogeneous Agent Macroeconomy". In: *International Economic Review* 47.1, pp. 1–27. URL: http://ideas.repec.org/a/ier/iecrev/v47y2006i1p1-27.html.

— (2007). "Heterogeneity and Aggregation: Implications for Labor-Market Fluctuations". In: *American Economic Review* 97.5, pp. 1939–1956. URL: http://ideas.repec.org/a/aea/aecrev/v97y2007i5p1939-1956.html.

Den Haan, Wouter J. (1997). "Solving dynamic models with aggregate shocks and heterogenous agents". In: *Macroeconomic Dynamics* 1, pp. 355–86.

Fernández-Villaverde, Jesús et al. (2016). "Solution and estimation methods for DSGE models". In: *Handbook of Macroeconomics* 2, pp. 527–724.

Fernández-Villaverde, Jesús et al. (2019). *Financial frictions and the wealth distribution*. Tech. rep. National Bureau of Economic Research.

Flury, Thomas and Neil Shephard (2011). "Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models". In: *Econometric Theory* 27.5, pp. 933–956.

Gilchrist, Simon et al. (2014). *Uncertainty, financial frictions, and investment dynamics*. Tech. rep. National Bureau of Economic Research.

Griewank, A. and A. Walther (2008). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, Second Edition*. SIAM e-books. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104). URL: http://books.google.at/books?id=xoiiLaRxcbEC.

Hasumi, Ryo and Hirokuni Iiboshi (2019). *A Bayesian Estimation of HANK models with Continuous Time Ap-*

proach: Comparison between US and Japan. Tech. rep. MPRA Working Papers.

Heathcote, Jonathan et al. (Jan. 2010). "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006". In: *Review of Economic Dynamics* 13.1, pp. 15–51.

Herbst, Ed and Frank Schorfheide (2017). "Metropolis-Hastings Algorithms for DSGE Models". `https://web.sas.upenn.edu/schorf/files/2017/05/chapter4_sl_0-1sxxi8g.pdf`.

Higham, Nicholas (2002). *Accuracy and Stability of Numerical Algorithms*. 2nd. SIAM.

Kaplan, Greg et al. (Mar. 2018). "Monetary Policy According to HANK". In: *American Economic Review* 108.3, pp. 697–743.

Khan, Aubhik and Julia K Thomas (2008). "Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics". In: *Econometrica* 76.2, pp. 395–436.

Krusell, Per and Anthony A. Smith (1998). "Income and Wealth Heterogeneity in the Macroeconomy". In: *Journal of Political Economy* 106.5, pp. 867–96.

Liu, Laura and Mikkel Plagborg-Møller (June 2019). "Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data". SED.

Mongey, Simon and Jerome Williams (2017). "Firm dispersion and business cycles: Estimating aggregate shocks using panel data". In: *Manuscript, New York University*.

Reiter, Michael (2009). "Solving Heterogenous Agent Models by Projection and Perturbation". In: *Journal of Economic Dynamics and Control* 33.3, pp. 649–665.

— (2010). "Approximate and Almost-Exact Aggregation in Dynamic Stochastic Heterogeneous-Agent Models". IHS Working Paper 258.

— (2019). "Solving Heterogeneous Agent Models with Non-convex Optimization Problems: Linearization and Beyond". Manuscript.

Rendahl, Pontus (2017). "Linear Time Iteration".

Takahashi, Shuhei (2019). "Time-Varying Wage Risk, Incomplete Markets, and Business Cycles". Manuscript.

Winberry, Thomas (2016). "A toolbox for solving and estimating heterogeneous agent macro models". `https://site.stanford.edu/sites/g/files/sbiybj8706/f/winberryalgorithm_0.pdf`.

— (2018). "A method for solving and estimating heterogeneous agent macro models". In: *Quantitative Economics* 9.3, pp. 1123–1151.

A Figures

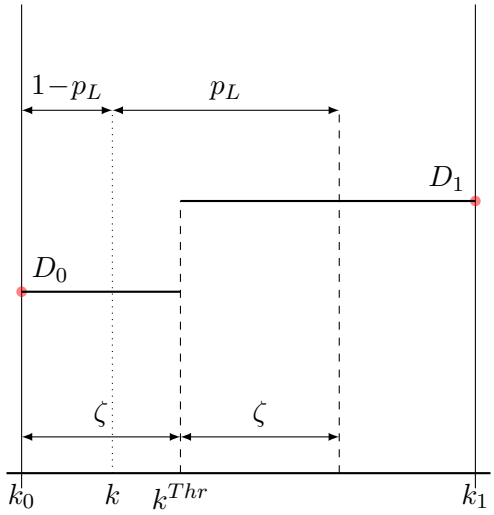

Figure 1: Illustration of endogenous transition, Case 2c (threshold bin, $k_0 = k^{Thr} - \zeta < k \leq k^{Thr}$). Probabilities p_L and $1-p_L$ just illustrate proportions and do not match scale otherwise.

B Iterative Solution of Linearized Model

We can write the linear model in the form

$$\Lambda x_{t-1}+\Gamma x_t+\mathrm{E}_t\Phi x_{t+1}+\Psi\epsilon_t=0 \tag{34}$$

Linearity implies that the expected value can be written as

$$\mathrm{E}_t x_{t+1}=F x_t \tag{35}$$

for an unknown matrix F. This implies

$$x_t=-(\Gamma+\Phi F)^{-1}(\Lambda x_{t-1}+\Psi\epsilon_t) \tag{36}$$

Consistency with (35) requires

$$F=-(\Gamma+\Phi F)^{-1}\Lambda \tag{37}$$

Algorithm:

1. Initialize $F=0$.
2. Iterate until convergence (Rendahl 2017; Higham 2002):

$$F\leftarrow(\Gamma+\Phi F)^{-1}\Lambda \tag{38}$$

Formula (38) is easy to write down but not computationally efficient. An efficient implementation takes the following into account.

- One need not compute the full matrix F, but only the part that refers to the forward looking variables. Only parts of the matrices Γ, Φ and Λ need to be used for that.
- The large matrix inverse $(\Gamma+\Phi F)^{-1}$ must never be computed. The linear equation system can be solved making use of a variant of the Sherman-Morrison-Woodbury formula.
- F need not be initialized to zero if an estimate of F is available from an earlier calculation with similar parameter values.

Some further optimizations are possible. For details, see the computer code accompanying this paper.