5-20-2021

Heterogeneity and Aggregate Fluctuations

Minsu Chang

Xiaohong Chen
Yale University

Frank Schorfheide

HETEROGENEITY AND AGGREGATE FLUCTUATIONS

By

Minsu Chang, Xiaohong Chen, and Frank Schorfheide

May 2021

COWLES FOUNDATION DISCUSSION PAPER NO. 2289

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Heterogeneity and Aggregate Fluctuations

Minsu Chang*
Georgetown University

Xiaohong Chen
Yale University,
Cowles Foundation

Frank Schorfheide
University of Pennsylvania,
CEPR, PIER, NBER

This Version: May 20, 2021

Abstract

We develop a state-space model with a state-transition equation that takes the form of a functional vector autoregression and stacks macroeconomic aggregates and a cross-sectional density. The measurement equation captures the error in estimating log densities from repeated cross-sectional samples. The log densities and the transition kernels in the law of motion of the states are approximated by sieves, which leads to a finite-dimensional representation in terms of macroeconomic aggregates and sieve coefficents. We use this model to study the joint dynamics of technology shocks, per capita GDP, employment rates, and the earnings distribution. We find that the estimated spillovers between aggregate and distributional dynamics are generally small, a positive technology shock tends to decrease inequality, and a shock that raises the inequality of earnings leads to a small but not significant increase in GDP. (JEL C11, C32, C52, E32)

Key words: Bayesian Model Selection, Econometric Model Evaluation, Earnings Distribution, Functional Vector Autoregressions, Heterogeneous Agent Models, State-space Model, Technology Shocks

* Correspondence: M. Chang: Department of Economics, Georgetown University, Washington, D.C., 20057-1036. Email: minsu.chang@georgetown.edu (Chang). F. Schorfheide: Department of Economics, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu (Schorfheide). X. Chen: Department of Economics and Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-8281. Email: xiaohong.chen@yale.edu. We thank Yongsung Chang, Mark Huggett, Dirk Krüger, Víctor Ríos-Rull as well as seminar participants at various seminars and conferences for helpful suggestions. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grant SES 1851634.

1 Introduction

Models with heterogeneity on the household side or the firm side have long been used to study distributional effects of macroeconomic policies. Heterogeneity evolves dynamically and in some of these models interacts closely with aggregate fluctuations. This is particularly true in models with financing constraints that try to capture the large downturn during the recent Great Recession. While the macroeconomics literature has demonstrated that dynamics in heterogeneous agent (HA) models can be different from their representative agent (RA) counterparts, it is an open question whether in the data there is strong evidence that the dynamics of macroeconomic aggregates interacts, at business cycle frequencies, with the evolution of the cross-sectional distribution of income and wealth on the household side and the distribution of productivity and capital on the firm side.

This paper develops and applies econometric tools that can provide semi-structural evidence about the interaction of aggregate and distributional dynamics. More specifically, we specify a state-space model with a state-transition equation that takes the form of a linear functional vector autoregression (fVAR) and stacks macroeconomic aggregates and a cross-sectional distribution. The cross-sectional distribution is represented as a log probability density function. This has the advantage that its law of motion is not constrained by non-negativity or monotonicity restrictions and hence can be linear. We specify measurement equations for macroeconomic aggregates as well as cross-sectional observations, treating the underlying densities as unobserved. The measurement equation for the cross-sectional observations is nonlinear because it needs to exponentiate the log density and normalize it such that it integrates one.

To make the functional analysis tractable and easy to implement, we approximate the log-densities of the cross-sectional distributions as well as the transition kernels in the functional autoregressive law of motion of the states by finite-dimensional sieves with fixed basis functions and time-varying coefficients that capture the dynamics. This conveniently turns the state-transition equation into a joint vector-autoregressive law of motion for the aggregate variables and the time-varying coefficients of the sieve approximations. To avoid nonlinear filtering, we effectively linearize the measurement equation for the cross-sectional observation. This approximation enables us to estimate the model in two steps. First, we estimate the coefficients of the density approximation for each time period based on the sample of cross-sectional observations. The approximation allows us to treat the coefficients estimates as noisy measures of their population counterparts. The measurement errors capture the

estimation uncertainty associated with the sieve coefficients. In the second step, we estimate a linear state-space model using Bayesian techniques.

As mentioned before, our procedure is able to account for estimation errors in the cross-sectional densities and can also be used with small cross-sectional sample sizes. Our implementation uses spline basis functions to approximate log densities, which dates back to Kooperberg and Stone (1990). We allow for seasonal adjustments of the cross-sectional densities if they are combined with seasonally-adjusted aggregate data. Moreover, we show how to adjust the measurement equation for top-coding of the cross-sectional observations. Prior distributions are used to regularize a potentially high-dimensional estimation problem. Finally, in order to determine the number of basis functions K for the sieve approximation in a data-driven way, we utilize our Bayesian framework to derive a marginal data density (MDD) approximation for the aggregate and cross-sectional data, akin to the widely-used Schwarz criterion. Moreover, we use the MDD approximation for model selection, but it could also be used for model averaging. We use the MDD not only to determine K, but also to determine the prior precision (and hence the degree of shrinkage) for the coefficients of the vector-autoregressive state-transition equation; see, for instance, Del Negro and Schorfheide (2004) and Giannone, Lenza, and Primiceri (2015).

We apply our econometric framework to simulated and actual data. In the simulation experiment, we generate data from a version of the Krusell and Smith (1998), henceforth KS, model, which we solve using the method proposed by Winberry (2018). Households' decisions in the KS model depend on the aggregate technology shock and the cross-sectional distribution of skills and wealth. In turn, the entirety of the household-level decisions determine the cross-sectional distribution, which leads to a joint law of motion of aggregate variables and cross-sectional distribution. We show that estimates of our functional state-space model can reproduce the evolution of the distribution of asset holdings and the response of the cross-sectional distribution to a technology shock. Our MDD model selection criterion chooses a low-dimensional sieve approximation that trades off model fit and complexity and shrinks the coefficients that control the effect of the lagged cross-sectional distribution on current technology to zero, capturing the fact that technology is indeed exogenous in the underlying KS economy.

In the empirical application we fit our model to aggregate total factor productivity (TFP) growth, GDP growth, employment, and cross-sectional data on labor earnings. The latter are scaled by the labor share of GDP. The empirical model is able to capture the time path of the cross-sectional densities. Based on the degree of shrinkage selected with the

MDD criterion, there is strong evidence that aggregate dynamics do not Granger-cause cross-sectional dynamics and vice versa. Our estimated covariance matrix of the fVAR innovations is, however, not block diagonal, and we are able to detect small but in many cases statistically significant effects of TFP shocks on the cross-sectional density of earnings. Most notably, the mass of individuals that earn less than GDP per capita rises in response to a TFP shock. From the response of the densities, we can also compute the response of percentiles and inequality measures. Here we find that earnings at the 10th percentile rise, whereas earnings at the 90th percentile fall. Thus, the expansion reduces inequality, which is also reflected in the response of the Gini coefficient. Overall, however, the effect is small.

We also consider the effect of a distributional shock that is constructed to increase inequality in the short run. This shock is associated with a slight increase in the employment rate and GDP, generating a positive correlation between inequality and aggregate economic activity. As a robustness exercise, we consider two simplified vector autoregressions (VARs) that combine the aggregate variables with a set of percentiles or with measures of inequality.[1] While these simplified VARs deliver qualitatively similar IRFs as the functional model, the responses are less precisely estimated and often feature odd long-run dynamics. A key problem with VARs that use quantiles of cross-sectional distributions is that nothing prevents the quantiles from crossing in a simulation or an impulse response function, which is clearly undesirable. The functional model, on the other hand, is coherent and parsimonious in the sense that any distributional statistic can be derived from it.

The structure of the transition equation in our functional state-space model resembles that of HA models solved with linearization techniques. This solution method was initially proposed by Reiter (2009) and has been further developed in several papers, including Kaplan and Violante (2018), Childers (2018), and Winberry (2018). The basic idea is to first solve for the stationary solution of the HA model with only idiosyncratic uncertainty, using a global solution technique. In a second step, one computes a first-order perturbation of the solution with respect to the aggregate shocks. This leads to a vector autoregressive law of motion that includes the time-varying coefficients associated with the basis functions that are used to approximate cross-sectional distributions in the first step. This solution technique is convenient for the likelihood-based estimation of HA models as in Mongey and Williams (2017), Acharya, Chen, Del Negro, Dogra, Matlin, and Sarfati (2019), Liu and Plagborg-Møller (2019), Bayer, Born, and Luetticke (2020), Cho (2020).

[1] An example of this approach is Coibion, Gorodnichenko, Kueng, and Silvia (2017).

In general, our econometric framework allows researchers to examine the following type of quantitative questions in a semi-structural framework that does not require the solution, calibration, or estimation of an HA model: First, does micro-level heterogeneity affect the propagation of aggregate shocks to aggregate variables?[2] Second, what is the effect of an aggregate shock on cross-sectional distributions? Third, what is the effect of a change in the cross-sectional distribution on macroeconomic aggregates?

The aforementioned questions are front and center in the quantitative macroeconomics literature. In order to address the first question, researchers often compare implications, such as impulse response functions, correlations, or forecasts from models with and without micro-level heterogeneity that share an otherwise identical identification structure. Recent examples include Ahn, Kaplan, Moll, Winberry, and Wolf (2018), Kaplan and Violante (2018), Ottonello and Winberry (2020), Acharya, Chen, Del Negro, Dogra, Matlin, and Sarfati (2019), Bayer, Born, and Luetticke (2020), Cho (2020), and Villalvazo (2021). Heterogeneity often leads to amplification of shocks and additional persistence; see also Krueger, Mitman, and Perri (2016). A key challenge is to find a parameterization for the HA and RA model that allows a meaningful comparison. In our setting we can easily make comparisons between the functional state-space model and VARs that only include aggregate variables.

The second question, i.e., the effect of aggregate shocks on cross-sectional distributions is also frequently studied in the context of HA models. For instance, the paper by Ahn, Kaplan, Moll, Winberry, and Wolf (2018) studies the effect of factor-specific productivity shocks on inequality dynamics, whereas Kaplan and Violante (2018) examine the distributional effects of monetary policy shocks. Bayer, Born, and Luetticke (2020) use their estimated HA model to construct a historical decomposition of an inequality measure with respect to a collection of aggregate shocks. Bhandari, Evans, Golosov, and Sargent (2021) report responses of the dispersion of assets to a TFP shock in a HANK model under optimal monetary-fiscal policy. Mongey and Williams (2017) analyze the effect of macro shocks on the dispersion of sales growth. Finally, examples of research examining the third question are papers by Huggett (1997) and Auclert and Rognlie (2020). The former considers a one-sector growth model with idiosyncratic shocks and shows that a redistribution of asset holdings among households while keeping the overall capital stock fixed at its steady state level, triggers a

[2]From the perspective of a linearized HA model, heterogeneity could affect the steady state (mean) and the elasticities (slope coefficients). That is something our framework is unable to detect. However, we can measure potential indirect effects: an aggregate shock may shift the cross-sectional distribution, which in turn triggers movements in the aggregate variables in subsequent periods.

response of the aggregate variables. Auclert and Rognlie (2020) show that a rise in inequality may lower aggregate output if monetary policy does not react to it.

There is an extensive literature on the statistical analysis of functional data. General treatments are provided in the books by Bosq (2000), Ramsey and Silverman (2005), and Horvath and Kokoszka (2012). Much of the literature assumes that the functions are observed without error. Our state-transition equation embodies the widely studied functional autoregressive model (see Bosq (2000) for a detailed analysis), but also includes a finite-dimensional vector of states. Functional autoregressive models can be estimated by functional principal component analysis, which approximates the functions by linear combinations of the eigenfunctions of the sample covariance operator associated with the K largest eigenvalues. Rather than using what the literature considers to be an optimal (in a least squares sense) *empirical* orthonormal basis, we use a spline basis that is chosen independently of the data in our analysis.

Applications of functional data analysis in macroeconometrics are growing steadily. Many of them are related to the yield curve. The state-space model Diebold and Li (2006) could be interpreted as a functional model for yield-curve data, but there is no infinite-dimensional aspect to the analysis in the sense that it is assumed (and empirically justified) that the yield-curves can be represented by three time-varying parameters. Inoue and Rossi (2020) estimate what they call a VAR with functional shocks, which uses a similar representation of the yield curve as in Diebold and Li (2006) and focuses on the identification of functional monetary policy shocks. Meeks and Monti (2019) use functional principal component regression to estimate a New Keynesian Phillips curve in which the distribution of inflation expectations appears on the right-hand side. Hu and Park (2017) develop an estimation theory for a functional autoregressive model with unit roots and fit it to yield curve data and Chang, Kim, and Park (2016) use a functional time series process to capture the evolution of earnings densities with a focus on unit-root components. Both papers use functional principal components analysis.

The remainder of this paper is organized as follows. In Section 2 we present our functional state-space model for a group of macroeconomic time series and a sequence of cross-sectional distributions. We develop an approximate filter for the state-space model that facilitates the likelihood-based estimation. We use Bayesian inference and derive an approximation to the marginal data density that is used for dimensionality and hyperparameter selection. Implementation details such as the choice of basis functions, the handling of top coding and seasonal adjustments, the choice of prior distributions, and the posterior sampler are

discussed in Section 3. Section 4 contains the results from the simulation experiment in which we estimate the functional state-space model based on data generated from the KS economy. The empirical application is presented in Section 5 and Section 6 concludes. Supplemental derivations, and additional computational details and empirical results are relegated to the Online Appendix.

2 A Functional State Space Model

VARs can be viewed as approximations to the equilibrium dynamics arising from linearized RA models and have proved to be useful for the evaluation and development of dynamic stochastic general equilibrium (DSGE) models. Moreover, VARs are widely used in empirical macroeconomics independently of DSGE models, to study business cycle fluctuations, the propagation of shocks, and to generate macroeconomic forecasts. In this paper, we develop a fVAR that is embedded in a state space model and can play a similar role as traditional VARs in environments in which macroeconomic aggregates interact with cross-sectional distributions. While we will not establish a formal link between our functional model and the solution of HA models, it does capture the salient features and provides a natural reference model for the evaluation of HA models. Moreover, just as VARs, our functional model can be used as a stand-alone tool for empirical work in macroeconomics.

The variables in the functional model comprise an $n_y \times 1$ vector of macroeconomic aggregates Y_t and a cross-sectional density $p_t(x)$. In our empirical application Y_t consists of (log) TFP growth, per-capita GDP growth, and the log employment rate. The cross-sectional variable x is earnings as a fraction of per-capita GDP. Throughout this paper, we will work with log densities defined as $\ell_t(x) = \ln p_t(x)$. We decompose Y_t and ℓ_t into a deterministic component $\big(Y_*, \ell_*(x)\big)$ and fluctuations around the deterministic component. Let

$$Y_t = Y_* + \tilde{Y}_t, \quad \ell_t = \ell_* + \tilde{\ell}_t. \tag{1}$$

For notational convenience we assumed that the deterministic component is time-invariant and could be interpreted as a steady state. This assumption could be easily relaxed by letting (Y_*, ℓ_*) depend on t. We assume that the deviations from the deterministic component $\big(Y_t, \ell_t(x)\big)$ evolve jointly according to the following linear fVAR law of motion:

$$\begin{aligned} \tilde{Y}_t &= B_{yy}\tilde{Y}_{t-1} + \int B_{yl}(\tilde{x})\tilde{\ell}_{t-1}(\tilde{x})d\tilde{x} + u_{y,t} \\ \tilde{\ell}_t(x) &= B_{ly}(x)\tilde{Y}_{t-1} + \int B_{ll}(x,\tilde{x})\tilde{\ell}_{t-1}(\tilde{x})d\tilde{x} + u_{l,t}(x). \end{aligned} \tag{2}$$

Here $u_{y,t}$ is mean-zero random vector with covariance Ω_{yy} and $u_{l,t}(x)$ is a random element in a Hilbert space with covariance function $\Omega_{ll}(x,\tilde{x})$. We denote the covariance function for $u_{y,t}$ and $u_{l,t}(x)$ by $\Omega_{yl}(x)$.

To condense the notation, we define integral operators with kernels $B_{yl}(\tilde{x})$ and $B_{ll}(x,\tilde{x})$ as

$$\boldsymbol{B}_{yl}[g] = \int B_{yl}(\tilde{x})g(\tilde{x})d\tilde{x}, \quad \boldsymbol{B}_{ll}[g](x) = \int B_{ll}(x,\tilde{x})g(\tilde{x})d\tilde{x}.$$

Using the operator notation, we can write (2) more compactly as

$$\begin{aligned} \tilde{Y}_t &= B_{yy}\tilde{Y}_{t-1} + \boldsymbol{B}_{yl}[\tilde{\ell}_{t-1}] + u_{y,t} \\ \tilde{\ell}_t(x) &= B_{ly}(x)\tilde{Y}_{t-1} + \boldsymbol{B}_{ll}[\tilde{\ell}_{t-1}](x) + u_{l,t}(x). \end{aligned} \tag{3}$$

For now, (3) should be interpreted as a reduced-form fVAR in which $u_{y,t}$ and $u_{l,t}(x)$ are one-step-ahead forecast errors. We could easily add more lags, but our simulation experiment and the empirical analysis will be based on a single lag. The system will subsequently serve as the state-transition equation in a functional state-space model.

Section 2.1 describes the sampling from the functional state-space model. Rather than focusing on an infinite-dimensional model in which the densities and operators are treated nonparametrically, we consider a collection of finite-dimensional models, presented in Section 2.2, in which log densities and kernels associated with integral operators are represented through finite-dimensional sieves. Section 2.3 develops an approximate linear filter for the finite-dimensional functional state-space models and in Section 2.4 we present a large-sample approximation of the marginal data density that can be used to select the degree of the sieve-approximation in a data-driven manner.

2.1 Sampling and Measurement

We assume that in every period $t = 1,\ldots,T$ an econometrician observes the macroeconomic aggregates Y_t as well as a sample of N_t draws x_{it}, $i = 1,\ldots,N_t$ from the cross-sectional density $p_t(x)$. In practice, N_t is likely to vary from period to period, but for the subsequent exposition it will be more convenient to assume that $N_t = N$ for all t. We collect the time t cross-sectional observations in the vector $X_t = [x_{1t},\ldots,x_{Nt}]'$. We also assume that the draws x_{it} are independently and identically distributed (iid) over the cross-section and independent over time. The measurement equation for the cross-section observations takes the form

$$x_{it} \sim \text{iid } p_t(x) = \frac{\exp\{\ell_t(x)\}}{\int \exp\{\ell_t(x)\}dx}, \quad i = 1,\ldots,N, \quad t = 1,\ldots,T. \tag{4}$$

The assumption of x_{it} being *iid* across i and t is consistent with data sets that comprise repeated cross sections.[3] It is also approximately consistent with panel data sets if the unit index i is randomly re-assigned in every period t. Thus, to the extent that the cross-sectional densities $p_t(x)$ are estimated from a panel data set, there is some potential loss of information in our functional modeling approach. However, on the positive side, the functional modeling approach does not require the econometrician to make assumptions about the evolution of x_{it} at the level of an individual, a household, or a firm.

2.2 A Collection of Finite-Dimensional Models

Equations (1), (3) and (4) define a state-space model for the observables $\{Y_t, X_t\}_{t=1}^T$. The log density $\ell_t(\cdot)$ is the state variable. To implement the estimation of the functional model we use a collection of finite-dimensional representations, indexed by a superscript (K). Let

$$\ell_t^{(K)}(x) = \sum_{k=1}^{K} \alpha_{k,t}\zeta_k(x) = \big[\zeta_1(x), \ldots, \zeta_K(x)\big] \cdot \begin{bmatrix} \alpha_{1,t} \\ \vdots \\ \alpha_{K,t} \end{bmatrix} = \zeta'(x)\alpha_t \tag{5}$$

and

$$\ell_*^{(K)}(x) = \zeta'(x)\alpha_*.$$

To simplify the notation a bit, we did not use (K) superscripts for the vectors $\zeta(x)$, α_t, and α_*. Here $\zeta_1(x), \zeta_2(x), \ldots$ is a sequence of basis functions. We define $\tilde{\alpha}_t = \alpha_t - \alpha_*$ such that $\tilde{\ell}^{(K)}(x) = \ell_t^{(K)}(x) - \ell_*^{(K)}(x)$. For theoretical considerations it is convenient to demean the vector of basis functions and assume that $\int \zeta(x)dx = 0$. For applications this normalization is not important.

To construct the measurement equation of the cross-sectional observations in (4), we define the K-dimensional vector of sufficient statistics

$$\bar{\zeta}(X_t) = \frac{1}{N}\sum_{i=1}^{N} \zeta(x_{it}).$$

This allows us to write a K'th order representation of the density of X_t:

$$\begin{aligned} p^{(K)}(X_t|\alpha_t) &= \exp\big\{N\mathcal{L}^{(K)}(\alpha_t|X_t)\big\}, \\ \mathcal{L}^{(K)}(\alpha_t|X_t) &= \bar{\zeta}'(X_t)\alpha_t - \ln \int \exp\{\zeta'(x)\alpha_t\}\, dx. \end{aligned} \tag{6}$$

[3] If the data exhibit spatial correlation, then our estimation approach below essentially replaces the likelihood function for $x_{1t}, \ldots, x_{Nt}$ by a composite likelihood function that ignores the spatial correlation; see Varin, Reid, and Firth (2011).

We represent the kernels $B_{ll}(x,\tilde{x})$ and $B_{yl}(\tilde{x})$, the function $B_{ly}(x)$, and the functional innovation $u_{l,t}(x)$ that appear in the state-transition equation (3) as follows:

$$\begin{aligned} B_{ll}^{(K)}(x,\tilde{x}) &= \zeta'(x)B_{ll}\xi(\tilde{x}), \quad B_{yl}^{(K)}(x) = B_{yl}\xi(\tilde{x}) \\ B_{ly}^{(K)}(x) &= \zeta(x)'B_{ly}, \quad u_{l,t}^{(K)}(x) = \zeta'(x)u_{\alpha,t}, \end{aligned} \tag{7}$$

where $\xi(x)$ is a second $K\times 1$ vector of basis functions and $u_{\alpha,t}$ is a $K\times 1$ vector of innovations. The matrix B_{ll} is of dimension $K\times K$, B_{yl} is of dimension $n_y\times K$, and B_{ly} is of dimension $K\times n_y$. Let $\boldsymbol{B}_{yl}^{(K)}[\cdot]$ and $\boldsymbol{B}_{ll}^{(K)}[\cdot](x)$ be the operators associated with the kernels $B_{yl}^{(K)}(x)$ and $B_{ll}^{(K)}(x,\tilde{x})$. Then we can write (3) as

$$\begin{aligned} \tilde{Y}_t &= B_{yy}\tilde{Y}_{t-1} + \boldsymbol{B}_{yl}^{(K)}[\tilde{\ell}_{t-1}^{(K)}] + u_{y,t}^{(K)} \\ \tilde{\ell}_t^{(K)}(x) &= B_{ly}^{(K)}(x)\tilde{Y}_{t-1} + \boldsymbol{B}_{ll}^{(K)}[\tilde{\ell}_{t-1}^{(K)}](x) + u_{l,t}^{(K)}(x). \end{aligned} \tag{8}$$

Combining (1), (5), (7), and (8) yields the following vector autoregressive system for the macroeconomic aggregates and the sieve coefficients (omitting K superscripts):

$$\begin{bmatrix} Y_t - Y_* \\ \alpha_t - \alpha_* \end{bmatrix} = \begin{bmatrix} B_{yy} & B_{yl}C_\alpha \\ B_{ly} & B_{ll}C_\alpha \end{bmatrix} \begin{bmatrix} Y_{t-1} - Y_* \\ \alpha_{t-1} - \alpha_* \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ u_{\alpha,t} \end{bmatrix}, \tag{9}$$

where $C_\alpha = \int \xi(\tilde{x})\zeta'(\tilde{x})d\tilde{x}$. Let $u_t' = [u_{y,t}', u_{\alpha,t}']$. We will subsequently assume that the innovations are Gaussian:

$$u_t \sim N(0,\Sigma). \tag{10}$$

2.3 Approximate Filtering

The K-dimensional approximation of the functional model has a state-space representation with state vector α_t, measurement equation (6), and state-transition equation (9). Due to the measurement equation for X_t, the state-space representation is nonlinear and the computation of the exact likelihood requires a nonlinear filter. To avoid the use of a nonlinear filter in the empirical application, we will conduct approximations justified by large-N approximations of Bayesian posteriors and marginal likelihoods.

We collect the parameters $(\alpha_*, Y_*, B, \Sigma)$ in the vector θ. Here the matrix B comprises B_{yy}, B_{yl}, B_{ly}, and B_{ll}. Although the dimension of θ depends on the degree of approximation K, we omit the (K) superscript. Let $Y_{1:t}$ denote the sequence $Y_1,\ldots,Y_t$. Starting from a

distribution $p^{(K)}(\alpha_{t-1}|Y_{1:t-1}, X_{1:t-1}, \theta)$, for each period t the filter computes:

$$\begin{aligned} p^{(K)}(Y_t, \alpha_t|Y_{1:t-1}, X_{1:t-1}, \theta) &= \int p^{(K)}(Y_t, \alpha_t|Y_{t-1}, \alpha_{t-1}, \theta) p^{(K)}(\alpha_{t-1}|Y_{1:t-1}, X_{1:t-1}, \theta) d\alpha_{t-1} \\ p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta) &= \int p^{(K)}(X_t|\alpha_t) p^{(K)}(Y_t, \alpha_t|Y_{1:t-1}, X_{1:t-1}, \theta) d\alpha_t \\ p^{(K)}(\alpha_t|Y_{1:t}, X_{1:t}, \theta) &\propto p^{(K)}(X_t|\alpha_t) p^{(K)}(Y_t, \alpha_t|Y_{1:t-1}, X_{1:t-1}, \theta). \end{aligned} \tag{11}$$

The first equation in display (11) iterates the state-transition equation (9) forward and integrates over the hidden state α_{t-1}. The second equation generates a forecast of the observables (Y_t, X_t) using the measurement equation (6). The third equation updates the density of the hidden state α_t using Bayes Theorem (here $\propto$ denotes proportionality). Note that the density $p^{(K)}(X_t|\alpha_t)$ defined in (6) does not depend on θ.

The state-transition density $p^{(K)}(Y_t, \alpha_t|Y_{t-1}, \alpha_{t-1}, \theta)$ is Normal under the assumption that the shock vector u_t is normally distributed. We will now approximate $p^{(K)}(X_t|\alpha_t)$ by conducting a second-order Taylor series expansion of $\mathcal{L}(\alpha_t|X_t)$ in (6) around the maximum likelihood estimator (MLE)

$$\hat{\alpha}_t = \text{argmax}_{\alpha_t} \, \mathcal{L}^{(K)}(\alpha_t|X_t). \tag{12}$$

Let $\hat{V}_t$ be the negative inverse Hessian associated with $\mathcal{L}(\alpha_t|X_t)$ evaluated at $\hat{\alpha}_t$. Then, we can write

$$p^{(K)}(X_t|\alpha_t) = \exp\left\{ N\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t) - \frac{N}{2}(\alpha_t - \hat{\alpha}_t)'\hat{V}_t^{-1}(\alpha_t - \hat{\alpha}_t) + N\mathcal{R} \right\}, \tag{13}$$

where $\mathcal{R}$ is the remainder term from the second-order Taylor series approximation. Abstracting from the remainder, the expression on the right-hand side of (13) is proportional to a measurement equation density from a model that treats the MLE $\hat{\alpha}_t(X_t)$ of the density coefficients as observables and uses the measurement equation

$$\hat{\alpha}_t(X_t) = \alpha_t + N^{-1/2}\eta_t, \quad \eta_t \sim N(0, \hat{V}_t). \tag{14}$$

in conjunction with the VAR in (9). By replacing (6) by (14) we obtain a linear Gaussian state-space model and the likelihood function can be evaluated with the Kalman filter.

To summarize, our large-N approximation of the filtering problem leads to a convenient two-step procedure. In the first step, the researcher computes the sequence of MLEs $\hat{\alpha}_t$ from the cross-sectional observations X_t, separately for each period $t = 1, \ldots, T$. In the second step, the researcher estimates a linear state-space model in which the $\hat{\alpha}_t$s are interpreted as noisy measures of the latent α_ts. Because the measurement error variance vanishes at rate N^{-1}, for large cross-sectional sample sizes, the estimation of a state-space model can be replaced by the estimation of a VAR in which the latent α_ts are replaced by their MLEs.

2.4 Bayesian Model Selection Criterion

In order to be able to determine the sieve order K in a data-driven way, we will now derive a large-N approximation to the Bayesian marginal data density (MDD) which for a sample of time dimension T is defined as

$$p^{(K)}(Y_{1:T}, X_{1:T}|\lambda) = \int \left(\prod_{t=1}^{T} p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta) \right) p(\theta|\lambda) d\theta, \tag{15}$$

where λ is a vector of hyperparameters that control the variance of the prior distribution. We will proceed by deriving a convenient representation for $p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta)$, which was defined in the second equation of display (11):

$$p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta) = \int p^{(K)}(X_t|\alpha_t) p^{(K)}(Y_t, \alpha_t|Y_{1:t-1}, X_{1:t-1}, \theta) d\alpha_t.$$

The density $p^{(K)}(X_t|\alpha_t)$ can be viewed as a likelihood function for α_t and $p(Y_t, \alpha_t|\cdot)$ characterizes the prior distribution for α_t. Note from the expansion in (13) that the likelihood function concentrates around $\hat{\alpha}_t$ as $N \longrightarrow \infty$. Using standard Bayesian asymptotic arguments, we obtain the Laplace approximation

$$\begin{aligned} p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta) &\approx \exp\left\{ N\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t) \right\} \left| \hat{V}_t/N \right|^{1/2} \\ &\quad \times (2\pi)^{K/2} p^{(K)}(Y_t, \hat{\alpha}_t|Y_{1:t-1}, X_{1:t-1}, \theta). \end{aligned} \tag{16}$$

The first term on the right-hand-side is the maximized likelihood function from the cross-sectional density estimation. The second term is the familiar penalty for model complexity which is asymptotically dominated by $N^{-K/2}$ and the last term can be interpreted as the prior density for α_t evaluated at $\hat{\alpha}_t$.

We now turn to $p^{(K)}(Y_t, \hat{\alpha}_t|Y_{1:t-1}, X_{1:t-1}, \theta)$ which is defined in the first equation of display (11). Standard Bayesian large-N asymptotics imply that

$$\alpha_{t-1}|(Y_{1:t-1}, X_{1:t-1}, \theta) \overset{\text{approx}}{\sim} N\big(\hat{\alpha}_{t-1}, \hat{V}_{t-1}/N\big). \tag{17}$$

By combining (17) with (9) and integrating out α_{t-1}, we can deduce that $p^{(K)}(Y_t, \hat{\alpha}_t|Y_{1:t-1}, X_{1:t-1}, \theta)$ is the density associated with the VAR

$$\begin{bmatrix} Y_t - Y_* \\ \hat{\alpha}_t - \alpha_* \end{bmatrix} = \begin{bmatrix} B_{yy} & B_{yl}C_\alpha \\ B_{ly} & B_{ll}C_\alpha \end{bmatrix} \begin{bmatrix} Y_{t-1} - Y_* \\ \hat{\alpha}_{t-1} - \alpha_* \end{bmatrix} + \begin{bmatrix} \tilde{u}_{y,t} \\ \tilde{u}_{\alpha,t} \end{bmatrix}, \tag{18}$$

where

$$\begin{bmatrix} \tilde{u}_{y,t} \\ \tilde{u}_{\alpha,t} \end{bmatrix} \sim N\big(0, \tilde{\Sigma}\big), \quad \tilde{\Sigma} = \Sigma + \frac{1}{N} \begin{bmatrix} B_{yl}C_\alpha \hat{V}_{t-1} C_\alpha' B_{yl}' & B_{yl}C_\alpha \hat{V}_{t-1} C_\alpha' B_{ll}' \\ B_{ll}C_\alpha \hat{V}_{t-1} C_\alpha' B_{yl}' & B_{ll}C_\alpha \hat{V}_{t-1} C_\alpha' B_{ll}' \end{bmatrix}.$$

Thus, replacing the latent α_t by the MLE $\hat{\alpha}_t$ added a $1/N$ term to the innovation covariance matrix, which we will ignore subsequently.[4] We deduce that we can write

$$p^{(K)}(Y_t, \hat{\alpha}_t | Y_{1:t-1}, X_{1:t-1}, \theta) \approx p^{(K)}(Y_t, \hat{\alpha}_t | Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \theta) \tag{19}$$

The approximation of the MDD can be completed as follows. Substituting (19) into the right-hand side of (16) and plugging the resulting expression into the marginal data density formula (15) leads to

$$\begin{aligned} p^{(K)}(Y_{1:T}, X_{1:T}|\lambda) &\approx (2\pi)^{KT/2} \left(\prod_{t=1}^{T} \exp\left\{ N \mathcal{L}^{(K)}(\hat{\alpha}_t | X_t) \right\} \left| \hat{V}_t / N \right|^{1/2} \right) \\ &\quad \times \int \left(\prod_{t=1}^{T} p^{(K)}(Y_t, \hat{\alpha}_t | Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \theta) \right) p(\theta|\lambda) d\theta. \end{aligned} \tag{20}$$

The MDD formula has two parts that can be evaluated independently. The expression in the first line of (20) can be viewed as a Bayesian information criterion for selecting the number of basis functions in the estimation of cross-sectional densities. An increase in the number of terms K improves the goodness-of-fit $\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t)$ but is penalized through the $O(N^{-K/2})$ determinant. The expression in the second line on the right-hand-side of the equation is the MDD associated with the VAR in (9) where the latent α_ts are replaced by the MLEs $\hat{\alpha}_t$. There exists a large literature on how to evaluate VAR MDDs either analytically or numerically for a variety of specifications and we will provide further details below. In Sections 4 and 5 we not only use the MDD to determine K, but also to determine the hyperparameter λ jointly with K.

3 Implementation Details

We now provide some of the implementation details. The choice of basis functions is described in Section 3.1 and some preliminary transformations of the estimated basis function coefficients is discussed in Section 3.2. Section 3.3 provides details on the specification of the measurement equation for the basis function coefficients and the state-transition equation of the empirical state-space model. Priors and the computation of posteriors for the parameters of the state-space model are discussed in Section 3.4 and Section 3.5 explains how forecasts and impulse response functions for the basis function coefficients can be converted back into cross-sectional densities.

[4] We consider applications in which the cross-sectional dimension N is much larger than the time series dimension T.

3.1 Basis Functions

A convenient basis for the log density is a spline which is a piecewise polynomial function with knots x_k, $k = 1, \ldots, K-1$. There is a large literature on log-spline density estimation dating back to Kooperberg and Stone (1990). A typical choice is to consider a cubic spline that is restricted to be linear and upward sloping on the interval $(-\infty, x_1)$ and linear and downward sloping on the interval $[x_{K-1}, \infty)$. Thus, the estimated density has the tails of a Laplace density, which are a bit thicker than Gaussian tails.

In our two application we restrict x to the interval $[0, \bar{x}]$. For the estimation based on simulated data in Section 4 we also use a cubic function for the last segment of the spline:

$$\begin{aligned} \zeta_1(x) &= x \\ \zeta_k(x) &= \left[\max\{x - x_{k-1},\, 0\}\right]^3, \quad k = 2, \ldots, K. \end{aligned} \tag{21}$$

We exclude the constant function $\zeta_0(x) = 1$ because it is redundant in light of the normalization imposed in the definition of $\mathcal{L}^K(\alpha_t|X_t)$ in (6). For the empirical application in Section 5 we construct the spline from $x = \bar{x}$ to $x = 0$, rather than from $x = 0$ to $x = \bar{x}$, using a linear element for the right tail

$$\begin{aligned} \zeta_K(x) &= \max\{\bar{x} - x,\, 0\} \\ \zeta_k(x) &= \left[\max\{x_{k-1} - x,\, 0\}\right]^3, \quad k = K-1, \ldots, 1. \end{aligned} \tag{22}$$

For both applications we compared three spline specifications for a given K: (21), (22), and the aforementioned linear-cubic-linear specification. For the simulated data (21) is preferred because the right tail of the simulated asset distributions is very thin. For the actual data it is desirable that the right-most segment of the spline is linear because under the cubic specification the density was increasing between the top coded value (see below) and the upper bound $\bar{x}$ in some periods. Near zero, on the other hand, the cubic segment was more successful approximating the density than the linear segment.

3.2 Sieve Coefficients

The first step in the estimation of the functional model is the computation of the MLEs $\hat{\alpha}_t$ for $t = 1, \ldots, T$. We discuss three adjustments that may be required to implement the empirical analysis.

Top Coding. In our empirical application the cross-sectional observations are top-coded. Because we standardize cross-sectional earnings by GDP per-capita which grows in real terms over time, the top-coded earnings value, which we take to be the maximum in every given time period t generally, falls over time until the statistical agency resets the maximum income level.

We define the censoring point $c_t = \max_{i=1,\ldots,N} x_{it}$ and let $N_{t,max} = \sum_{i=1}^{N} \mathbb{I}\{x_{it} = c_t\}$. If $N_{t,max} = 1$ we assume that the observed sample is not constrained by the top-coding and use the standard likelihood function described above. We introduce the unknown parameter $\pi_t = \mathbb{P}\{x_{it} \geq c_t\}$ and drop the top-coded observations from the definition of $\bar{\zeta}(X_t)$ by letting

$$\bar{\zeta}(X_t) = \frac{1}{N_t} \sum_{i=1}^{N} \zeta(x_{it}) \mathbb{I}\{x_{it} < c_t\}.$$

The log-likelihood function (divided by N_t, see (6)) under top coding is given by

$$\begin{aligned} \mathcal{L}^{(K)}(\alpha_t, \pi_t | X_t) &= \frac{N_{t,max}}{N_t} \ln \pi_t + \frac{N_t - N_{t,max}}{N_t} \ln(1 - \pi_t) \\ &\quad + \bar{\zeta}'(X_t)\alpha_t - \frac{N_t - N_{t,max}}{N_t} \ln \int_0^{c_t} \exp\left\{\zeta'(x)\alpha_t\right\}. \end{aligned} \tag{23}$$

The first line captures the likelihood of sampling top-coded observations and the second line represents the continuous density of the observations that are not top coded. Notice that log-likelihood function can be independently maximized with respect to π_t and α_t, where $\hat{\pi}_t = N_{t,max}/N_t$. The modified definition of $\mathcal{L}^{(K)}(\alpha_t, \hat{\pi}_t | X_t)$ requires a slight rescaling of $\hat{V}_t$. Further details are provided in the Online Appendix.

In our application, the top-coded values exceed the largest knot in every period t, i.e., $\bar{x}_t > x_{K-1}$ for all t, which means that all spline coefficients remain identifiable. If, on the other hand, the top-coded value is less than x_{K-1} in some period t, then some elements of the α_t vector are not identifiable from the cross-sectional information. These elements can be treated as missing values in the estimation of the linear state-space representation described in Section 3.4 below and will be implicitly imputed by the VAR law of motion (9).[5]

Compression. The vector $\hat{\tilde{\alpha}}_t = \hat{\alpha}_t - \alpha_*$ may exhibit collinearity. Even though K basis functions may be necessary to approximate the cross-sectional densities, the time variation might be concentrated in a lower-dimensional space, because, for instance, only the means

[5]There is a large literature on handling missing values in state-space models. Textbook treatments are available in Harvey (1989) and Durbin and Koopman (2001). An application to the estimation of mixed-frequency VARs is provided in Schorfheide and Song (2015).

of the cross-sectional distributions are varying over time. This feature can be captured by assuming that the time-variation is captured by a $\tilde{K} < K$ dimensional factor a_t:

$$(\alpha_t - \alpha_*)' = a_t'\Lambda, \tag{24}$$

where Λ is a $\tilde{K} \times K$ matrix. As is well known from the factor model literature, Λ and a_t are only identified up to a $\tilde{K} \times \tilde{K}$ dimensional invertible matrix. In principle, the matrix Λ and the sequence of vectors a_t, $t = 1, \ldots, T$ have to be estimated simultaneously under this factor structure,

To avoid the simultaneous estimation of the cross-sectional densities, we take the following short cut. First, we compute the $\hat{\alpha}_t$s period-by-period without imposing any restrictions. Second, conditional on α_* we compute the demeaned (and potentially seasonally adjusted) MLEs $\hat{\tilde{\alpha}}_t = \hat{\alpha}_t - \alpha_*$ and arrange them in a $T \times K$ matrix $\hat{\tilde{\alpha}}$ with rows $\hat{\tilde{\alpha}}_t'$. Third, we conduct a principal components analysis which is based on the eigenvalue decomposition of the sample covariance matrix $\hat{\tilde{\alpha}}'\hat{\tilde{\alpha}}/T$. Let $\hat{M}$ be $K \times \tilde{K}$ matrix of eigenvectors associated with the $\tilde{K}$ non-zero eigenvalues (in practice greater than 10^{-10}). Then, let

$$\hat{a} = \hat{\tilde{\alpha}}\hat{M}, \quad \hat{\Lambda} = (\hat{a}'\hat{a})^{-1}\hat{a}'\hat{\tilde{\alpha}}, \tag{25}$$

where $\hat{a}$ is the $T \times \tilde{K}$ matrix with rows $\hat{a}_t'$.

We can now replace (13) by[6]

$$p^{(K)}(X_t|a_t, \alpha_*, \hat{\Lambda}) = \exp\left\{N\mathcal{L}^{(K)}(\alpha_* + \hat{\Lambda}'\hat{a}_t|X_t) - \frac{N}{2}(a_t - \hat{a}_t)'\hat{\Lambda}\hat{V}_t^{-1}\hat{\Lambda}'(a_t - \hat{a}_t) + N\mathcal{R}\right\}.$$

To evaluate the MDD formula in (20), we replace $(2\pi)^{KT/2}$ by $(2\pi)^{\tilde{K}T/2}$, $\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t)$ by $\mathcal{L}^{(K)}(\alpha_* + \hat{\Lambda}'\hat{a}_t|X_t)$, and we change the penalty term from $|\hat{V}_t/N|^{1/2}$ to $|(\hat{\Lambda}\hat{V}_t^{-1}\hat{\Lambda}')^{-1}/N|^{1/2}$.

Seasonal Adjustments. In our empirical application x_{it} is based on quarterly earnings data from the Current Population Survey (CPS). Unlike the macroeconomic variables stacked in Y_t, the quarterly earnings data are not seasonally adjusted. Deterministic seasonal adjustments of the cross-sectional densities can be incorporated in the model by replacing the vector of constants $\alpha_* = \alpha_t - \tilde{\alpha}_t$ by a time-varying process. In our application the time period t is a quarter. We let $\alpha_{*,t} = \sum_{q=1}^{4} \alpha_{q,t}s_q(t)$, where $s_q(t) = 1$ if period t is associated with quarter q and $s_q(t) = 0$ otherwise.

[6]Because our goal is to eliminate perfect collinearities, we choose an eigenvalue cut-off that yields $\alpha_* + \hat{\Lambda}'\hat{a}_t = \hat{\alpha}_t$ in Sections 4 and 5.

3.3 State-Space Representation

After subtracting α_* (or $\alpha_{*,t}$ in case of the seasonal adjustment) and compressing the α_t coefficients, we can express the state-space representation in terms of the lower-dimensional vectors a_t and $\hat{a}_t$. The measurement equation (14) is replaced by

$$\hat{a}_t = a_t + N^{-1/2}\eta_t, \quad \eta_t \sim N\big(0, (\hat{\Lambda}\hat{V}_t^{-1}\hat{\Lambda}')^{-1}\big). \tag{26}$$

The state transition is essentially given by (9) but we need to adjust it for the compression of the α_t vector. Moreover, we now absorb the matrix C_α into the matrices of regression coefficients:

$$\begin{bmatrix} Y_t - Y_* \\ a_t \end{bmatrix} = \begin{bmatrix} \Phi_{yy} & \Phi_{ya} \\ \Phi_{ay} & \Phi_{aa} \end{bmatrix} \begin{bmatrix} Y_{t-1} - Y_* \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ u_{a,t} \end{bmatrix}. \tag{27}$$

We assume that the innovations are normally distributed and write the state transition more compactly as

$$W_t = \Phi_1 W_{t-1} + u_t, \quad u_t \sim N(0, \Sigma), \tag{28}$$

where $W_t = [(Y_t - Y_*)', a_t']'$.

3.4 Priors and Posteriors

The estimation of the state-space model is done conditional on the sequence $\hat{V}_t$, $t = 1, \ldots, T$, and the compression matrix $\hat{\Lambda}$. Moreover, in our implementation we also condition on estimates of the deterministic components, $\hat{Y}_*$ and $\hat{a}_*$. Thus, the remaining unknown coefficients are concentrated in the state-transition equation (28), which takes the form of a multivariate linear Gaussian regression model. The state transition can be expressed in matrix form as

$$W = Z\Phi + U, \tag{29}$$

where W, Z, and U have rows W_t', W_{t-1}', and u_t', respectively and for a VAR(1) without intercept $\Phi = \Phi_1'$. Defining $\phi = \text{vec}(\Phi)$ we use a prior distribution of the form

$$\Sigma \sim IW(\underline{\nu}, \underline{S}), \quad \phi|\lambda \sim N\big(\underline{\mu}_\phi, \underline{P}_\phi^{-1}(\lambda)\big), \tag{30}$$

where $IW(\cdot)$ is the Inverse-Wishart distribution with $\underline{\nu}$ degrees of freedom and scale matrix $\underline{S}$.

The prior precision matrix $\underline{P}_\phi(\lambda)$ is a function of a vector of hyperparameters $\lambda = [\lambda_1, \lambda_2, \lambda_3]'$ and takes the form

$$\underline{P}_\phi(\lambda) = \lambda_1 \begin{bmatrix} (\underline{\Sigma}^{-1})_{yy} \otimes \begin{bmatrix} \hat{D}_y & 0 \\ 0 & \lambda_2 \hat{D}_a \end{bmatrix} & (\underline{\Sigma}^{-1})_{ya} \otimes \begin{bmatrix} \sqrt{\lambda_3}\hat{D}_y & 0 \\ 0 & \sqrt{\lambda_2}\hat{D}_a \end{bmatrix} \\ (\underline{\Sigma}^{-1})_{ay} \otimes \begin{bmatrix} \sqrt{\lambda_3}\hat{D}_y & 0 \\ 0 & \sqrt{\lambda_2}\hat{D}_a \end{bmatrix} & (\underline{\Sigma}^{-1})_{aa} \otimes \begin{bmatrix} \lambda_3 \hat{D}_y & 0 \\ 0 & \hat{D}_a \end{bmatrix} \end{bmatrix}. \quad (31)$$

The partitions of Σ^{-1} conform with the partition $W_t' = [(Y_t - Y_*)', a_t']$. The matrices $\hat{D}_y$ and $\hat{D}_a$ are diagonal matrices of dimension $n_y \times n_y$ and $\tilde{K} \times \tilde{K}$ that are used to rescale the prior variances based on the variability of the regressors. For instance, one can set the jj element of $\hat{D}_y$ equal to a (pre-)sample variance $\hat{\sigma}^2(Y_{jt})$. This scaling is common for Minnesota-type VAR priors.

The hyperparameter λ_1 controls the overall precision of the prior distribution; λ_2 scales the relative precision of the prior distribution for the coefficients that control the effect of a_{t-1} on Y_t; likewise, λ_3 scales the relative precision of the prior distribution for the coefficients that control the effect of Y_{t-1} on a_t. Unlike the more commonly used matrix-Normal Inverse-Wishart prior that mimics the Kronecker structure of the likelihood function, the prior in (30) allows us to control the degree of spillovers from distributional dynamics to the aggregate dynamics and vice versa. If the prior mean $\underline{\mu}_\phi$ is zero, then as $\lambda_2, \lambda_3 \longrightarrow \infty$, the posterior distributions of Φ_{ay} and Φ_{ya} concentrate around zero, which shuts down spillover effects.

Conditional on λ, it is straightforward to sample from the posterior distribution of (ϕ, Σ) using a Gibbs sampler following the approach in Carter and Kohn (1994) that iterates over the blocks:

$$\phi|(\Sigma, W_{1:T}, \lambda), \quad \Sigma|(\phi, W_{1:T}, \lambda) \quad a_{1:T}|(\phi, \Sigma, \tilde{Y}_{1:T}, \hat{a}_{1:T}). \quad (32)$$

Here it is important to recall that $W_{1:T} = [\tilde{Y}_{1:t}, a_{1:T}]$ and note that conditional on $a_{1:T}$ the MLEs $\hat{a}_{1:T}$ do not contain any information about (ϕ, Σ). The Online Appendix describes how to sample from $\phi|(\Sigma, W_{1:T}, \lambda)$ and $\Sigma|(\phi, W_{1:T}, \lambda)$. Sampling from $a_{1:T}|(\phi, \Sigma, \tilde{Y}_{1:T}, \hat{a}_{1:T})$ is implemented by a standard forward filtering and backward simulation approach that utilizes the Kalman filter in the forward iteration and the simulation smoother for the backward simulation. The Online Appendix also discusses how the second part of the MDD formula – the one associated with the VAR in (29), see the second line on the right-hand side of (20) – can be evaluated using a combination of the Rao-Blackwellization approach in Fuentes-Albero and Melosi (2013) and Geweke (1999)'s modified harmonic mean estimator.

3.5 Recovering Cross-Sectional Densities

Based on the estimated state-transition equation (27) we can generate forecasts and impulse response functions for the compressed coefficients a_t. However, the dynamics of these coefficients in itself are not particularly interesting. Thus, we have to convert them back into densities using the following steps (which can be executed for each draw of a_t from the relevant posterior distribution). First, use (24) with $\Lambda = \hat{\Lambda}$ to transform a_t into α_t. If the estimation is based on a seasonal adjustment, α_* can be replaced by $\alpha_{*,t}$, or, if the goal is to compute impulse responses, one could use the average of the seasonal dummies as intercept. Second, compute

$$p^{(K)}(x|\alpha_t) = \frac{\exp\left\{\zeta'(x)\alpha_t\right\}}{\int \exp\left\{\zeta'(\tilde{x})\alpha_t\right\} d\tilde{x}}.$$

4 A Simulation Experiment

To examine our functional state-space model's ability to capture the joint dynamics of aggregate variables and a cross-sectional distribution, we first estimate it based on artificial data simulated from a Krusell and Smith (1998) economy. The model economy and its approximate solution is described in Section 4.1. The estimation results are summarized in Section 4.2.

4.1 Model Economy and Data Generating Process

The model economy consists of a continuum of households $j \in [0, 1]$. Household j chooses consumption and asset holdings to maximize

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \frac{c_{jt}^{1-\sigma} - 1}{1-\sigma} \right]$$

subject to the budget constraint

$$x_{jt+1} + c_{jt} = (1-\tau)W_t\epsilon_{jt} + bW_t(1-\epsilon_{jt}) + x_{jt}(1+R_t), \quad x_{jt+1} \geq \underline{x}. \tag{33}$$

Here $\epsilon_{jt} \in \{0, 1\}$ is an exogenous two-state Markov process that determines the efficiency units of labor supplied by the household j in period t. Households with $\epsilon_{jt} = 1$ receive after-tax labor income $(1-\tau)W_t$ and households with $\epsilon_{jt} = 0$ receive unemployment benefits bW_t. The total labor supply $L = \int \epsilon_{jt} dj$ is fixed over time. We assume that the government

balances its budget constraint in each period by setting $\tau = b(1-L)/L$. The asset x_{jt} is a claim on the aggregate capital stock and generates a risky return R_t and households face the borrowing constraint $x_{jt+1} \geq \underline{x}$.

The representative firm produces output Y_t according to the production function

$$Y_t = \exp\{z_t\} K_t^{\alpha} L^{1-\alpha}, \quad z_t = \rho_z z_{t-1} + \sigma_z \omega_t, \quad \omega_t \sim N(0,1),$$

where z_t is an exogenous aggregate productivity shock that follows an AR(1) law of motion, and K_t is the aggregate capital stock. Factor prices are given by

$$R_t = \alpha \exp\{z_t\} K_t^{\alpha-1} L^{1-\alpha} - \delta, \quad W_t = (1-\alpha) \exp\{z_t\} K_t^{\alpha} L^{-\alpha}, \tag{34}$$

where δ is the depreciation rate of capital. In equilibrium, the net supply of assets equals the capital stock: $K_t = \int x_{jt} dj$.

The aggregate state of the economy is $S_t = (z_t, \mu_t)$, where μ_t is the distribution of households over (ϵ_{jt}, x_{jt}) pairs. Because ϵ_{jt} takes only two values, it is convenient to use $\mu_{t,\epsilon}$ to denote the conditional distribution of assets given the employment status ϵ_t. To simulate data from the model economy, we construct an approximate solution in which the density associated with $\mu_{t,\epsilon}$ can be written as a mixture of a discrete and continuous part:

$$q_{t,\epsilon}(x) = \widehat{m}_{t,\epsilon} \Delta_{\underline{x}}(x) + (1 - \widehat{m}_{t,\epsilon}) p_{t,\epsilon}(x). \tag{35}$$

Here, $\widehat{m}_{t,\epsilon}$ is the mass of individuals for whom the borrowing constraint $\underline{x}$ is binding. $\Delta_{\underline{x}}(x)$ is the Dirac function with the property that $\Delta_{\underline{x}}(x) = 0$ for $x \neq \underline{x}$ and $\int \Delta_{\underline{x}}(x) dx = 1$. This function captures the mass of households for which the borrowing constraint is binding. The continuous part of the asset distribution is represented by the (proper) density $p_{t,\epsilon}(x)$.

Following Winberry (2018) we represent the density $p_{t,\epsilon}(x)$ as follows:

$$p_{t,\epsilon}(x) = \exp\left\{ \gamma_{t,\epsilon,0} + \gamma_{t,\epsilon,1}(x - m_{t,\epsilon,1}) + \sum_{k=2}^{3} \gamma_{t,\epsilon,k} \left[(x - m_{t,\epsilon,1})^k - m_{t,\epsilon,k} \right] \right\}. \tag{36}$$

Here the $m_{t,\epsilon,k}$s are the time-varying centralized moments of the distribution. The $\gamma_{t,\epsilon,k}$s are a set of time-varying coefficients that can be determined from the moments by solving the system of equations

$$m_{t,\epsilon,1} = \int x p_{t,\epsilon}(x) dx; \quad m_{t,\epsilon,k} = \int (x - m_{t,\epsilon,1})^k p_{t,\epsilon}(x) dx, \; k = 2, 3; \quad \int p_{t,\epsilon}(x) dx = 1.$$

The desired consumption of household j in absence of the borrowing constraint is given by

$$c_{jt}^{*} = \left[\beta \mathbb{E}_t[(1 + R_{t+1}) c_{jt+1}^{-\sigma}]\right]^{-1/\sigma}.$$

According to (33) the return R_{t+1} depends on the capital stock K_{t+1}, which in turn depends on the cross-sectional distribution of assets:

$$K_{t+1} = \sum_{h=0,1} \mathbb{P}\{\epsilon = h\} \left[\underline{x} \cdot \widehat{m}_{t+1,\epsilon=h} + (1 - \widehat{m}_{t+1,\epsilon=h}) \int x p_{t+1,\epsilon=h}(x) dx\right].$$

Actual consumption may be lower than desired consumption of the borrowing constraint is binding. Asset holdings in period $t+1$ can be determined from the budget constraint (33). Doing this for each household j determines the $t+1$ distribution of asset holdings. Thus, the model generates a joint autoregressive law of motion for

$$\varsigma_t = \left(z_t, \widehat{m}_{t,\epsilon=0}, \widehat{m}_{t,\epsilon=1}, m_{t,\epsilon=0}, m_{t,\epsilon=1}\right), \tag{37}$$

where $m_{t,\epsilon} = [m_{t,\epsilon,1}, m_{t,\epsilon,2}, m_{t,\epsilon,3}]'$ represents $p_{t,\epsilon}(x)$.

We use the method of Winberry (2018) to solve the KS economy. The method involves two steps. First, we construct a nonlinear solution for the version of the model without aggregate uncertainty. To do so, we replace the unknown decision rule for (transformed) desired consumption $(-1/\sigma) \ln c_{jt}^{*}$ with a Chebychev polynomial and integrals through quadrature approximations. This leads to a rational expectations system in ς_t, (K_t, R_t, W_t), the Chebychev polynomial coefficients, the quadrature approximation points of the asset distribution, and the density evaluated at these points.[7] The system is then linearized around the steady state and solved with a standard solver for linearized rational expectations (LRE) systems. The solution to the LRE system can be written as a VAR in ς_t, defined in (37), and a set of equations that relate the remaining variables to the state vector ς_t. To generate data, we simulate the law of motion for ς_t for 2,000 periods and draw *iid* cross-sectional observations from (35), where $p_{t,\epsilon}(x)$ is defined in (36).

We calibrate the KS economy to loosely match features of annual U.S. data. The parameterization is summarized in Table 1. The left panel of Figure 1 depicts the time series of the aggregate capital stock from $t = 1$ to $t = 200$. The capital stock peaks in period $t = 27$ and reaches a trough in period $t = 150$. The densities of asset holdings (normalized to one) for the unemployed and employed households are plotted in the center panel and the right panel of the figure. We condition on $x > \underline{x} = 0$ and plot $p_{t,\epsilon}(x)$ given in (36). The

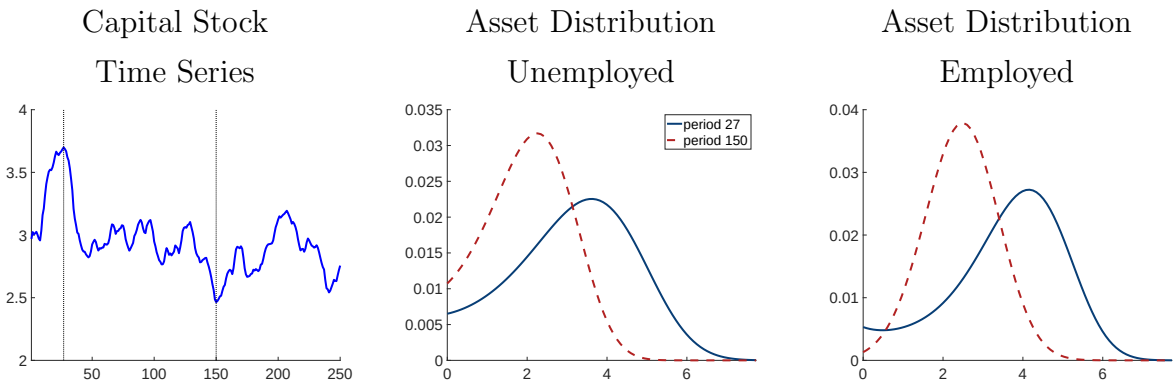
[7] For the subsequent simulations we use 25 quadrature nodes and a Chebychev polynomial of order three.

Table 1: Calibration of Krusell-Smith Economy

Parameter		Value	Parameter		Value
β	Discount factor	.93	b	Unempl. benefits	.15
σ	Utility curvature	1	$\pi(\epsilon_0 \to \epsilon_1)$	Unempl. to Empl.	.5
$\underline{x}$	Borrowing constraint	0	$\pi(\epsilon_1 \to \epsilon_0)$	Empl. to Unempl.	.038
α	Capital share	.36	ρ_z	TFP Persistence	.859
δ	Capital depreciation	.10	σ_z	TFP Innovation StdDev	.028

Notes: Annual parameterization

Figure 1: Features of Simulated Data



Notes: Aggregate capital attains its maximum in period $t = 27$ and its minimum in period $t = 150$. Each density is normalized so that it integrates to one.

densities illustrate that employed households hold more assets than unemployed households and that in a boom period (high capital stock due to favorable technology shocks) the asset distribution shifts to the right.

4.2 Functional Model Estimation

We now estimate the functional model based on the simulated time series and cross-sectional data. Because according to the calibration in Table 1 the employment rate is 93%, we focus on the asset distribution of the employed households and omit information about the assets of the unemployed from the estimation. Because due to precautionary savings behavior the mass of employed individuals with zero assets is essentially zero and does not vary over time, the aggregate variable in our estimation is simply log productivity z_t and we provide the

Table 2: Knot Placement

	Percentiles																				
K	0.01	0.025	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
4							X					X					X				
6				X			X					X					X			X	
8			X	X			X					X					X			X	X
10	X	X	X	X			X					X					X			X	X
14	X	X	X	X	X		X		X			X			X		X		X	X	X
22	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Notes: Percentiles refer to distribution of pooled observations (across $i = 1, \ldots, N$ and $t = 1, \ldots, T$) for N=10,000 and T=2,000.

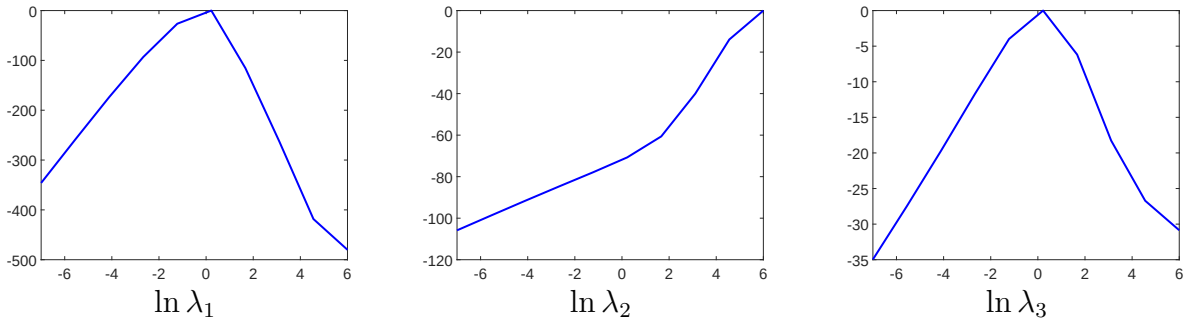
econometrician with iid draws from the cross-sectional distribution $x_{it} \sim p_{t,\epsilon=1}(x)$, where $p_{t,\epsilon=1}(x)$ was defined in (36).

The empirical model differs from the data generating process (DGP) in the following dimensions. First, strictly speaking, the DGP takes the form of a first-order autoregressive process only if the asset distribution of the unemployed is also included in the vector of endogenous variables. Second, the representation of the densities in the DGP are based on polynomial basis functions whereas the density in the empirical model is based on cubic splines. Third, the DGP is based on a VAR that includes the moments $m_{t,\epsilon,k}$ whereas the empirical model is based on a VAR that includes the spline coefficient vector a_t. The relationship between the $m_{t,\epsilon,k}$s and the polynomial basis function coefficients $\gamma_{t,\epsilon,k}$ in the DGP is nonlinear. Overall, for a sufficiently large K, we expect the empirical model to deliver a good approximation to the DGP, but for any given K it does not nest the DGP.

The first step in the analysis is the determination of the approximation order K, and the hyperparameter λ based on the MDD approximation. We use the spline basis in (21) and place the knots at pre-determined percentiles of the empirical distribution of the simulated x_{it}s, pooled across a large simulation with $N_{sim} = 10,000$ and $T_{sim} = 2,000$. Table 2 summarizes the knot locations as a function of K. For $K = 4$, we use the 25^{th}, 50^{th}, and 75^{th} percentiles. As K increases, we add lower and upper percentiles. Moving from $K = 8$ to $K = 10$, we only add percentiles in the left tail of the distribution, because this part of the distribution is most affected by business cycle variations.

For the subsequent estimation we fix the time series dimension at $T = 250$ and consider three choices of N: 1,000, 5,000, and 10,000. Conditional on N we evaluate the approximate

Figure 2: Normalized Log Marginal Data Density



Notes: Based on N=10,000, T=250, and K=10. The log MDD peaks at $\hat{\lambda}_1 = \exp(0.22) = 1.25$, $\hat{\lambda}_2 = \exp(6) = 403.43$, and $\hat{\lambda}_3 = \exp(0.22) = 1.25$. Each panel shows a one-dimensional cut of the log MDD surface fixing the other two λ values at the argmax. The x-axis is on a log scale. We normalize the maximum value of the log MDD to zero.

marginal data density $p^{(K)}(Y_{1:T}, X_{1:T}|\lambda)$ in (20) over a grid of K and λ values.[8] For each element of the λ vector, we consider 10 equally-spaced values of $\ln \lambda_j$ between -5 and 6. For each K we maximize the MDD with respect to λ to obtain $\hat{\lambda}(K, N)$. Figure 2 shows three cuts of the log MDD as a function of λ. The figure is based on $K = 10$ and $N = 10,000$. In the left panel we vary λ_1 and keep λ_2 and λ_3 fixed at $\hat{\lambda}_2$ and $\hat{\lambda}_3$, respectively. In the remaining panels we vary the other two λ elements. The maximum of the log MDD function is normalized to zero. The log MDD is quite sensitive to λ_1 and the maximum with respect to λ_2 is attained at the boundary of the grid.

In Table 3 we report $\hat{\lambda}(K, N)$ for various choices of K and the cross-sectional sample size N. The table also contains the values of the log MDD at $\hat{\lambda}(K, N)$. Rather than reporting raw log MDD values, in each column we report log differentials with respect to $\hat{\lambda}(K = 4)$. Several important results emerge. First, as N increases from 1,000 to 5,000 or 10,000 the selected dimension of the spline approximation increases from 6 to 10 because the improvement in the fit of the cross-sectional densities outweighs the dimensionality penalty. Because for any fixed K the log-spline density specification does not nest the true cross-sectional density, there is no "true" K in this simulation design.

Second, the hyperparameter λ_1 controls the overall precision (inverse variance) of the prior distribution for the vector autoregressive coefficients ϕ. The $\hat{\lambda}_1(K)$ are weakly increas-

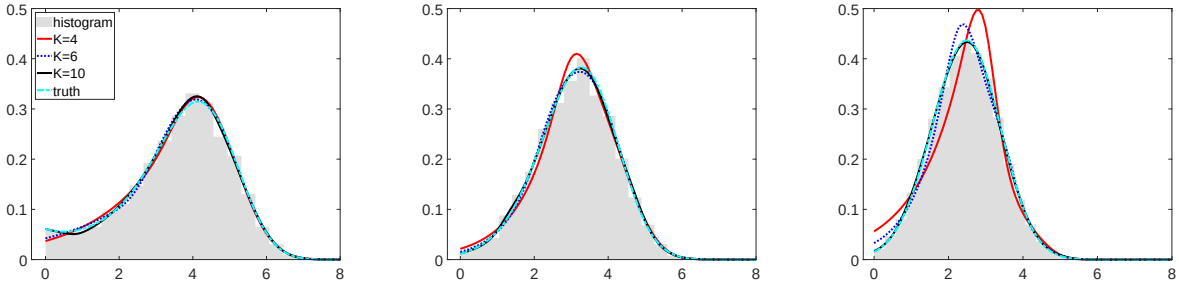
[8]For each (K, λ) combination we generate 5,500 draws from the posterior distribution $(\phi, \Sigma)|(\tilde{Y}_{1:T}, \hat{a}_{1:T})$ using a Gibbs sampler. After discarding the first 500 draws we use a modified harmonic mean estimator to construct a numerical approximation of $p^{(K)}(\tilde{Y}_{1:T}, \hat{a}_{1:T}|\lambda)$. Details are provided in the Appendix.

Table 3: Hyperparameter Estimates and Log MDD Differentials for $T = 250$

	N=1,000				N=5,000				N=10,000			
K	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	MDD	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	MDD	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	MDD
4	1.2	403	5.3	0	0.3	403	5.3	0	0.1	403	5.3	0
6	**1.2**	**403**	**5.3**	**1,374**	0.3	403	5.3	7,300	0.3	403	5.3	14,978
8	5.3	403	1.2	978	1.2	403	1.2	8,137	0.3	403	5.3	16,853
10	5.3	403	1.2	915	**1.2**	**403**	**1.2**	**8,187**	**1.2**	**403**	**1.2**	**17,039**
14	5.3	403	1.2	795	5.3	403	0.3	7,908	5.3	403	0.1	16,740

Notes: For each K we maximized the MDD with respect to λ_1, λ_2, and λ_3. The table reports $\hat{\lambda}$ estimates for each (K, N) pair. The log MDD differentials are computed with respect to $K = 4$.

Figure 3: Cross-sectional Fit ($N = 10,000, T = 250, K = 10$ Selected)

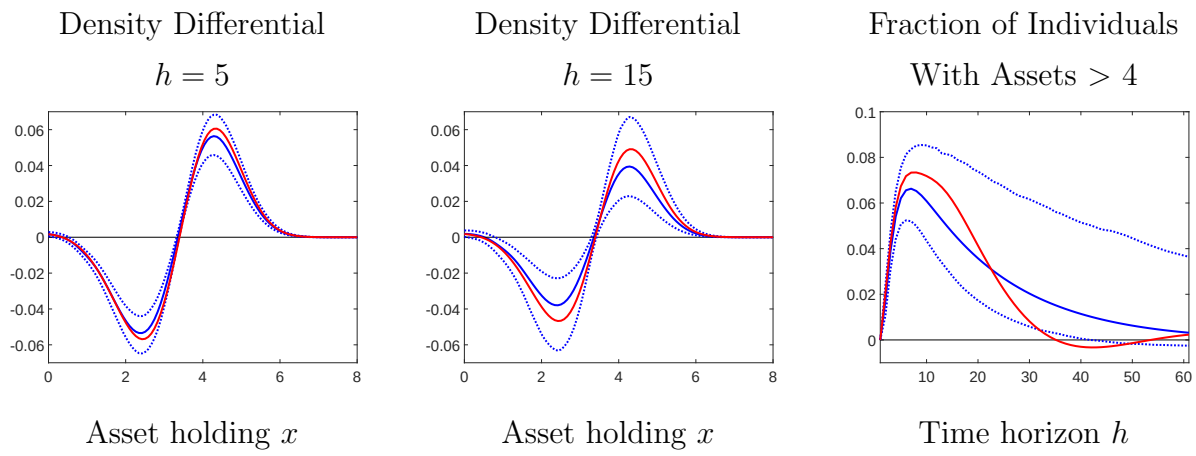


Notes: Each panel overlays: a histogram of the empirical asset distribution of the employed, the log-spline density estimates for various K, and the true density from which the asset values were sampled.

ing in K, indicating that as the dimension of the vector $W_t' = [(Y_t - Y_*)', a_t']$ increases more shrinkage toward the prior mean is desirable. Third, for every (K, N) combination $\hat{\lambda}_2 = 403$ which is the maximum value of the λ grid. Recall that λ_2 scales the relative precision of the prior distribution for the coefficients that control the effect of the cross-sectional density represented by a_{t-1} on the aggregate variable Y_t. A value of $\hat{\lambda}_2 = 403$ implies that the resulting estimate of the spill-over effect is essentially zero. The estimate $\hat{\lambda}_3$ is substantially smaller than $\hat{\lambda}_2$. Thus, we deduce that based on the estimated model aggregate dynamics spill into distributional dynamics, but not vice versa. This is consistent with the KS model economy in which technology is exogenous, but productivity fluctuations affect the asset distribution.

In Figure 3 we compare the fit of the estimated cross-sectional densities (various K) for

Figure 4: Impulse Responses to a 3-Standard-Deviation Technology Shock



Notes: Impulse responses are generated from the estimated state-space representation. Results are based on $T = 250$, $N = 10,000$, $K = 10$, and $\hat{\lambda}(K)$ (see Table 3). "True" response from the KS model appears in red, responses from the estimated functional state-space model are in blue (pointwise median is solid, 5th and 95th percentiles are dotted).

$N = 10,000$

$$p^{(K)}(x|\hat{\alpha}_t) = \exp\left\{\zeta'(x)\hat{\alpha}_t - \ln \int \exp\left\{\zeta'(\tilde{x})\hat{\alpha}_t\right\} d\tilde{x}\right\} \tag{38}$$

to the "true" cross-sectional density and a histogram of the cross-sectional observations. The three panels correspond to different time periods. Recall from Figure 1 that the capital stock peaks in period 27 and reaches a trough in period $t = 150$. The fit of the estimated densities varies over time as the shape of the "true" density changes. For $K = 4$ and $K = 6$ there is some mismatch in the left tail in periods $t = 27$ and $t = 150$. The fit generally improves as K increases and for the selected $K = 10$ we essentially have a perfect fit.

We proceed by estimating the state-space representation for $K = 10$ and $\lambda = \hat{\lambda}(10)$ on the $N = 10,000$ sample using the Gibbs sampler described in Section 3.4.[9] Figure 4 shows impulse responses to a 3-standard-deviation technology shock. We compute "true" responses from the approximate solution of the KS economy and estimated responses from the state-space representation given by (26) and (27). Because the technology process is exogenous in the KS economy, its innovation can be easily identified using a Cholesky factorization of Σ. For each (Φ, Σ) we generate an impulse response function for a_t by iterating (27) forward. Once we have the impulse response sequence $\{a_{t+h}\}_{h=1}^{H}$ we reconstruct the density $p^K(x|\alpha_{t+h})$ by reversing the compression step in Section 2.2 to transform $\{a_{t+h}\}_{h=1}^{H}$ into $\{\alpha_{t+h}\}_{h=1}^{H}$ and

[9] We generate 11,000 draws from the posterior and drop the first 1,000.

plugging the coefficients into (38).

The "shocked" density is compared to the steady state density that is obtained by using the α_* coefficients. The left and center panels of Figure 4 show the difference between the shocked and the steady state densities at horizons $h = 5$ and $h = 15$, respectively.[10] We refer to this difference as the response of the density. In the right panel of Figure 4, we show the response of the fraction of individuals with assets greater than 4 as a function of h. This fraction is computed from the density response. To capture parameter uncertainty, we execute the calculations for each draw (Φ, Σ) from the posterior distribution and plot (in blue) the posterior median responses and bands that represent pointwise 5 and 95 percentiles. The "true" responses from the KS model are plotted in red.

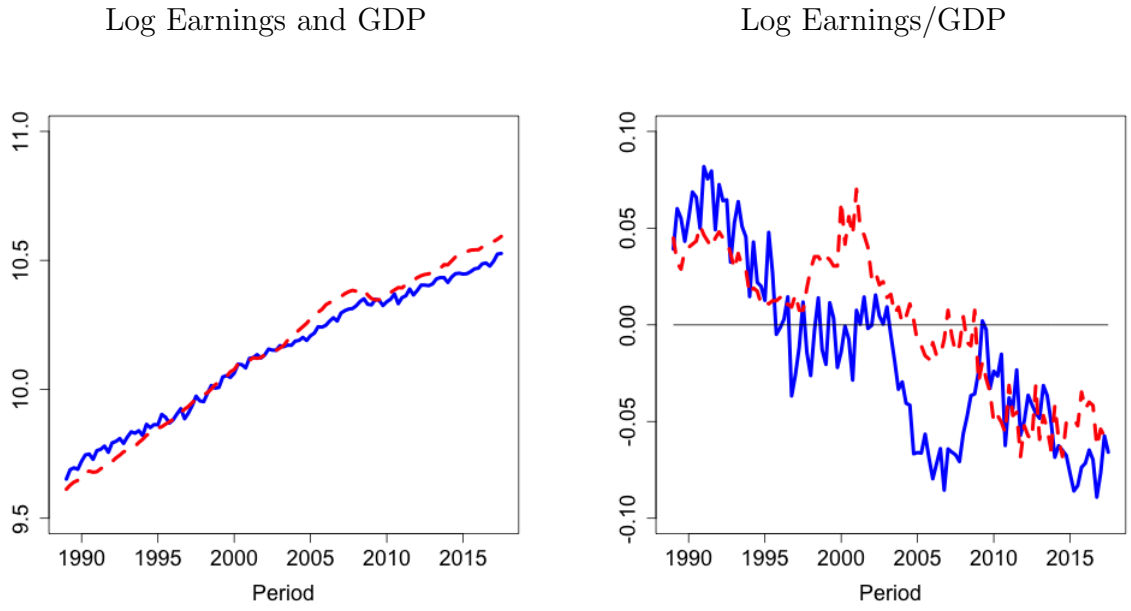
According to the KS model, a positive technology shock increases asset holdings and shifts the asset distribution to the right. In the left and center panels of the figure this is reflected in the increase of the density for x-values above 3.3 and a drop for asset holdings below 3.3. Moreover, it can be seen in the right panel, that the fraction of individuals with asset holdings greater than 4 increases by roughly six percentage points (0.06). The estimated responses capture the "true" dynamics well. The posterior median density response decays slightly slower than the "true" response, but, accounting for parameter uncertainty, the bands that delimit pointwise 90% credible bands cover the "true" response. To summarize, the estimated functional state-space model is successful in capturing the joint dynamics of the cross-sectional asset distribution and the technology shock.

5 Empirical Analysis

The empirical analysis focuses on the joint dynamics on total factor productivity, real per-capita GDP, and employment at the aggregate level, and the cross-sectional distribution of earnings at the micro level. Estimation results for the functional state-space model are presented in Section 5.1. In Section 5.2 we report impulse responses of aggregate variables, the cross-sectional distribution, and inequality measures derived from the cross-sectional distribution to aggregate shocks. Two robustness exercises are conducted in Sections 5.3 and 5.4. Finally, in Section 5.5 we examine the response to a distributional shock that raises inequality.

[10] Timing convention: the system is in steady state in period $h = 1$ and the shock occurs at $h = 2$.

Figure 5: Earnings and GDP



Notes: Left panel: average log earnings (blue, solid) and log per capita GDP (red, dashed). Right panel: average log earnings-to-GDP ratio (blue, solid) and demeaned log labor share (red, dashed) of the nonfarm business sector. In both panels per-capita GDP is scaled by 2/3 to account for the labor share.

5.1 Data and Model Estimation

Data. We use three macroeconomic aggregates in our empirical analysis: total factor productivity (TFP), real per-capita GDP, and the employment rate. In addition, we use cross-sectional data on earnings. Real per-capita GDP (*A939RX0Q048SBEA*) is provided by the Federal Reserve Bank of St. Louis' FRED database and the TFP series (*dtfp*) is obtained from Fernald (2012). Weekly earnings (*PRERNWA*) are obtained from the monthly Current Population Survey (CPS) through the website of the National Bureau of Economic Research (NBER). Weekly earnings are scaled to annual earnings by multiplying with 52. Based on the CPS variable *PREXPLF* "Experienced Labor Force Employment" we construct an employment indicator which is one if the individual is employed and zero otherwise. This indicator is used to compute the aggregate employment rate.

In the left panel of Figure 5 we plot average log nominal earnings computed from the cross-sectional data and log nominal per-capita GDP. We scale per-capita GDP by a factor of 2/3 to account for the labor share.[11] After this re-scaling the mean of log earnings and log

[11]Nominal per-capita GDP is obtained by multiplying real per-capita GDP by the GDP deflator

per-capita GDP have approximately the same level. However, the mean log earnings grow more slowly than per-capita GDP. In the right panel of the Figure we plot the average log earnings-to-GDP ratio (here per-capita GDP is again scaled by 2/3) and the demeaned log labor share of the nonfarm business sector (obtained from the Bureau of Labor Statistics). The drop in the log earnings-to-GDP ratio is of the same order of magnitude as the fall in the labor share over the sample period.

In the remainder of this paper we simply standardize individual-level earnings by (2/3) of nominal per-capita GDP. Rather than taking a logarithmic transformation of the earnings data, we apply the inverse hyperbolic sine transformation, which is given by

$$x = g(z|\theta) = \frac{\ln(\theta z + (\theta^2 z^2 + 1)^{1/2})}{\theta} = \frac{\sinh^{-1}(\theta z)}{\theta}, \quad z = \frac{\text{Earnings}}{(2/3) \cdot \text{per-capita GDP}}. \tag{39}$$

The function is plotted in the Online Appendix. We set $\theta = 1$. For small values of z the function is approximately equal to z and for large values of z it is equal to $\ln(z) + \ln(2)$. This transformation avoids the thorny issue of applying a log transformation to earnings that are close to zero. Below we will refer to x as transformed data and to z as original data.

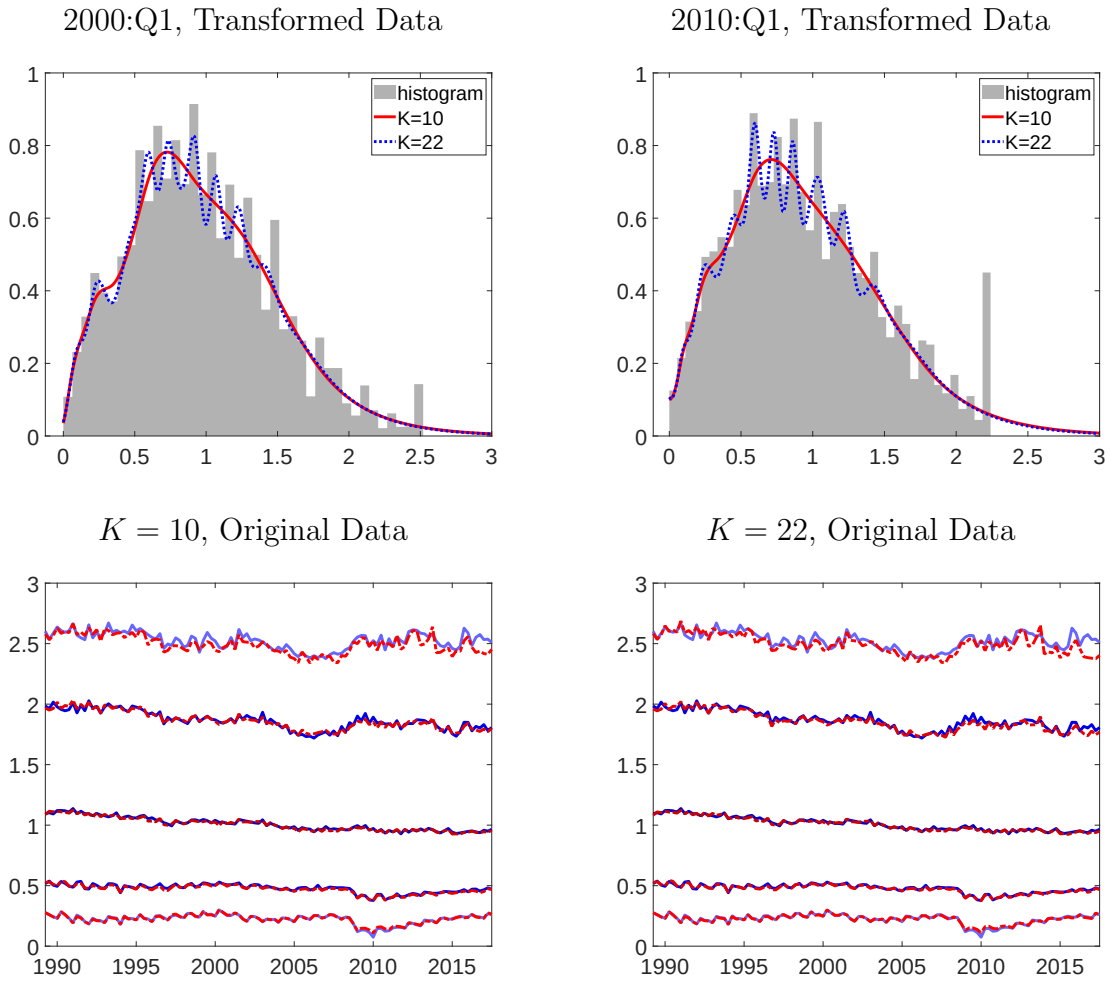
Density Estimation. We take the time period t to be a quarter. For each t we estimate a cross-sectional density for the transformed earnings-to-GDP ratio; see (39). As in the simulation study in Section 4, we consider different approximation orders K. We use the spline basis in (22) and place the knots at pre-determined percentiles of the empirical distribution of the x_{it}s pooled across i and t; see Table 2.

In the top row of Figure 6 we show three types of density estimates for the transformed earnings in 2000:Q1 and 2010:Q1: log spline estimates for $K = 10$ and $K = 22$, and histograms. While the $K = 10$ density estimates are smooth, the $K = 22$ estimates capture the jaggedness of the histograms. The spike in the right tail visible in the histograms captures the top coding. It is more pronounced in 2010:Q1 because incomes rise over time and more households reach the top income level (until the top values gets reset by the statistical agency). By construction, the log-spline density estimates extrapolate the top-coded income values.

In the bottom row of the figure we overlay the sample percentiles of the earnings/GDP distribution and percentiles computed from the log spline density estimates $p^{(K)}(x|\hat{\alpha}_t)$ for $K = 10$ and $K = 22$. The earnings/GDP distribution has a pointmass at zero, representing

(*GDPDEF* from FRED). The factor 2/3 is a rule-of-thumb number that happens to align the levels in the left panel. The average labor share of the nonfarm business sector over the sample period is 0.6.

Figure 6: Fitted Densities and Percentiles of Earnings/GDP Distribution



Notes: Top row: fitted cross-sectional densities and histograms. Bottom row: sample percentiles (10%, 20%, 50%, 80%, and 90%) are blue, percentiles from estimated densities are red.

the unemployed individuals, and a continuous part, representing individuals with non-zero earnings. We normalize the estimated density $p^{(K)}(x|\hat{\alpha}_t)$ so that it integrates to the fraction of employed individuals in the sample and apply the change-of-variable formula to convert the density for transformed earnings $x = g(z|\theta)$ into actual earnings z; see (39) and the Appendix for further details.

Overall, the estimated cross-sectional densities are able to reproduce the time path of the empirical percentiles well, except for some small inaccuracies in the 90th percentile toward the end of the sample. The estimated percentiles for the two specifications are virtually indistinguishable, despite the difference in the density plots in the top panels.

Median earnings (relative to the approximate labor share of per-capita GDP) fall steadily from above one in 1990 to below one in 2017. A similar pattern is observed for the 80% percentile. During the Great Recession, there is a noticeable drop of earnings at the 10% and 20% percentile, followed by a slow and steady rise from 2010 to 2017. Earnings at the 80% and 90% percentiles rise during the Great Recession relative to their 2005 levels.

Functional State-Space Model Estimation. The log-spline density estimation yields a sequence of coefficient vectors $\{\hat{\alpha}_t^{(K)}\}$. We then apply the seasonality adjustment and the compression described in Section 3.2 to obtain $\{\hat{a}_t^{(K)}\}$. The vector Y_t of aggregate variables is composed of TFP growth, real per-capita GDP growth, and the (un)employment rate from the CPS data. We set Y_* equal to the mean of these series. After computing growth rates for TFP and GDP our sample ranges from 1989:Q2 to 2017:Q3. We consider 1990:Q1 as period $t = 1$ and use 1989:Q4 to initialize the lag in the first-order vector autoregressive law of motion. The estimation of the functional state-space model is based on the prior distribution in (30) and (31). We set $\underline{\nu} = n_w + 5$, where n_w is the dimension of the vector W_t in the state-transition equation (28), and let $\underline{\Sigma} = \underline{\nu}\hat{\Sigma}$, where $\hat{\Sigma}$ is the OLS estimator of Σ in (28) that is obtained when the latent a_t in the definition of W_t is replaced by the estimate $\hat{a}_t$. The prior for ϕ is centered at $\underline{\mu}_\phi = 0$.

Hyperparameter Selection and Granger Causality. We proceed by evaluating the (approximate) log MDD as a function of the model dimension $K \in \{4, 6, 8, 10, 14, 22\}$ and the hyperparameter vector λ. For each element of the λ vector we consider ten equally-spaced values of $\ln \lambda_j$ on the interval $[-5, 6]$. Recall that the hyperparameter λ_1 controls the overall precision of the prior distribution. The remaining two hyperparameters control the relative precision of the prior for the submatrices Φ_{ya} (λ_2) and Φ_{ay} (λ_3).

Results are summarized in Table 4. For each value of K we report the optimal $\hat{\lambda}$ (columns 2-4) and the log MDD differentials for $K(\hat{\lambda})$ relative to $K = 4$ (column 5). With respect to K, the log MDD is maximized for the largest value considered, $K = 22$. That is consistent with the visual impression from Figure 6. The additional knots are used to capture the jagged pattern of the histograms and the improvement in fit still outweighs the dimensionality penalty induced by the MDD.

The overall degree of selected shrinkage captured by $\hat{\lambda}_1(K)$ is weakly increasing in the dimensionality K. Moreover, there is generally stronger shrinkage of the off-diagonal blocks to zero. The optimal values $\hat{\lambda}_2(K)$ and $\hat{\lambda}_3(K)$ are either 95 or 403, indicating the absence of

Table 4: Hyperparameter Estimates and Log MDD Differentials

	Optimal λ				Alternative λ			
K	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	MDD	λ_1	λ_2	λ_3	MDD
4	1.2	95	403	0	1.2	1	1	-38
6	1.2	95	403	8,697	1.2	1	1	8,651
8	1.2	403	403	9,594	1.2	1	1	9,527
10	*5.3*	*403*	*95*	*9,707*	5.3	1	1	9,634
14	5.3	95	403	10,177	5.3	1	1	10,089
22	**5.3**	**403**	**95**	**13,667**	5.3	1	1	13,559

Notes: The log MDD differentials are computed with respect to $(K = 4, \lambda = \hat{\lambda})$. For each K we maximized the MDD with respect to λ to obtain $\hat{\lambda}(K)$. We also report log MDD differentials for a set of alternative λ values.

strong Granger-causal relationships between the aggregate variables and the cross-sectional income distribution.

The last column reports log MDD differentials obtained by setting $\lambda_1 = \hat{\lambda}_1$ and $\lambda_2 = \lambda_3 = 1$, meaning that the degree of shrinkage is identical for the coefficients in the diagonal blocks (Φ_{yy}, Φ_{aa}) and the off-diagonal blocks (Φ_{ya}, Φ_{ay}). The differentials are also computed relative to $(K = 4, \lambda = \hat{\lambda})$. This configuration allows for greater spillovers between aggregate and cross-sectional dynamics. Holding K fixed, relaxing the shrinkage for the off-diagonal blocks leads to a deterioration of the log MDD between 38 ($K = 4$) and 108 ($K = 22$). Thus, the larger the dimensionality of the approximation of the cross-sectional density, the stronger the empirical case for shrinking the spillover coefficients to zero to balance the trade-off between in-sample fit and model complexity.

The mechanical application of the MDD criterion suggests to proceed with $K = 22$. However, we are concerned that the jagged pattern of the histograms that the $K = 22$ (and higher) specification is approximating – see Figure 6 – is more an artifact of the data collection (e.g., survey respondents rounding their earnings) than a genuine feature of the earnings distribution. Moreover, because the earnings are standardized by the continuously evolving GDP per capita, in terms of x-coordinates, the spikes shift from period to period, inducing spurious dynamics. Thus, in the remainder of this section we first present results for $K = 10$, which delivers smooth estimates of the cross-sectional densities. We then proceed in Section 5.3 with comparing $K = 10$ to $K = 22$ results. While the impulse responses of the cross-sectional densities computed based on $K = 22$ inherit the saw-tooth pattern visible

in the density estimates, the responses for percentiles and inequality statistics derived from the density responses are indeed very similar for $K = 10$ and $K = 22$.

5.2 Effects of Aggregate Shocks

Identification. In the vector Y_t we order TFP growth first, GDP growth second, and the employment rate third. Let Σ_{tr} be the lower-triangular Cholesky factor of Σ such that $\Sigma = \Sigma_{tr}\Sigma_{tr}'$ and let Ω be an orthonormal matrix. The relationship between the reduced-form innovations u_t and the structural innovations ϵ_t is given by:

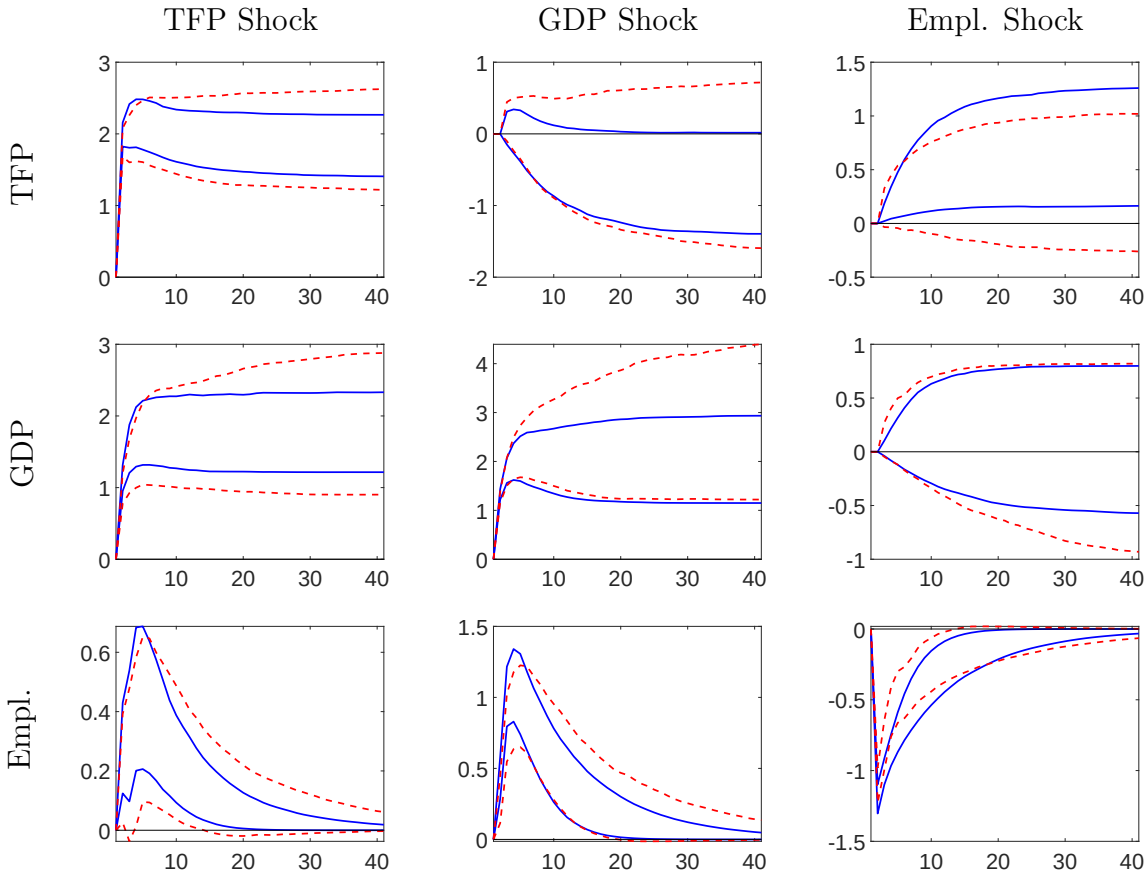
$$u_t = \Sigma_{tr}\Omega\epsilon_t. \tag{40}$$

It is well known that Ω is not identifiable from the data. Denote the jth column of Ω by $\Omega_{\cdot j}$. We label the first structural shock as technology shock and assume that it is the only shock that affects measured TFP contemporaneously. Thus, $\Omega_{\cdot 1} = \iota_1$, where ι_j is an $n_y \times 1$ vector whose jth element is one and all other elements are zero. Moreover, we let $\Omega_{\cdot j} = \iota_j$ for $j = 2, 3$. We refer to shocks $\epsilon_{1,t}$, $\epsilon_{2,t}$, and $\epsilon_{3,t}$ as aggregate shocks because they do not affect the cross-sectional distribution contemporaneously. The shocks $\epsilon_{2,t}$ and $\epsilon_{3,t}$ do not have a strict economic interpretation. We refer to them as shocks to GDP growth and the employment rate. Much of the subsequent discussion will focus on the propagation of technology shocks.

Response of Aggregate Variables. Impulse response bands (delimited by the 10th and 90th percentile of the posterior distribution) computed conditional on the posterior mean estimates of the state-space representation for $K = 10$ are plotted in Figure 7. Because the distributional responses are generally small, we scale the IRFs by a factor of three and consider three-standard-deviation shocks throughout this section. We compare responses based on $\lambda_2 = \hat{\lambda}_2$, $\lambda_3 = \hat{\lambda}_3$ against responses obtained by estimating the state-space model with $\lambda_2 = \lambda_3 = 1$. Recall that under the estimated hyperparameters, we are essentially shrinking the off-diagonal blocks Φ_{ay} and Φ_{ya} to zero. By setting λ_2 and λ_3 equal to one, we are allowing for more flexibility.

First, the TFP growth shock raises the level of TFP permanently. GDP also rises permanently and employment shows a positive response (real business cycle instead of New Keynesian dynamics). Allowing for more flexibility by setting $\lambda_2 = \lambda_3 = 1$ does not change the IRF bands, except that they become slightly wider. Second, the GDP growth shock raises GDP permanently, creates a temporary employment boom and a drop in measured

Figure 7: Responses of Aggregate Variables to Aggregate Shocks



Notes: IRFs for three-standard deviation aggregate shocks (orthogonalized via Cholesky factorization; see (40)). Panels depict responses of the log level of TFP and GDP, scaled by 100, and responses of the employment rate in percent. The bands correspond to pointwise 10th and 90th percentiles of the posterior distribution for $K = 10$. Solid blue responses are based on $\hat{\lambda}_2 = 403$ and $\hat{\lambda}_3 = 95$; dashed red responses are based on $\lambda_2 = \lambda_3 = 1$.

total factor productivity in the long run. While the 10th percentiles of the responses look very similar across hyperparameter settings, the 90th percentiles are shifted upward for $\lambda_2 = \lambda_3 = 1$, in particular for the medium-run and long-run response of GDP to its own shock. Finally, the third shock leads to a drop in the employment rate and it raises TFP and GDP with a one-period delay. While the effect on employment is very similar across the two hyperparameter configurations, the TFP and GDP bands widen and shift downward as the shrinkage of the off-diagonal blocks is reduced.

To summarize, the most noticeable effect of the change in hyperparameters is the widening of the IRF bands, i.e., the decrease of the prior precision slightly reduces the posterior precision. In the Online Appendix we compare IRFs of aggregate variables to aggregate

shocks for $K = 10$ and $K = 22$. Recall that according to Table 4 $K = 22$ attains the largest MDD among the specifications considered. The IRF bands for these two specifications are almost identical, which is consistent with the system being approximately block diagonal.

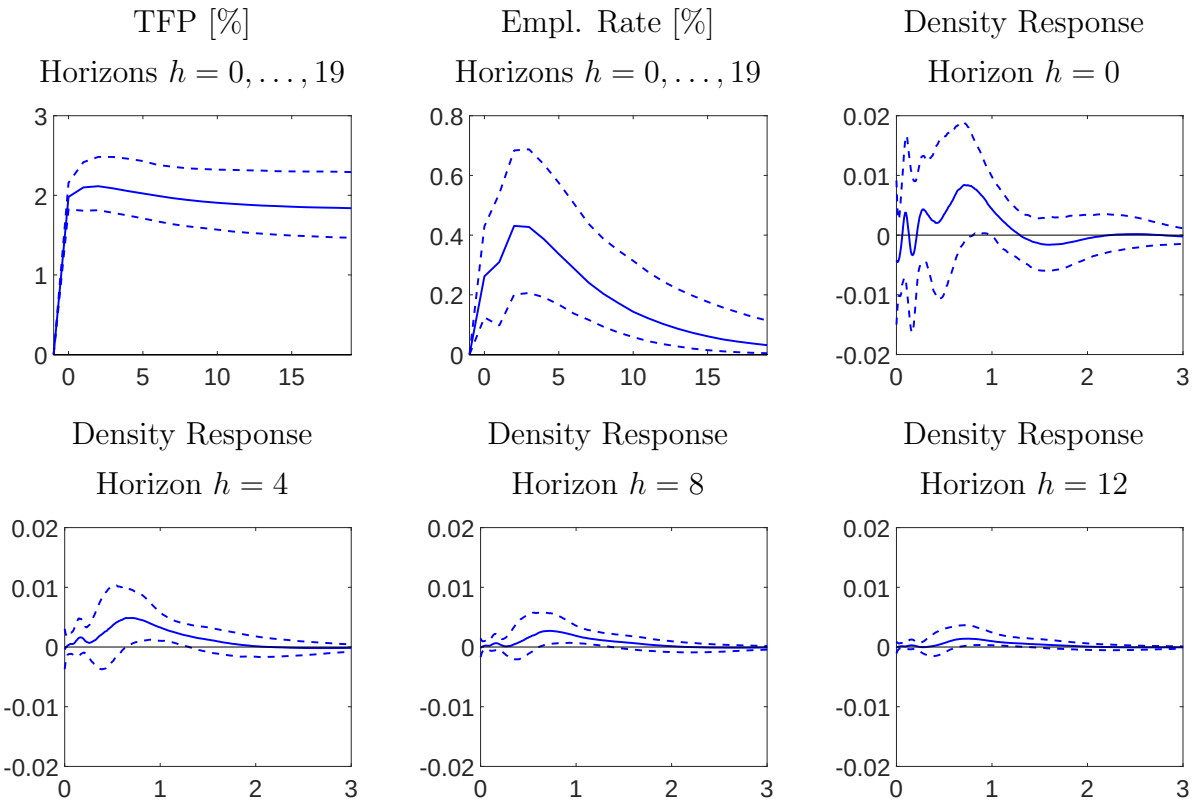
We have also compared the $\hat{\lambda}$ IRFs against IRFs in which we force of off-diagonal blocks of the Φ matrix to be exactly equal to zero (not shown in the figure). We found the two sets of IRFs to be effectively non-distinguishable. Thus, we conclude that in order to measure the response of GDP and employment to an aggregate TFP growth shock, we do not need to keep track of the earnings distribution. Krusell and Smith (1998) showed that in a benchmark stochastic growth model with household heterogeneity and uninsurable idiosyncratic risk, the aggregate dynamics of output, consumption, and investment in response to a TFP shock are almost identical to their RA counterpart. This HA model feature is consistent with our empirical result.

Distributional Responses. We will now examine the response of the cross-sectional distribution of earnings (original data) to a TFP shock. Even though the off-diagonal blocks Φ_{ay} and Φ_{ya} of the VAR coefficient matrix are close to zero, the TFP shock can affect the earnings distribution through the impact vector $\Sigma_{tr}\iota_1$. Figure 8 shows posterior median responses to a three-standard deviation TFP innovation and bands delimited by the 10th and 90th percentiles of the posterior distribution. The top and center left panels reproduce the responses of TFP and employment to a technology shock previously shown in Figure 7. The employment response is hump shaped. At its peak the employment rate increases between 0.3% and 0.7%.

The remaining panels show the response of the earnings distribution at various horizons. As in Section 4, the density responses are constructed as follows. First, we compute the impulse response of a_t from the vector autoregressive state transition equation. Second, we convert that a_t sequence into a sequence of densities $p^{(K)}(x|\alpha_t)$. We normalize these densities so that they integrate to the employment rate. Third, we compute the differential between the shocked density and the steady state density.

Because the employment rate rises in response to a technology shock, the area under the density differential function is positive. According to the median response the mass of individuals earning less than the labor share of GDP per capita increases substantially and initially there is a slight drop in the mass of individuals earning between 1.3 and 2. On impact the 80% bands are wide and include both positive and negative responses. For

Figure 8: Earnings Density (Transformed Data) Response to a TFP Shock

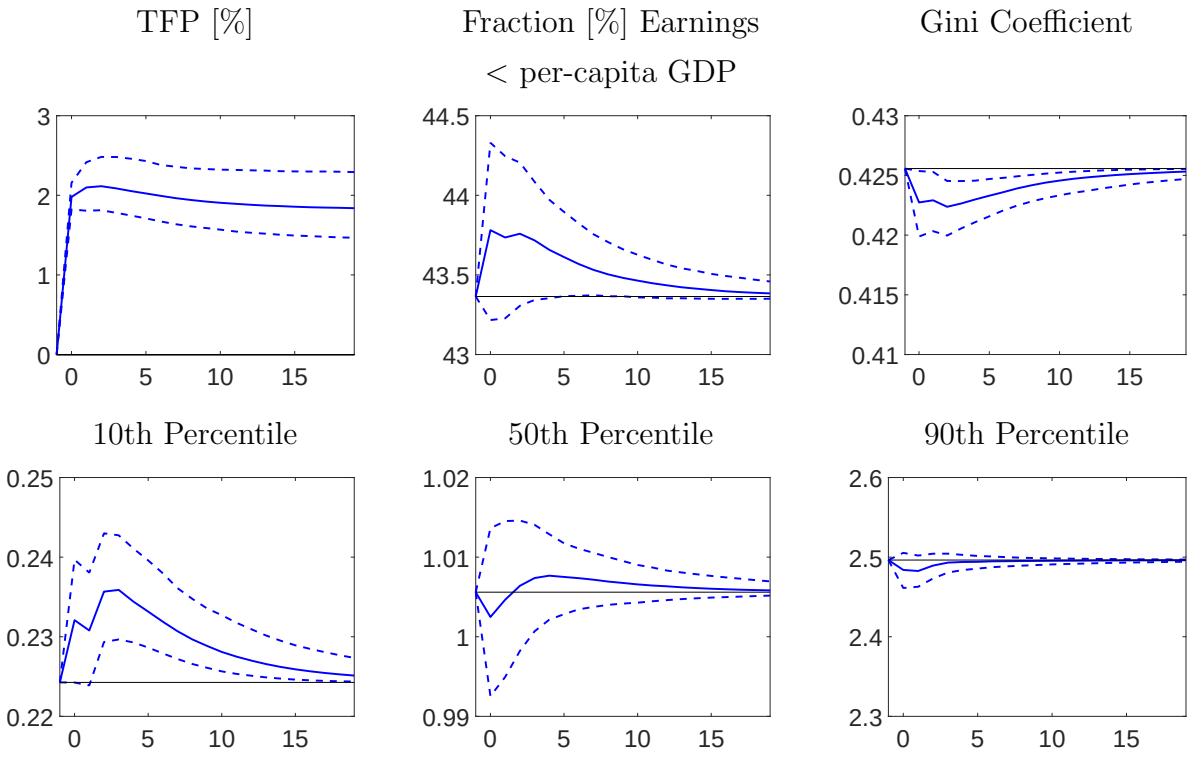


Notes: Responses to a 3-standard-deviations shock to TFP for $K = 10$. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plot depicts 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. As distributional responses we depict differences between the shocked and the steady state cross-sectional density at various horizons.

horizons $h = 4$, 8, and 12 and earnings below GDP per capita, the density differential bands include mostly positive values.

Most of the probability mass is added between 0.5 and 1.0. This mass comes from two directions: first, unemployed individuals who find jobs and receive strictly positive instead of zero labor earnings, and, second, individuals whose earnings do not rise as strongly as GDP per capita. The first effect is consistent with a model in which individuals are heterogeneous with respect to their idiosyncratic productivity and only individuals whose productivity exceeds a state-dependent threshold work; see, for instance, Chang and Kim (2006) and Chang, Kim, Kwon, and Rogerson (2019). In response to an expansionary TFP shock previously unemployed low productivity individuals are hired. If wages per efficiency unit are constant, these individuals are likely to earn less than GDP per capita. The second effect requires adjustments on the intensive margin or some heterogeneity in wages per efficiency

Figure 9: Inequality Measure (Original Data) Responses to a TFP Shock



Notes: Responses to a 3-standard-deviations shock to TFP for $K = 10$. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plot depicts 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution.

unit. A strong wealth effect that leads wealthy individuals to reduce their hours or a relative fall of efficiency unit-specific wages for some individuals is required to shift their earnings from above 1.0 to below 1.0.

A key advantage of the functional modeling approach is that in any period t the cross-sectional earnings density fully summarizes the earnings distribution. Based on the impulse response of the cross-sectional density, we can now compute impulse response functions of various summary statistics. Figure 9 shows responses for the fraction of individuals with an earnings-to-per-capita-GDP ratio less than one, the Gini coefficient, and the 10th, 50th, and 90th percentiles of the distribution. The impulse responses are computed relative to an average level of these statistics, indicated by a solid black line. All statistics are computed after assigning zero earnings to the unemployed individuals.

According to the posterior median of the IRF, overall the fraction of individuals earning less than per-capita GDP increases from 43.5% to 43.8%. While the density differential

response for $h = 0$ depicted in Figure 8 looked "insignificant," converted into the fraction earning less than per-capita GDP, the response is positive with a probability of almost 90% even on impact. As individuals move from being unemployed (zero earnings) to being employed (positive earnings) the Gini coefficient falls from 0.4256 to 0.4224. Earnings at the 10th percentile of the cross-sectional distribution also increase from 0.224 to 0.236, at the median they stay roughly constant around 1.005, with a band that ranges upon impact from approximately 0.995 to 1.015. Earnings at the 90th percentile fall very slightly, from 2.496 to 2.483. The posterior median of the ratio of earnings at the 90th and 10th percentile drops by 5.1% from 11.13 to 10.56 two periods after the impact of the shock. This compression of the earnings distribution is consistent with the decrease in the Gini coefficient.

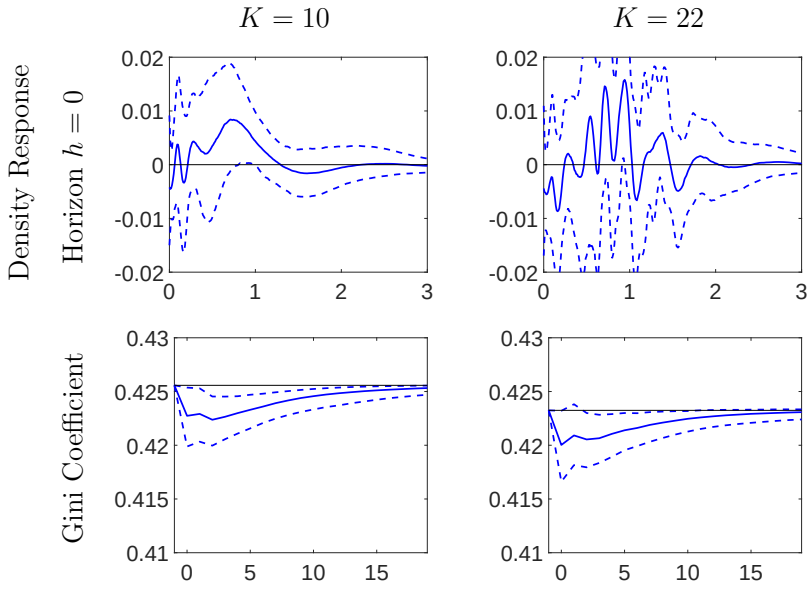
5.3 Robustness Analysis 1: Results for $K = 22$

According to Table 4 the MDD is maximized with respect to K for the largest value considered, namely $K = 22$. We previously presented results for $K = 10$, because they were based on a smoother density. As a robustness exercise, we compare selected IRFs for $K = 10$ and $K = 22$ in Figure 10. The density responses look quite different for the two model specifications. For $K = 22$ the posterior median appears to be much more volatile as a function of the earnings. The swings trace out movements of the jagged density estimates shown in Figure 6. The Gini coefficient responses on the other hand, are almost identical for the two values of K, with the exception of a small level shift. In the Online Appendix we also plot the responses of the percentiles and the fraction of individuals with earnings less than GDP per capita. It turns out that the $K = 22$ IRFs are quantitatively very similar to the responses depicted in Figure 9, confirming that our results are robust with respect to the dimensionality K of the density approximation.

5.4 Robustness Analysis 2: Comparison to Simple VARs

The advantage of the functional approach is that once the dynamics of the cross-sectional densities have been estimated, it is straightforward to compute the dynamics of any statistic associated with the densities. We previously considered impulse responses of the fraction of individuals earning less than GDP per capita, the Gini coefficients, and the quantiles of the earnings distribution, all computed from the original data. As a robustness check, we estimate two simple VARs. The first VAR combines the aggregate variables with the

Figure 10: Earnings Density and Gini Responses to a TFP Shock: $K = 10$ vs. $K = 22$



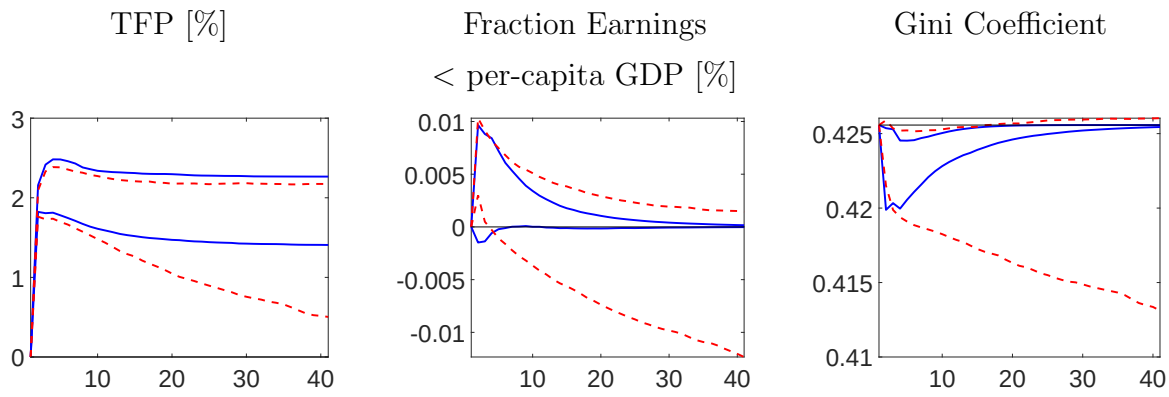
Notes: Responses to a 3-standard-deviations shock to TFP. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plot depicts 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. As distributional responses we depict differences between the shocked and the steady state cross-sectional density at various horizons. Density responses are reported for transformed data, Gini responses are reported for original data.

fraction of individuals earning less than GDP per capita and the Gini coefficient series directly computed from the cross-sectional observations. The second VAR combines the aggregate variables with the 10th, 20th, 50th, 80th and 90th percentiles.

Both VARs take the form of (29) with the prior in (30), where W_t is redefined to include the relevant distributional statistics. The estimated hyperparameters for the Gini coefficient VAR are $\hat{\lambda}_1 = 0.29$, $\hat{\lambda}_2 = 403$, and $\hat{\lambda}_3 = 403$. The estimated λs for the percentile VAR are $\hat{\lambda}_1 = 1.25$, $\hat{\lambda}_2 = 403$, and $\hat{\lambda}_3 = 5.29$. For the Gini specification the MDD criterion shrinks toward a block diagonal Φ matrix. For the percentile VAR, on the other hand, the MDD criterion leads to less shrinkage on the parameter block that controls the effect of lagged aggregate variables on the percentiles of the current cross-sectional distribution. In fact the $\hat{\lambda}$ pattern is similar to the one obtained from the functional VAR applied to data simulated from the KS economy.

Figure 11 overlays impulse responses to a three standard deviation technology shock for TFP, the fraction of individuals earning less than per-capita GDP, and the Gini coefficient from the functional state-space model and the Gini coefficient VAR. While the 90th per-

Figure 11: Functional vs. Alternative VAR: Inequality Statistics Responses to a TFP Shock



Notes: Responses to a 3-standard-deviations shock to TFP. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The bands correspond to pointwise 10th and 90th percentiles of the posterior distribution. Solid blue responses are based on the functional state-space model for $K = 10$; dashed red responses are based on the alternative VAR. Inequality responses are computed for the original data.

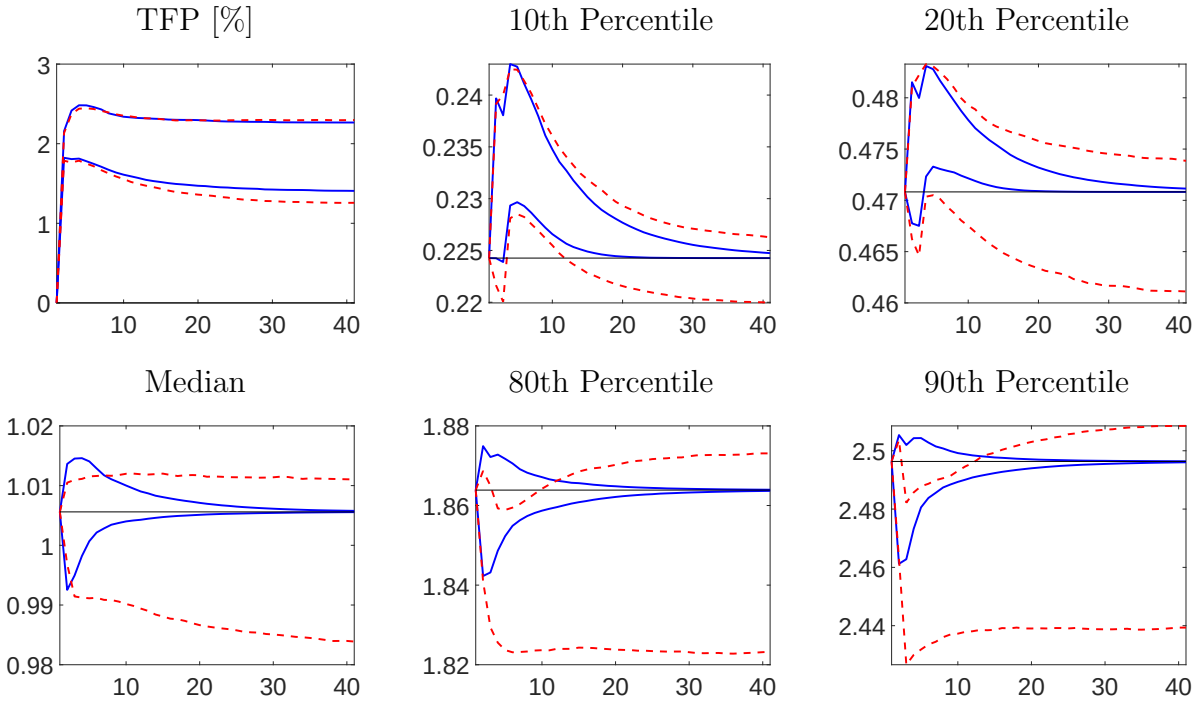
centiles of the posterior distributions of impulse responses line up quite accurately, the 10th percentiles differ substantially. Under the simple VAR, the 10th percentiles for the fraction of earnings and the Gini coefficient continue to fall as the horizon increases, which is not particularly plausible.

Figure 12 overlays impulse responses to a three standard deviation technology shock for TFP and various percentiles of the earnings distribution from the functional state-space model and the percentile VAR. The bands of the TFP responses line up almost perfectly. While the responses for the 10th and 20th are quite similar, the functional model delivers substantially tighter bands than the finite-dimensional VAR that contains the sample quantiles, in particular after 20 to 30 quarters when the widths of the functional VAR bands is close to zero. The bottom edges of the bands for the responses of the 50th, 80th, and 90th percentiles are much lower for the simplified VAR than the functional VAR. Overall, the width of the IRF bands suggests that the functional approach allows for a much sharper inference, despite a higher but still parsimonious parameter count (ten series instead of two or five series to represent the dynamics of the cross section), and is therefore preferable.

5.5 Effects of a Distributional Shock

The shocks considered in the previous subsection only affected the aggregate variables upon impact. We will now consider the effects of a distributional shock which we define to be a

Figure 12: Functional vs. Alternative VAR: Quantile Responses to a TFP Shock



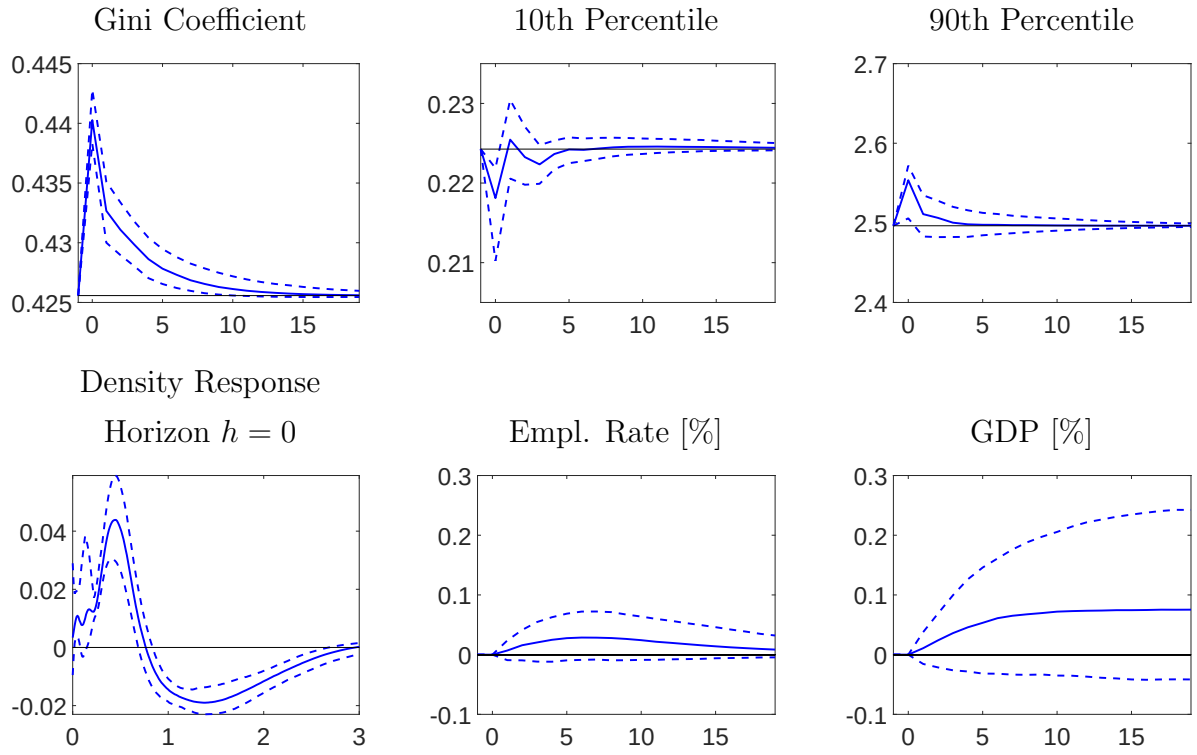
Notes: Responses to a 3-standard-deviations shock to TFP. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The bands correspond to pointwise 10th and 90th percentiles of the posterior distribution. Solid blue responses are based on the functional state-space model for $K = 10$; dashed red responses are based on the alternative VAR. Percentile responses are computed for the original data.

shock that does not move the aggregate variables upon impact. A distributional shock could reflect, for instance, an unanticipated revenue neutral change in fiscal policy that triggers a redistribution of earnings, or a change in the underlying idiosyncratic earnings processes such as an increase in earnings risk.

The aggregate effect depends on the nature of the distributional shock. To the extent that the marginal propensity to consume is negatively correlated with income, higher inequality may lower aggregate consumption. If inequality comes from a rise in idiosyncratic volatility, then it could raise precautionary savings. Auclert and Rognlie (2020) show that in general equilibrium a falling interest rate may weaken the negative relationship between inequality and aggregate output, and that the net effect is sensitive to monetary and fiscal policy.

In our fVAR setting the distributional shock is an abstract object that is not directly tied to idiosyncratic earnings processes. Building on the notation in (40), let $Q_{\cdot 4} = q$, where $q' = [0, q'_\alpha]$ and the partition of q conforms with the partition of $u'_t = [u'_{y,t}, u'_{\alpha,t}]$. By construction, $Q_{\cdot 4}$ is orthogonal to $Q_{\cdot j}$, $j = 1, \ldots, 3$. We choose the unit-length vector q_α as

Figure 13: Responses to a Distributional Innovation ϵ_t^* (Maximize Gini-Coefficient)



Notes: Responses to a 3-standard-deviation distributional innovation ϵ_t^* for $K = 10$. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plot depicts 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. As distributional responses we depict differences between the shocked and the steady state cross-sectional density at various horizons. Density response is reported for transformed data, all other responses for original data.

a function of (Φ, Σ) to generate a maximal increase in inequality upon impact, as measured by the Gini coefficient.

The results are depicted in Figure 13. The average response of the Gini coefficient spikes upon impact by construction of the distributional shock and then dies out fairly quickly according to the autoregressive dynamics. Earnings at the 10th percentile fall while earnings at the 90th percentile rise relative to the labor share of GDP per capita. The posterior median ratio of earnings at the 90th and 10th percentiles rises by 5.3% from 11.13 to 11.72, confirming that the shock increases inequality. The bottom left panel shows the density response upon impact of the shock. Probability mass shifts from the 1.0 to 2.5 range into the 0 to 1.0 range, leaving essentially only the labor income of individuals who are earning more than 2.5 times of GDP per capita unaffected.

The bottom center and right panels show the response of two aggregate variables: the

employment rate and GDP. By construction, neither employment nor GDP respond to the distributional shock upon impact. At the posterior median both employment and GDP increase, but not by much. The employment response is hump shaped and peaks at 0.03%, whereas the effect on GDP is an increase of 0.08% in the long run. While the TFP shock generates a negative relationship between economic activity and inequality, the particular distributional shock constructed here generates a weak positive relationship. The bands for the GDP and employment rate responses do include zero, which is consistent with the response that Auclert and Rognlie (2020) find for standard neoclassical models.

6 Conclusion

We developed a functional state-space model that stacks macroeconomic aggregates and cross-sectional distributions to provide semi-structural evidence about the interaction of aggregate and distributional dynamics. We documented that the model estimated on simulated data is able to reproduce the impulse response dynamics of the underlying KS model economy. In our empirical analysis we found that adding the earnings distribution to a VAR in TFP growth, GDP growth, and employment does not affect the propagation of aggregate variables to aggregate shocks, which is consistent with model-based findings reported in Krusell and Smith (1998). We find that an expansionary TFP shock decreases earnings inequality in our sample because it raises earnings at the bottom of the earnings distribution. Finally, we show that a distributional shock that raises inequality has a small positive, albeit not significant, effect on aggregate output. We expect the techniques developed in this paper to be widely applicable to study the interaction between macroeconomic aggregates and cross-sectional distributions and useful for the evaluation of the most recent vintage of HA models. Useful extensions left for future work include the introduction of time-varying volatility and allowing for mixed-frequency observations.

References

Acharya, S., W. Chen, M. Del Negro, K. Dogra, E. Matlin, and R. Sarfati (2019): "Estimating HANK: Macro Time Series and Micro Moments," *Manuscript, Federal Reserve Bank of New York*.

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2018): "When Inequality Matters for Macro and Macro Matters for Inequality," in *NBER Macroeconomics Annual 2017*, ed. by M. Eichenbaum, and J. Parker, pp. 1 – 75. University of Chicago Press.

Auclert, A., and M. Rognlie (2020): "Inequality and Aggregate Demand," *Manuscript, Stanford University*.

Bayer, C., B. Born, and R. Luetticke (2020): "Shocks, Frictions, and Inequality in US Business Cycles," *Working paper*.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021): "Inequality, Business Cycles, and Monetary-Fiscal Policy," *Econometrica*, forthcoming.

Bosq, D. (2000): *Linear Processes in Function Spaces*. Springer Verlag, New York.

Carter, C., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Chang, Y., C. S. Kim, and J. Park (2016): "Nonstationarity in Time Series of State Densities," *Journal of Econometrics*, 192, 152–167.

Chang, Y., and S.-B. Kim (2006): "From Individual to Aggregate Labor Supply: A Quantitative Analysis Based on a Heterogeneous Agent Macroeconomy," *International Economic Review*, 47, 1–27.

Chang, Y., S.-B. Kim, K. Kwon, and R. Rogerson (2019): "2018 Klein Lecture: Individual and Aggregate Labor Supply in Heterogeneous Agent Economies with Intensive and Extensive Margins," *International Economic Review*, 60(1), 3–24.

Childers, D. (2018): "Solution of Rational Expectations Models with Function Valued States," *Manuscript, Carnegie Mellon University*.

Cho, D. (2020): "Unemployment Risk, MPC Heterogeneity, and Business Cycles," *Manuscript, University of Melbourne*.

Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia (2017): "Innocent Bystanders? Monetary Policy and Inequality," *Journal of Monetary Economics*, 88, 70–89.

Del Negro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45(2), 643 – 673.

Diebold, F. X., and C. Li (2006): "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130, 337–364.

Durbin, J., and S. J. Koopman (2001): *Time Series Analysis by State Space Methods*. Oxford University Press.

Fernald, J. G. (2012): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," *FRBSF Working Paper*, 2012-19.

Fuentes-Albero, C., and L. Melosi (2013): "Methods for Computing Marginal Data Densities from the Gibbs Output," *Journal of Econometrics*, 175, 132–141.

Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–126.

Giannone, D., M. Lenza, and G. Primiceri (2015): "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics*, 97(2), 436–451.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge University Press.

Horvath, L., and P. Kokoszka (2012): *Inference for Functional Data with Applications*. Springer Verlag, New York.

Hu, B., and J. Y. Park (2017): "Econometric Analysis of Functional Dynamics in the Presence of Persistence," *Manuscript, Department of Economics, Indiana University*.

Huggett, M. (1997): "The one-sector growth model with idiosyncratic shocks: steady states and dynamics," *Journal of Monetary Economics*, 39, 385–403.

Inoue, A., and B. Rossi (2020): "The Effects of Conventional and Unconventional Monetary Policy: A New Approach," *Manuscript, Vanderbilt University and Pompeu Fabra*.

Kaplan, G., and L. Violante, Giovanni (2018): "Microeconomic Heterogeneity and Macroeconomic Shocks," *Journal of Economic Perspective*, 32, 167–194.

Kooperberg, C., and C. J. Stone (1990): "A Study of Logspline Density Estimation," *Computational Statistics & Data Analysis*, 12, 327–347.

Krueger, D., K. Mitman, and F. Perri (2016): "Macroeconomics and Household Heterogeneity," in *Handbook of Macroeconomics*, ed. by H. Uhlig, and J. Taylor, vol. 2A, pp. 843–921. Elsevier (Amsterdam).

Krusell, P., and A. A. Smith (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106(5), 867–896.

Liu, L., and M. Plagborg-Møller (2019): "Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data," *Manuscript, Indiana University and Princeton University*.

Meeks, R., and F. Monti (2019): "Heterogeneous Beliefs and the Phillips Curve," *Bank of England Staff Working Paper*, 807.

Mongey, S., and J. Williams (2017): "Firm Dispersion and Business Cycles: Estimating Aggregate Shocks Using Panel Data," *Manuscript, University of Chicago*.

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

Ramsey, J. O., and B. W. Silverman (2005): *Functional Data Analysis*. Springer Verlag, New York, 2nd edn.

Reiter, M. (2009): "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics & Control*, 33, 649–665.

Schorfheide, F., and D. Song (2015): "Real-Time Forecasting with a Mixed-Frequency VAR," *Journal of Business and Economic Statistics*, 33(3), 366–380.

Varin, C., N. Reid, and D. Firth (2011): "An Overview of Composite Likelihood Methods," *Statistica Sinica*, 21, 5–42.

Villalvazo, S. (2021): "Inequality and Asset Prices during Sudden Stops," *Manuscript, University of Pennsylvania*.

Winberry, T. (2018): "A Toolbox for Solving and Estimating Heterogeneous Agent Macro Models," *Quantitative Economics*, 9, 1123–1151.

Online Appendix: Heterogeneity and Aggregate Fluctuations

Minsu Chang, Xiaohong Chen, and Frank Schorfheide

A Functional State-Space Model – Further Details

A.1 Hessian Matrix

An important object in the analysis is the Hessian matrix of the log-likelihood function for the cross-sectional observations. Let $\zeta_k(x)$, $k = 1, \ldots, K$ be the sequence of basis functions for the approximation of the cross-sectional density $p_t^{(K)}(x)$ and let $X_t = \{x_{1t}, \ldots, x_{Nt}\}$. To simplify the notation, we will subsequently drop the (K) superscript. Recall that $\bar{\zeta}_k(X_t) = \frac{1}{N}\sum_{i=1}^N \zeta_k(x_{it})$. The log likelihood function of X_t has the form

$$\mathcal{L}(\alpha_t|X_t) = \sum_{k=1}^K \alpha_{k,t}\bar{\zeta}_k(X_t) - \ln \int \exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\} dx. \tag{A.1}$$

The first-order derivatives with respect to $\alpha_{k,t}$ for $k = 1, \ldots, K$ are given by

$$\begin{aligned}\mathcal{L}_k^{(1)}(\alpha_t|X_t) &= \bar{\zeta}_k(X_t) - \frac{\int \zeta_k(x)\exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\}dx}{\int \exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\}dx} \\ &= \bar{\zeta}_k(X_t) - \int \zeta_k(x)p(x|\alpha_t)dx.\end{aligned} \tag{A.2}$$

The second-order derivatives with respect to $(\alpha_{k,t}, \alpha_{l,t})$ are given by

$$\begin{aligned}\mathcal{L}_{kl}^{(2)}(\alpha_t|X_t) &= -\int \zeta_k(x)\frac{\partial p(x|\alpha_t)}{\partial \alpha_{l,t}}dx \\ &= -\mathbb{E}\left[\zeta_k(x_{it})\frac{\partial \ln p(x_{it}|\alpha_t)}{\partial \alpha_{l,t}}\right] \\ &= -\mathbb{E}\left[\zeta_k(x_{it})\left(\zeta_l(x_{it}) - \int \zeta_l(x)p(x|\alpha_t)dx\right)\right].\end{aligned}$$

Thus, we can write

$$\begin{aligned}&\mathcal{L}_{kl}^{(2)}(\alpha_t|X_t) \\ &= -\int \left(\zeta_k(x) - \int \zeta_k(x)p(x|\alpha_t)dx\right)\left(\zeta_l(x) - \int \zeta_l(x)p(x|\alpha_t)dx\right)p(x|\alpha_t)dx.\end{aligned} \tag{A.3}$$

Note that the Hessian elements only depend on α_t but not on the data X_t.

A.2 Top Coding

Likelihood Function with Censoring. We define the censoring point c_t as

$$c_t = \max_{i=1,\ldots,N} x_{it}$$

Moreover, we let

$$N_{t,max} = \sum_{i=1}^{N} \mathbb{I}\{x_{it} = c_t\}.$$

If $N_{t,max} = 1$, we assume that the observed sample is not constrained by the top-coding and use the standard likelihood function described in the main text. If $N_{t,max} > 1$ we use a likelihood function that assumes that any earnings value exceeding c_t is coded as c_t.

Recall that in the main text we ignored the dependence of the cross-sectional sample size N on t in the notation and defined $p^{(K)}(X_t|\alpha_t) = \exp\left\{N\mathcal{L}^{(K)}(\alpha_t|X_t)\right\}$, where

$$\mathcal{L}^{(K)}(\alpha_t|X_t) = \bar{\zeta}'(X_t)\alpha_t - \ln\int_0^\infty \exp\left\{\zeta'(x)\alpha_t\right\}dx, \quad \bar{\zeta}(X_t) = \frac{1}{N}\sum_{i=1}^{N_t}\zeta(x_{it}).$$

We introduce the unknown parameter $\pi_t = \mathbb{P}\{x_{it} \geq c_t\}$. We drop the top-coded observations from the definition of $\bar{\zeta}(X_t)$ und make the time dependence explicit in the notation. Let

$$\bar{\zeta}_t(X_t) = \frac{1}{N_t}\sum_{i=1}^{N_t}\zeta(x_{it})\mathbb{I}\{x_{it} < c_t\}. \tag{A.4}$$

The log likelihood function is obtained as follows: the sample contains $N_{t,max}$ top-coded observations where the probability of sampling a top-coded observation is π_t. The probability of samping an observation that is not top-coded is $(1-\pi_t)$. Conditional on not being top-coded, the observation $x_{it} < c_t$ is sampled from a continuous density with a domain that is truncated at c_t. Thus, dividing the log-likelihood by the sample size N_t, we obtain

$$\begin{aligned}\mathcal{L}^{(K)}(\alpha_t, \pi_t|X_t) &= \frac{N_{t,max}}{N_t}\ln\pi + \frac{N_t - N_{t,max}}{N_t}\ln(1-\pi_t) \\ &\quad + \bar{\zeta}_t'(X_t)\alpha_t - \frac{N_t - N_{t,max}}{N_t}\ln\int_0^{c_t}\exp\left\{\zeta'(x)\alpha_t\right\}dx.\end{aligned} \tag{A.5}$$

Notice that regardless of the value of α_t, the MLE of π_t is

$$\hat{\pi}_t = \text{argmax}_{\pi\in[0,1]}\,\mathcal{L}^{(K)}(\alpha_t, \pi_t|X_t) = N_{t,max}/N_t. \tag{A.6}$$

Moreover, regardless of the value of π_t, the MLE of α_t is given by

$$\begin{aligned}\hat{\alpha}_t &= \text{argmax}_{\alpha_t}\, \mathcal{L}^{(K)}(\alpha_t, \pi_t | X_t) \\ &= \text{argmax}_{\alpha_t}\, \bar{\zeta}_t'(X_t)\alpha_t - \frac{N_t - N_{t,max}}{N_t} \ln \int_0^{c_t} \exp\left\{\zeta'(x)\alpha_t\right\} dx.\end{aligned} \tag{A.7}$$

The objective function for α_t is almost identical to what we had without top coding, except for a definition of $\bar{\zeta}_t(X_t)$ that drops the top-coded observations in the summation and the factor of $(N_t - N_{t,max})/N_t$ in front of the normalization constant of the density.

Recovering the Density for Uncensored Observations. To reconstruct the full density we can use

$$p(x|\alpha_t) = \frac{\exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\}}{\int_0^\infty \exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\} dx}. \tag{A.8}$$

Note that here we dropped the censoring indicator function and the integration is now from 0 to ∞. Once the α_t's have been estimated based on the censored observations, we work with the full density in the functional state-space model and its K-dimensional approximation.

Modification of Hessian Matrix. We now re-compute the score and the Hessian. Dropping the (K) superscript we obtain the following first derivatives with respect to α_k for $k = 1, \ldots, K$:

$$\mathcal{L}_k^{(1)}(\alpha_t|\pi_t, X_t) = \bar{\zeta}_{t,k}(X_t) - \left(\frac{N_t - N_{t,max}}{N_t}\right) \int_0^{c_t} \zeta_k(x)\bar{p}(x|\alpha_t)dx,$$

where

$$\bar{p}(x|\alpha_t) = \frac{\exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\}}{\int_0^{c_t} \exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\} dx} \mathbb{I}\{x < c_t\}.$$

We can now deduce from our previous calculations that

$$\begin{aligned}&\mathcal{L}_{kl}^{(2)}(\alpha_t|\pi_t, X_t) \\ &= -\left(\frac{N_t - N_{t,max}}{N_t}\right) \int_0^{c_t} \left(\zeta_k(x) - \int_0^{c_t} \zeta_k(x)\bar{p}(x|\alpha_t)dx\right) \\ &\qquad \times \left(\zeta_l(x) - \int_0^{c_t} \zeta_l(x)\bar{p}(x|\alpha_t)dx\right) \bar{p}(x|\alpha_t)dx.\end{aligned} \tag{A.9}$$

Thus, compared to the standard case in (A.3), the limits of integration change and there is an additional factor $(N_t - N_{t,max})/N_t$.

A.3 VAR State-Transition Equation

In this section we describe how to sample from the conditional posterior distributions $\phi|(\Sigma, W, \lambda)$ and $\Sigma|(\phi, W, \lambda)$ that appear in the Gibbs sampler (20) and how to approximate the aggregate component of the marginal data density in (20) given by:

$$\int \left(\prod_{t=1}^{T} p^{(K)}(Y_t, \hat{\alpha}_t | Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \theta) \right) p(\theta|\lambda) d\theta.$$

Throughout this section we adopt the convention that $\phi = \text{vec}(\Phi)$. Recall that the prior distribution for the VAR parameters was given in (30) and has the form

$$\Sigma \sim IW(\underline{\nu}, \underline{S}), \quad \phi|\lambda \sim N\big(\underline{\mu}_\phi, \underline{P}_\phi^{-1}(\lambda)\big).$$

Posteriors. The state-transition equation in matrix form was given by $W = Z\Phi + U$; see (29). Let n_w be the dimension of W_t, then the likelihood function for (Φ, Σ) is given by

$$\begin{aligned} & p(W|\Phi, \Sigma) \qquad (A.10) \\ &= (2\pi)^{-n_w T/2} |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} \text{tr}\big[\Sigma^{-1}(\Phi - \hat{\Phi})' Z' Z (\Phi - \hat{\Phi})\big] \right\} \exp\left\{ -\frac{1}{2} \text{tr}\big[\Sigma^{-1} \hat{S}\big] \right\}, \end{aligned}$$

where

$$\hat{\Phi} = (Z'Z)^{-1} Z'W, \quad \hat{S} = W'W - W'Z(Z'Z)^{-1}Z'W.$$

The prior density takes the form

$$\begin{aligned} p(\phi, \Sigma|\lambda) &= (2\pi)^{-n_w n_z/2} |\underline{P}_\phi|^{1/2} \exp\left\{ -\frac{1}{2}(\phi - \underline{\mu}_\phi)' \underline{P}_\phi (\phi - \underline{\mu}_\phi) \right\} \qquad (A.11) \\ &\quad \times \underline{C}_{IW} |\Sigma|^{-(\underline{\nu} + n_w + 1)/2} \exp\left\{ -\frac{1}{2} \text{tr}\big[\Sigma^{-1} \underline{S}\big] \right\}, \end{aligned}$$

where $\underline{C}_{IW}$ is the normalization constant of the IW prior.

The conditional posteriors $\phi|(W, \Sigma, \lambda)$ and $\Sigma|(W, \phi, \lambda)$ can be obtained as follows:

$$\begin{aligned} p(\phi|W, \Sigma, \lambda) &\propto p(W|\Phi, \Sigma) p(\phi|\lambda) \qquad (A.12) \\ &\quad \times \exp\left\{ -\frac{1}{2}(\phi - \hat{\phi})' (\Sigma^{-1} \otimes Z'Z)(\phi - \hat{\phi}) \right\} \\ &\quad \times \exp\left\{ -\frac{1}{2}(\phi - \underline{\mu}_\phi)' \underline{P}_\phi (\phi - \underline{\mu}_\phi) \right\}. \end{aligned}$$

Define

$$\bar{P}_\phi = \underline{P}_\phi + \Sigma^{-1} \otimes Z'Z, \quad \bar{\mu}_\phi = \bar{P}_\phi^{-1}\big(\underline{P}_\phi \underline{\mu}_\phi + (\Sigma^{-1} \otimes Z'Z)\hat{\phi}\big).$$

We deduce that the conditional posterior distribution of ϕ takes the form

$$\phi|(W,\Sigma) \sim N\big(\bar{\mu}_\phi, \bar{P}_\phi\big). \tag{A.13}$$

The conditional posterior of Σ can be obtained as follows:

$$\begin{aligned} p(\Sigma|W,\phi,\lambda) &\propto p(W|\Phi,\Sigma)p(\Sigma|\lambda) \\ &\propto |\Sigma|^{-T/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\big[(W-Z\Phi)'(W-Z\Phi)\big]\right\} \\ &\quad \times|\Sigma|^{-(\underline{\nu}+n_w+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\big[\Sigma^{-1}\underline{S}\big]\right\}. \end{aligned} \tag{A.14}$$

Define

$$\bar{\nu} = \underline{\nu} + T, \quad \bar{S} = \underline{S} + (W - Z\Phi)'(W - Z\Phi)$$

and deduce that

$$\Sigma|(W,\phi) \sim IW\big(\bar{\nu}, \bar{S}\big). \tag{A.15}$$

Marginal Data Density. Treating W as observed, we can approximate the marginal data density $p(W|\lambda)$ for the VAR using Geweke (1999)'s modified harmonic mean estimator. We begin by integrating out ϕ directly:

$$\begin{aligned} p(W|\Sigma,\lambda) &= \int p(W|\phi,\Sigma)p(\phi|\lambda)d\phi \\ &= (2\pi)^{-n_wT/2}|\Sigma|^{-T/2}\int \exp\left\{-\frac{1}{2}(\phi-\hat{\phi})'(\Sigma^{-1}\otimes Z'Z)(\phi-\hat{\phi})\right\} \\ &\quad \times\exp\left\{-\frac{1}{2}\mathrm{tr}\big[\Sigma^{-1}\hat{S}\big]\right\}(2\pi)^{-n_wn_z/2}|\underline{P}_\phi|^{1/2}\exp\left\{-\frac{1}{2}(\phi-\underline{\mu}_\phi)'\underline{P}_\phi(\phi-\underline{\mu}_\phi)\right\}d\phi. \\ &= (2\pi)^{-n_wT/2}|\Sigma|^{-T/2}|\underline{P}_\phi|^{1/2}|\bar{P}_\phi|^{-1/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\big[\Sigma^{-1}\hat{S}\big]\right\} \\ &\quad \times\exp\left\{-\frac{1}{2}\big[\hat{\phi}'(\Sigma^{-1}\otimes Z'Z)\hat{\phi} + \underline{\mu}_\phi'\underline{P}_\phi\underline{\mu}_\phi - \bar{\mu}_\phi'\bar{P}_\phi\bar{\mu}_\phi\big]\right\}. \end{aligned} \tag{A.16}$$

We now can express

$$p(W|\lambda) = \int p(W|\Sigma,\lambda)p(\Sigma|\lambda)d\Sigma. \tag{A.17}$$

To implement the modified harmonic mean estimator, we need to evaluate the prior density

$$p(\Sigma|\lambda) = \underline{C}_{IW}|\Sigma|^{-(\underline{\nu}+n_w+1)/2}\exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}\underline{S}\right]\right\},$$

where

$$\underline{C}_{IW} = \frac{|\underline{S}|^{\underline{\nu}/2}}{2^{n_w\underline{\nu}/2}\pi^{n_w(n_w-1)/4}\prod_{i=1}^{n_w}\Gamma((\underline{\nu}+1-i)/2)},$$

in addition to $p(W|\Sigma,\lambda)$ in (A.16).

We use a modified harmonic mean estimator of the following form

$$p(W|\lambda) \approx \left[\frac{1}{N_{sim}}\sum_{i=1}^{N_{sim}}\frac{f\big(\text{vech}(\Sigma^i)\big)}{p(W|\Sigma^i,\lambda)p(\Sigma^i|\lambda)}\right]^{-1}, \tag{A.18}$$

where $\text{vech}(A)$ stacks the non-redundant elements of a symmetric matrix A and $\{\Sigma^i\}_{i=1}^{N_{sim}}$ are draws from the posterior distribution $p(\Sigma|W,\lambda)$. We choose the following function $f(\theta)$:

$$\begin{aligned} f(\theta) &= \tau^{-1}(2\pi)^{-n_\theta/2}|\bar{V}_\theta|^{-1/2}\exp[-0.5(\theta-\bar{\theta})'\bar{V}_\theta^{-1}(\theta-\bar{\theta})] \\ &\quad \times I\left\{(\theta-\bar{\theta})'\bar{V}_\theta^{-1}(\theta-\bar{\theta}) \leq F^{-1}_{\chi^2_{n_\theta}}(\tau)\right\}, \end{aligned}$$

where $\theta = \text{vech}(\Sigma)$, $n_\theta = n_w(n_w+1)/2\times 1$, and $\bar{\theta}$ and $\bar{V}_\theta$ are numerical approximations of the posterior mean and covariance matrix of θ, computed from the posterior draws $\theta^i = \text{vech}(\Sigma^i)$.

In order to evaluate the MDD approximation, it is important to extract a normalization constant from the likelihood $p(W|\Sigma^i,\lambda)$ in (A.16). Let $p(W|\Sigma^i,\lambda) = C^i g(W|\Sigma^i,\lambda)$ where C^i is the normalization constant for draw i. Denote the maximum of the normalization constants as C_{max}. Then

$$\begin{aligned} p(W|\lambda) &\approx \left[\frac{1}{N_{sim}}\sum_{i=1}^{N_{sim}}\frac{f\big(\text{vech}(\Sigma^i)\big)}{p(W|\Sigma^i,\lambda)p(\Sigma^i|\lambda)}\right]^{-1} \\ &= \left[\frac{1}{N_{sim}}\sum_{i=1}^{N_{sim}}\frac{f\big(\text{vech}(\Sigma^i)\big)}{C^i g(W|\Sigma^i,\lambda)p(\Sigma^i|\lambda)}\right]^{-1} \\ &= \left[\frac{1}{C_{max}}\times\frac{1}{N_{sim}}\sum_{i=1}^{N_{sim}}\frac{f\big(\text{vech}(\Sigma^i)\big)}{\frac{C^i}{C_{max}}g(W|\Sigma^i,\lambda)p(\Sigma^i|\lambda)}\right]^{-1} \end{aligned}$$

Since the object of interest is $\ln p(W)$,

$$\begin{aligned} \ln p(W|\lambda) &\approx \ln(C_{max}) - \ln\left(\frac{1}{N_{sim}}\sum_{i=1}^{N_{sim}}\exp\Big\{\ln f\big(\text{vech}(\Sigma^i)\big) - \ln g(W|\Sigma^i,\lambda)\right. \\ &\quad \left.-(\ln C^i - \ln C_{max}) - \ln p(\Sigma^i|\lambda)\Big\}\right), \end{aligned}$$

which avoids exponentiating a large numbers.

B Solution of the KS Model

The aggregate state of the economy is $s = (z, \mu)$, where μ is the distribution of households over (ϵ, x) pairs. We write μ_ϵ to denote the conditional distribution of the assets given the employment status ϵ. Expectations of test functions $h(x)$ under this measure are denoted by $\mathbb{E}_{\mu_\epsilon}[h(x)] = \int h(x) d\mu_\epsilon$. Note that z is an exogenous state variable and μ an endogenous state variable. A recursive competitive equilibrium is a list of functions

$$x'(\epsilon, x; s), \quad R(s), \quad W(s), \quad \mu'(s). \tag{A.19}$$

We will subsequently construct approximations to these functions.

B.1 Evolution of Asset Holdings

We begin with the evolution of asset holdings conditional on the exogenous two-state ϵ process. The distribution of asset holdings at the beginning of the next period can be determined as follows. For all measurable sets $\mathcal{A}$,

$$\int \mathbb{I}\{x \in \mathcal{A}\} d\mu'_\epsilon = \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \int \mathbb{I}\{x'(\tilde{\epsilon}, x; s) \in \mathcal{A}\} d\mu_{\tilde{\epsilon}}. \tag{A.20}$$

There is always a mass of individuals $\hat{m}_\epsilon$ at the borrowing constraint $\underline{x}$. The evolution of this mass can be characterized as follows:

$$\widehat{m}'_\epsilon = \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \left(\int_{x > \underline{x}} \mathbb{I}\{x'(\tilde{\epsilon}, x; s) = \underline{x}\} d\mu_{\tilde{\epsilon}} + \mathbb{I}\{x'(\tilde{\epsilon}, \underline{x}; s) = \underline{x}\} \widehat{m}_{\tilde{\epsilon}} \right). \tag{A.21}$$

It is temping to assume that the distribution of asset holdings conditional on $x > \underline{x}$ is continuous. Unfortunately, that is not the case if ϵ is a discrete random variable and z is a continuous random variable. Suppose that z is large, then all individuals with $\underline{x}$ assets that transition from unemployment to employment will make the same choice of asset holdings $\tilde{x}$ which generates a point mass at $\tilde{x}$. The solution method described subsequently will ignore

this difficulty and characterize the distribution of x given $x > \underline{x}$ by the first three moments and then approximate it with a continuous density.

We will write the density associated with μ_ϵ as

$$q_\epsilon(x) = \widehat{m}_\epsilon \Delta_{\underline{x}}(x) + (1 - \widehat{m}_\epsilon) p_\epsilon(x).$$

The discrete part corresponds to a point mass of $\widehat{m}_\epsilon$ at $\underline{x}$. Using $\Delta_{\underline{x}}(x)$ to denote the Dirac function with the property that $\Delta_{\underline{x}}(x) = 0$ for $x \neq 0$ and $\int \Delta_{\underline{x}}(x)dx = 1$. The continuous part is represented by the (proper) density $p_\epsilon(x)$.

Next period's point masses are given by

$$\begin{aligned} \widehat{m}'_\epsilon &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \Bigg[(1 - \widehat{m}_{\tilde{\epsilon}}) \int \left(\int \mathbb{I}\{\eta_i \leq \underline{x} - x'(\tilde{\epsilon}, x; s)\} p_\eta(\eta) d\eta \right) p_{\tilde{\epsilon}}(x) dx \\ &\quad + \widehat{m}_{\tilde{\epsilon}} \int \mathbb{I}\{\eta_i \leq \underline{x} - x'(\tilde{\epsilon}, \underline{x}; s)\} p_\eta(\eta) d\eta \Bigg], \end{aligned} \tag{A.22}$$

where, according to Bayes Theorem, $\pi(\tilde{\epsilon}|\epsilon) = \pi(\tilde{\epsilon})\pi(\epsilon|\tilde{\epsilon})/\pi(\epsilon)$. Note that the updating formula for the point mass has two parts. The first part captures households that were unconstrained at the beginning of the period, but are constrained at the end of the period. The second part captures households that remain at the borrowing constraint.

The continuous part of next period's asset distribution is given by

$$\begin{aligned} p'_\epsilon(x) &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \Bigg[(1 - \widehat{m}_{\tilde{\epsilon}}) \frac{\int p_\eta(x - x'(\tilde{\epsilon}, \tilde{x}; s)) \mathbb{I}\{x > \underline{x}\} p_{\tilde{\epsilon}}(\tilde{x}) d\tilde{x}}{\int \int p_\eta(x - x'(\tilde{\epsilon}, \tilde{x}; s)) \mathbb{I}\{x > \underline{x}\} p_{\tilde{\epsilon}}(\tilde{x}) d\tilde{x} dx} \\ &\quad + \widehat{m}_{\tilde{\epsilon}} \frac{p_\eta(x - x'(\tilde{\epsilon}, \underline{x}; s)) \mathbb{I}\{x > \underline{x}\}}{\int p_\eta(x - x'(\tilde{\epsilon}, \underline{x}; s)) \mathbb{I}\{x > \underline{x}\} dx} \Bigg], \end{aligned} \tag{A.23}$$

with the understanding that the decision rule $x'(\cdot)$ is through s also a function of $\widehat{m}_\epsilon$ and $p_\epsilon(x)$. Equations (A.22) and (A.23) define a law of motion for the cross-sectional density $q_\epsilon(x)$.

B.2 Firms and Households

Based on the asset distribution approximation, we can re-define the aggregate state as $s = \left(z, \widehat{m}_e, p_e, \widehat{m}_u, p_u\right)$. Technology evolves according to

$$z' = \rho_z z + \sigma_z \omega'. \tag{A.24}$$

The capital stock has to equal the net asset holdings:

$$K(s) = \sum_{\epsilon} \pi(\epsilon) \left[(1 - \widehat{m}_{\epsilon}) \int x p_{\epsilon}(x) dx + \widehat{m}_{\epsilon} \underline{x} \right]. \tag{A.25}$$

Profit maximization of the representative firm implies

$$\begin{aligned} R(s) &= \alpha e^{z} K^{\alpha-1}(s) L^{1-\alpha} - \delta \\ W(s) &= (1-\alpha) e^{z} K^{\alpha}(s) L^{-\alpha} . \end{aligned} \tag{A.26}$$

We now turn to the optimization problem of the households. They take $R(\cdot)$, $W(\cdot)$, and $\mu(\cdot)$ as given. Define the conditional expectation

$$\psi(\epsilon, x; z, \widehat{m}_{\epsilon}, p_{\epsilon}) = \beta \mathbb{E} \left[\big(1 + R(z', \widehat{m}'_{\epsilon}, p'_{\epsilon})\big) c\big(\epsilon', x'; z', \widehat{m}'_{\epsilon}, p'_{\epsilon}\big)^{-\sigma} \middle| \epsilon, x; z, \widehat{m}_{\epsilon}, p_{\epsilon} \right]. \tag{A.27}$$

The desired asset holdings in the next period can be obtained by substituting the consumption that satisfies the Euler equation into the budget constraint:

$$x'_{*}(\epsilon, x; s) = W(s)\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1 + R(s)\big)x - \psi^{-1/\sigma}(\epsilon, x; s). \tag{A.28}$$

The actual asset holdings have to take into account the borrowing constraint:

$$x'(\epsilon, x; s) = \max \ \big\{ \underline{x}, \ x'_{*}(\epsilon, x; s) \big\}. \tag{A.29}$$

Once the asset holdings are determined, consumption is given by

$$c(\epsilon, x; s) = W(s)\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1 + R(s)\big)x - x'(\epsilon, x; s). \tag{A.30}$$

B.3 Finite-dimensional Approximation

Going forward, we will transition to using time subscripts for all aggregate states. We approximate the density $p_{t,\epsilon}(x)$ using the following finite-dimensional – denoted by (K) superscript – representation:

$$p^{(K)}_{t,\epsilon}(x) = \exp \left\{ \gamma_{t,\epsilon,0} + \gamma_{t,\epsilon,1}(x - m_{t,\epsilon,1}) + \sum_{k=2}^{K} \gamma_{t,\epsilon,k} \left[(x - m_{t,\epsilon,1})^{k} - m_{t,\epsilon,k} \right] \right\}. \tag{A.31}$$

Here $m_{t,\epsilon,k}$ are centralized moments of the distribution. As it will become apparent below, we will essentially discretize the approximate density. The moments $m_{t,\epsilon,k}$ are then used to summarize the discretized distribution and reduce the dimensionality of the state space.

The parameters $\gamma_{t,\epsilon,k}$ and the moments $m_{t,\epsilon,k}$ must be consistent with each other:

$$\begin{aligned} m_{t,\epsilon,1} &= \int x p_{t,\epsilon}^{(K)}(x)dx \\ m_{t,\epsilon,k} &= \int (x - m_{t,\epsilon,1})^k p_{t,\epsilon}^{(K)}(x)dx. \end{aligned} \tag{A.32}$$

We also require that the approximate density integrates to one:

$$\gamma_{t,\epsilon,0} = -\ln \int \exp\left\{ \gamma_{t,\epsilon,1}(x - m_{t,\epsilon,1}) + \sum_{k=2}^{K} \gamma_{t,\epsilon,k}\left[(x - m_{t,\epsilon,1})^k - m_{t,\epsilon,k}\right] \right\} dx. \tag{A.33}$$

Conditional on the moments $m_{t,\epsilon,k}$, we can use (A.32) and (A.33) to recover the $\gamma'_{t,\epsilon,k}s$.

We now approximate the law of motion for the probability masses $\widehat{m}_{t,\epsilon}$ and the moments $m_{t,\epsilon,k}$. First,

$$\begin{aligned} \widehat{m}_{t+1,\epsilon} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\int \mathbb{I}\{x'_t(\tilde{\epsilon}, x) = \underline{x}\}(1 - \widehat{m}_{t,\tilde{\epsilon}}) p_{t,\tilde{\epsilon}}^{(K)}(x)dx \\ &\quad + \mathbb{I}\{x'_t(\tilde{\epsilon}, x = \underline{x}) = \underline{x}\}\widehat{m}_{t,\tilde{\epsilon}} \Bigg) \end{aligned} \tag{A.34}$$

The mass of households at the borrowing constraint in period $t+1$ consists of the households that were unconstrained in period t and then hit the constraint in period $t+1$ and those who were constrained in period t and remained constrained. One also has to account for the employment transitions: $\pi(\tilde{\epsilon}|\epsilon)$ is the probability of having been in employment status $\tilde{\epsilon}$ in period t given that the period $t+1$ employment status is ϵ.

Second, the moments of the continuous part of the asset distribution have to satisfy

$$\begin{aligned} m_{t+1,\epsilon,1} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\int \mathbb{I}\{x'_t(\tilde{\epsilon}, x) > \underline{x}\} x'_t(\tilde{\epsilon}, x)(1 - \widehat{m}_{t,\tilde{\epsilon}}) p_{t,\tilde{\epsilon}}^{(K)}(x)dx \\ &\quad + \mathbb{I}\{x'_t(\tilde{\epsilon}, \underline{x}) > \underline{x}\} x'_t(\tilde{\epsilon}, \underline{x})\widehat{m}_{t,\tilde{\epsilon}} \Bigg) \\ m_{t+1,\epsilon,k} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\int \mathbb{I}\{x'_t(\tilde{\epsilon}, x) > \underline{x}\} [x'_t(\tilde{\epsilon}, x) - m_{t+1,\epsilon,1}]^k (1 - \widehat{m}_{t,\tilde{\epsilon}}) p_{t,\tilde{\epsilon}}^{(K)}(x)dx \\ &\quad + \mathbb{I}\{x'_t(\tilde{\epsilon}, \underline{x}) > \underline{x}\} [x'_t(\tilde{\epsilon}, \underline{x}) - m_{t+1,\epsilon,1}]^k \widehat{m}_{t,\tilde{\epsilon}} \Bigg). \end{aligned} \tag{A.35}$$

Conditional on the decision rule $x'_t(\epsilon, x)$ and the initial density approximation $p^{(K)}_{t,\tilde{\epsilon}}(x)$, Equations (A.34) and (A.35) define a law of motion for $m_{t,\epsilon,k}$ and $\widehat{m}^k_{t,\epsilon}$. Combined with (A.31) and (A.32) one obtains a transition equation for $p^{(K)}_{t,\tilde{\epsilon}}(x)$.

The characterization of the law of motion for $\hat{m}_{t,\epsilon}$ and $m_{t,\epsilon,k}$ involves integrals of the form

$$\int h(x)p^{(K)}_{t,\epsilon}(x)dx.$$

These integrals are approximated using Gauss-Legendre quadrature. Let $\{x_j, \omega_j\}_{j=1}^J$ be a collection of quadrature nodes and weights, then

$$\int h(x)p^{(K)}_{t,\epsilon}(x)dx \approx \sum_{j=1}^J h(x_j)\omega_j p^{(K)}_{t,\epsilon}(x_j).$$

Thus, for instance, we can define the quadrature approximation

$$\begin{aligned} m^Q_{t+1,\epsilon,1} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\sum_{j=1}^J \mathbb{I}\{x'_t(\tilde{\epsilon}, x_j) > \underline{x}\}x'_t(\tilde{\epsilon}, x_j)(1-\widehat{m}_{t,\tilde{\epsilon}})\omega_j p^{(K)}_{t,\tilde{\epsilon}}(x_j) \\ &\quad + \mathbb{I}\{x'_t(\tilde{\epsilon}, \underline{x}) > \underline{x}\}x'_t(\tilde{\epsilon}, \underline{x})\widehat{m}_{t,\tilde{\epsilon}}\Bigg). \end{aligned} \tag{A.36}$$

In order to implement the integration, we are effectively discretizing the cross-sectional density of assets. However, rather than treating the $p^{(K)}_{t,\epsilon}(x_j)$ directly as state variables and eliminating the moments, we treat the lower dimensional vector of moments as state variables. This imposes some parsimony on the characterization of the law of motion of the cross-sectional densities by reducing the state space from J to K (in the numerical illustration $J = 25$ and $K = 3$) and through (A.31) we can easily interpolate the density in-between the grid points x_j. We will subsequently work with the quadrature approximation and drop the Q superscript.

Recall that in our notation $p^K_{t,\epsilon}(x)$ is a properly normalized density. The aggregate capital stock can be obtained from the moments of the asset distribution:

$$K_t = \sum_{\epsilon} \pi(\epsilon)\left[(1-\widehat{m}_{t,\epsilon})m_{t,\epsilon,1} + \widehat{m}_{t,\epsilon}\underline{x}\right]. \tag{A.37}$$

In turn, the factor prices can be written as

$$R_t = \alpha e^{z_t} K_t^{\alpha-1} L^{1-\alpha} - \delta \tag{A.38}$$

$$W_t = (1-\alpha)e^{z_t} K_t^{\alpha} L^{-\alpha}. \tag{A.39}$$

Aggregate total factor productivity evolves according to

$$z_t = \rho_z z_{t-1} + \sigma_z \omega_t. \tag{A.40}$$

We approximate the conditional expectation in the Euler equation using Chebychev polynomials:

$$\psi_t^{(L)}(\epsilon, x) = \exp\left\{\sum_{l=1}^{L} \theta_{t,\epsilon,l} T_l(\xi(x))\right\}, \tag{A.41}$$

where $T_l(\cdot)$ is the l'th order Chebychev polynomial and $\xi(x) = 2(x-\underline{x})/(\bar{x}-\underline{x})-1$ transforms the interval $[\underline{x}, \bar{x}]$ into the interval $[-1, 1]$. Using the $\psi(\cdot)$ we write that asset and consumption choices as

$$\begin{aligned} x'_{*,t}(\epsilon, x) &= W_t\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1+R_t\big)x - \big[\psi_t^{(L)}(\epsilon, x)\big]^{-1/\sigma} \\ x'_t(\epsilon, x) &= \max\ \big\{\underline{x},\ x'_{*,t}(\epsilon, x)\big\} \\ c_t(\epsilon, x) &= W_t\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1+R_t\big)x - x'_t(\epsilon, x). \end{aligned} \tag{A.42}$$

The coefficients for the Chebychev polynomial are determined by collocation. Define a grid $\{x_l\}_{l=1}^L$, then $\{\theta_{t,\epsilon,l}\}_{l=1}^L$ are obtained by solving the system of equations:

$$\exp\left\{\sum_{i=1}^{L} \theta_{t,\epsilon,i} T_i(\xi(x_l))\right\} = \beta \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \mathbb{E}_t\big[(1+R_{t+1}) c_{t+1}^{-\sigma}\big(\tilde{\epsilon}, x'_t(\tilde{\epsilon}, x_l)\big)\big], \quad l = 1, \ldots, L. \tag{A.43}$$

B.4 A Nonlinear Rational Expectations System

We now collect the equations that characterize the equilibrium approximation. For simplicity, we assume that $\underline{x} = 0$ which allows us to drop some indicator functions. Using (A.31), define

$$p_{t,\epsilon,j} = p_{t,\epsilon}^{(K)}(x_j)$$

so that we can write

$$p_{t,\epsilon,j} = f_p\big(x_j; m_{t,\epsilon,1}, \ldots, m_{t,\epsilon,K}, \gamma_{t,\epsilon,1}, \ldots, \gamma_{t,\epsilon,K}\big), \quad j = 1, \ldots, J, \tag{A.44}$$

where the function $f_p(\cdot)$ is given by (A.31). Using the quadrature approximations, we can express the consistency conditions between the $\gamma_{t,\epsilon,k}$s and the $m_{t,\epsilon,k}$s as

$$\begin{aligned} m_{t,\epsilon,1} &= \sum_{j=1}^{J} x_j \omega_j f_p\big(x_j; m_{t,\epsilon,1}, \ldots, m_{t,\epsilon,K}, \gamma_{t,\epsilon,1}, \ldots, \gamma_{t,\epsilon,K}\big) \qquad (A.45) \\ m_{t,\epsilon,k} &= \sum_{j=1}^{J} (x_j - m_{t,\epsilon,1})^k \omega_j f_p\big(x_j; m_{t,\epsilon,1}, \ldots, m_{t,\epsilon,K}, \gamma_{t,\epsilon,1}, \ldots, \gamma_{t,\epsilon,K}\big), \quad k = 2, \ldots, K. \end{aligned}$$

This set of equations is used to determine the $\gamma_{t,\epsilon,k}$s as a function of the $m_{t,\epsilon,k}$s.

Now define

$$x'_{t,\epsilon,j} = x'(\epsilon, x_j; s_t), \quad j = 1, \ldots, J.$$

Moreover, assume that the first value of the x grid corresponds to the lower bound on asset holdings: $x_1 = 0$. Then we can write

$$\begin{aligned} m_{t+1,\epsilon,1} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \left(\sum_{j=1}^{J} x'_{t,\tilde{\epsilon},j} (1 - \widehat{m}_{t,\tilde{\epsilon}}) \omega_j p_{t,\tilde{\epsilon},j} + x'_{t,\tilde{\epsilon},1} \widehat{m}_{t,\tilde{\epsilon}} \right) \qquad (A.46) \\ m_{t+1,\epsilon,k} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \left(\sum_{j=1}^{J} [x'_{t,\tilde{\epsilon},j} - m_{t+1,\epsilon,1}]^k (1 - \widehat{m}_{t,\tilde{\epsilon}}) \omega_j p_{t,\tilde{\epsilon},j} + [x'_{t,\tilde{\epsilon},1} - m_{t+1,\epsilon,1}]^k \widehat{m}_{t,\tilde{\epsilon}} \right), \\ &\qquad k = 2, \ldots, K \\ \widehat{m}_{t+1,\epsilon} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \left(\sum_{j=1}^{J} (1 - \widehat{m}_{t,\tilde{\epsilon}}) \omega_j p_{t,\tilde{\epsilon},j} + \mathbb{I}\{x'_{t,\tilde{\epsilon},1} = 0\} \widehat{m}_{t,\tilde{\epsilon}} \right). \end{aligned}$$

The capital stock, the factor prices, and TFP are given by

$$\begin{aligned} K_t &= \sum_{\epsilon} \pi(\epsilon)(1 - \widehat{m}_{t,\epsilon}) m_{t,\epsilon,1} \qquad (A.47) \\ R_t &= \alpha e^{z_t} K_t^{\alpha-1} L^{1-\alpha} - \delta \\ W_t &= (1-\alpha) e^{z_t} K_t^{\alpha} L^{-\alpha} \\ z_t &= \rho_z z_{t-1} + \sigma_z \epsilon_{z,t}. \end{aligned}$$

We now turn to the households' asset holding and consumption decision. With a slight change in notation write

$$\psi_t^{(L)}(\epsilon, x; \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}) = \exp\left\{ \sum_{l=1}^{L} \theta_{t,\epsilon,l} T_l(\xi(x)) \right\}.$$

The desired asset holdings, actual asset holdings, and consumption can be summarized with the following functions:

$$\begin{aligned} x'_{*,t}(\epsilon, x) &= f_{x'_*}(\epsilon, x; W_t, r_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}) \\ &= W_t\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1+R_t\big)x - \big[\psi_t^{(L)}(\epsilon, x; \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L})\big]^{-1/\sigma} \\ x'_t(\epsilon, x) &= f_{x'}(\epsilon, x; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}) \\ &= \max\ \big\{0,\ f_{x'_*}(\epsilon, x; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L})\big\} \\ c_t(\epsilon, x) &= f_c(\epsilon, x; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}) \\ &= W_t\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1+R_t\big)x - f_{x'}(\epsilon, x; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}). \end{aligned} \tag{A.48}$$

Thus,

$$x'_{t,\epsilon,j} = f_{x'}(\epsilon, x_j; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}), \quad j = 1, \ldots, J. \tag{A.49}$$

Finally, we can use the definition of $\psi_t^{(L)}(\epsilon, x)$:

$$\begin{aligned} &\psi_t^{(L)}(\epsilon, x_l; \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}) \\ &= \beta \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \mathbb{E}_t\big[(1+R_{t+1}) f_c^{-\sigma}(\tilde{\epsilon}, x_l; W_{t+1}, R_{t+1}, \theta_{t+1,\epsilon,1}, \ldots, \theta_{t+1,\epsilon,L})\big], \quad l = 1, \ldots, L. \end{aligned} \tag{A.50}$$

Overall, we obtain a rational expectations system in the following variables:

$$\underbrace{\{p_{t,\epsilon,j}\}}_{2J},\ \underbrace{\{\gamma_{t,\epsilon,k}\}}_{2K},\ \underbrace{\{m_{t,\epsilon,k}\},\ \{\hat{m}_{t,\epsilon}\}}_{2(K+1)},\ \underbrace{K_t,\ R_t,\ W_t,\ z_t}_{4},\ \underbrace{\{x'_{t,\epsilon,j}\}}_{2J},\ \underbrace{\{\theta_{t,\epsilon,l}\}}_{2L}.$$

Note that (A.44) delivers 2J equations that determine $\{p_{t,\epsilon,j}\}$; (A.45) delivers 2K equations that implicitly determine $\{\gamma_{t,\epsilon,k}\}$; (A.46) generates $2(K+1)$ equations that determine the evolution of $\{m_{t,\epsilon,k}\}$ and $\{\hat{m}_{t,\epsilon}\}$; (A.47) comprises 4 equations that determine the aggregate variables K_t, R_t, W_t, and z_t; (A.49) delivers 2J equations that determine $\{x'_{t,\epsilon,j}\}$; and, finally, (A.50) generates 2L equations that determine $\{\theta_{t,\epsilon,l}\}$. Thus, the system contains as many equations as variables.

B.5 Steady State and Local Dynamics

The model can now be solved by finding the steady state of the system defined by Equations (A.44) to (A.50), which amounts to solve the model without aggregate shocks using a

projection approach. The system can then be log-linearized around the steady state and the first-order dynamics can be obtained with a standard algorithm that solves linear rational expectations models, as provided by DYNARE. Winberry's MATLAB code treats $\{m_{t,\epsilon,k}\}$, $\{\hat{m}_{t,\epsilon}\}$ and z as state (pre-determined) variables and includes W, R, $\{p_{t,\epsilon,j}\}$, and $\{\theta_{t,\epsilon,l}\}$ as non-predetermined variables. The variables $\{\gamma_{t,\epsilon,k}\}$ and $\{x'_{t,\epsilon,j}\}$ are substituted out.

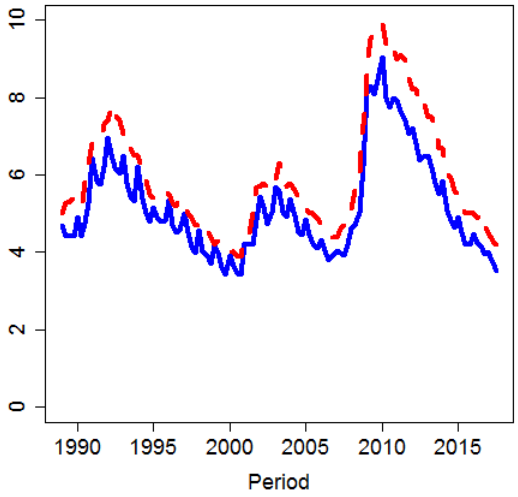
C More About the Empirical Analysis

C.1 Data Construction

The observations on real per capita GDP, GDP deflator, and the unemployment rate are downloaded from the Federal Reserve Bank of St. Louis' FRED database:
https://fred.stlouisfed.org/.
The TFP series is available from the Federal Reserve Bank of San Francisco:
https://www.frbsf.org/economic-research/indicators-data/total-factor-productivity-tfp/.
The labor share series is available from the Bureau of Labor Statistics, labor productivity and cost measures: https://www.bls.gov/lpc/.
The CPS raw data are downloaded from
http://www.nber.org/data/cps_basic.html.
The raw data files are converted into STATA using the do-files available at:
http://www.nber.org/data/cps_basic_progs.html.
We use the series PREXPLF ("Experienced Labor Force Employment"), which is the same as in the raw data, and the series PRERNWA ("Weekly Earnings"), which is constructed as PEHRUSL1 ("Hours Per Week at One's Main Job") times PRHERNAL ("Hourly Earnings") for hourly workers, and given by PRWERNAL for weekly workers. STATA dictionary files are available at:
http://www.nber.org/data/progs/cps-basic/

We pre-process the cross-sectional data as follows. We drop individuals if (i) the employment indicator is not available; and (ii) if they are coded as "employed" but the weekly earnings are missing. In addition, we re-code individuals with non-zero earnings as employed and set earnings to zero for individuals that are coded as not employed. A CPS-based unemployment rate is computed as the fraction of individuals that are coded as not employed. By construction this is one minus the fraction of individuals with non-zero weekly earnings, which is used to normalize the cross-sectional density of earnings. It turns out that the CPS-based unemployment rate tracks the aggregate unemployment rate (*UNRATE* from FRED)

Figure A-1: CPS Unemployment



Notes: CPS unemployment rate (blue, solid) and aggregate unemployment rate (red, dashed).

very closely; see the left panel of Figure A-1. The levels of the two series are very similar, but the CPS unemployment rate exhibits additional high-frequency fluctuations, possibly due to seasonals that have been removed from the aggregate unemployment rate.

C.2 Data Transformations

We transform the raw earnings-GDP ratio, denoted by z below, using the inverse hyperbolic sine transformation, which is given by

$$x = g(z|\theta) = \frac{\ln(\theta z + (\theta^2 z^2 + 1)^{1/2})}{\theta} = \frac{\sinh^{-1}(\theta z)}{\theta}. \tag{A.51}$$

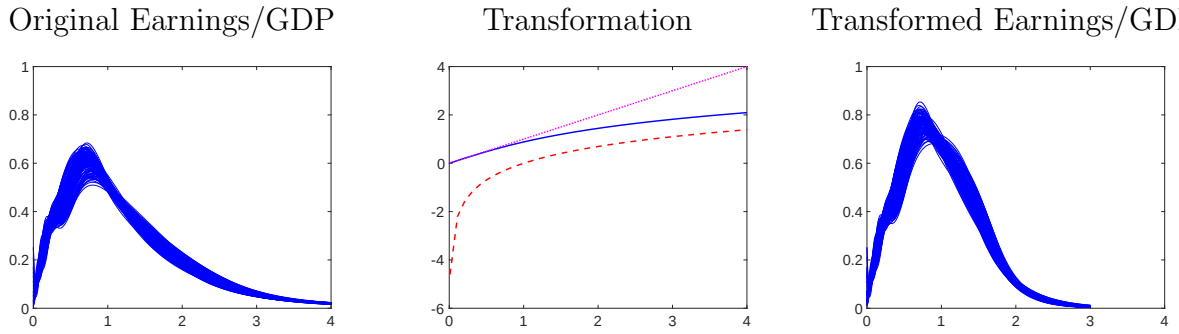
The transformation is plotted in the center panel Figure A-2 for $\theta = 1$. Note that $g(0|\theta) = 0$ and $g^{(1)}(0|\theta) = 1$, that is, for small values of z the transformation is approximately linear. For large values of z the transformation is logarithmic:

$$g(z|\theta) \approx \frac{1}{\theta}\ln(2\theta z) = \frac{1}{\theta}\ln(2\theta) + \frac{1}{\theta}\ln(z).$$

The inverse of the transformation takes the form

$$z = g^{-1}(x|\theta) = \frac{1}{\theta}\sinh(\theta x) = \frac{1}{2\theta}(e^{\theta x} - e^{-\theta x}).$$

Figure A-2: Estimated Log Earnings Distributions



Notes: Center panel: inverse hyperbolic sine transformation (blue, solid) for $\theta = 1$, logarithmic transformation (red, dashed), and 45-degree line (magenta, dotted). Left and right panels: each hairline corresponds to the estimated density of earnings for a particular quarter t, where t ranges from 1989:Q1 to 2017:Q3.

Most of the calculations in the paper are based on $p_x(x)$. But in some instances, it is desirable to report for $p_z(z)$. From a change of variables (omitting the θ), we get

$$p_z(z) = p_x(g(z))|g'(z)|,$$

where

$$g'(z) = \frac{1 + \frac{\theta z}{(\theta^2 z^2+1)^{1/2}}}{\theta z + (\theta^2 z^2 + 1)^{1/2}} = \frac{1}{(\theta^2 z^2 + 1)^{1/2}}.$$

Whenever we do convert the estimated densities back from z to x, we recycle the density evaluations at x_j. Thus, we evaluate $p_z(z)$ for grid points $z_j = g^{-1}(x_j)$, which leads to

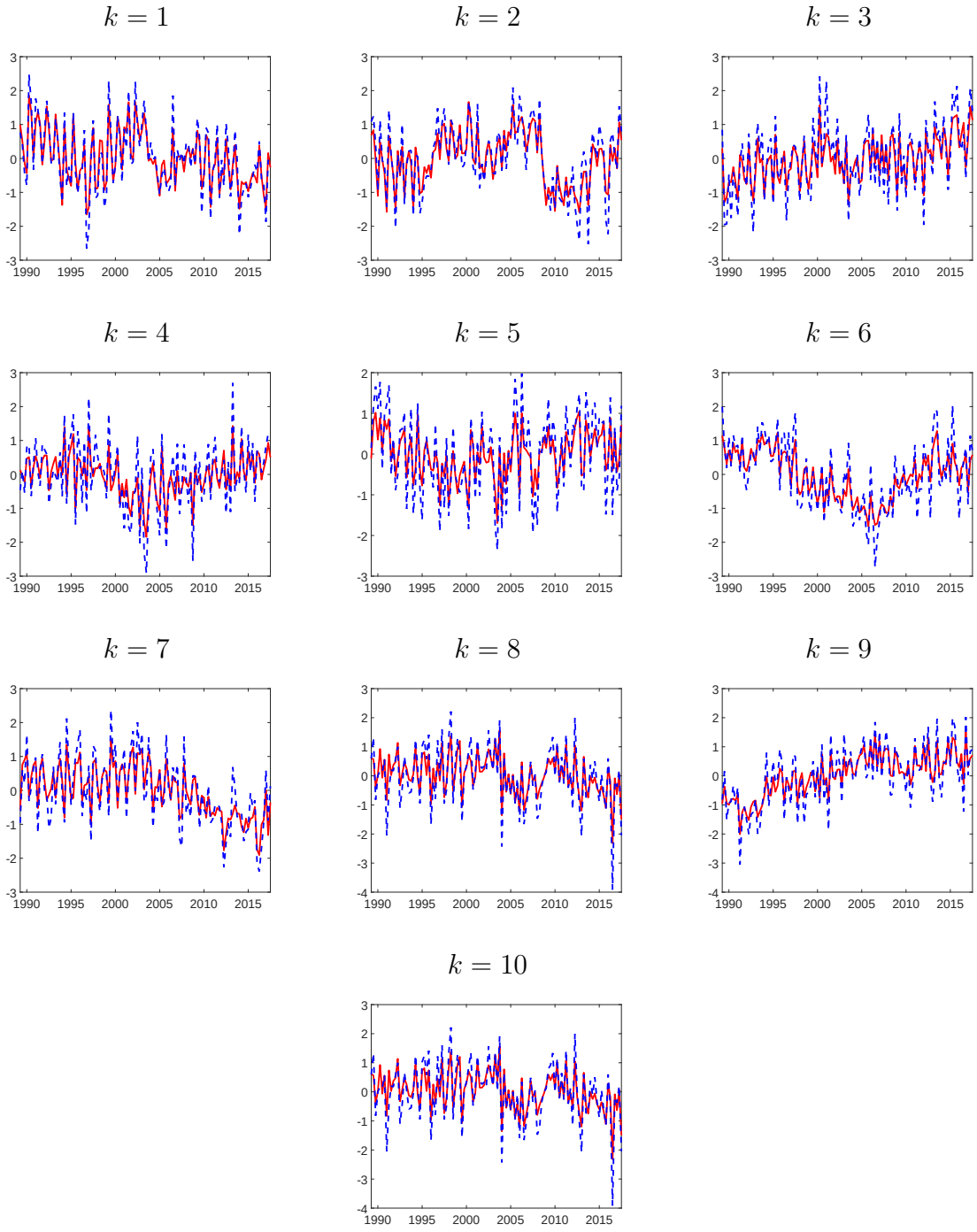
$$p_z(z_j) = p_x(x_j)\big|g'(g^{-1}(x_j))\big|,$$

where

$$\big|g'(g^{-1}(x_j))\big| = \frac{1}{\left(\frac{1}{4}(e^{\theta x_j} - e^{-\theta x_j})^2 + 1\right)^{1/2}} = \frac{2}{(e^{2\theta x_j} + e^{-2\theta x_j} + 2e^{2\theta x_j}e^{-2\theta x_j})^{1/2}} = \frac{2}{e^{\theta x_j} + e^{-\theta x_j}}.$$

In the left and right panels of Figure A-2 we overlay the log-spline estimates of the cross-sectional densities. The left panel shows the density of the original earnings whereas the right panel shows the densities of transformed earnings-to-GDP ratio which is obtained by the change-of-variables.

Figure A-3: Estimated $\hat{a}_{k,t}$ versus Smoothed $a_{k,t}$



Notes: The red solid lines correspond to the smoothed $a_{k,t}$ series, whereas the blue dashed lines represent the estimated series $\hat{a}_{k,t}$.

C.3 Estimated versus Smoothed Coefficients

Figure A-3 overlays the estimated $\hat{a}_{k,t}$ versus the smoothed $a_{k,t}$'s generated as output of the Gibbs sampler. Recall that as part of the transformation from $\hat{\alpha}_t$ into $\hat{a}_t$ we demean and

orthogonalize the series. The discrepancy is the measurement error $\eta_{k,t}$, which is generally small. All of the series show low frequency movements around zero in combination with some high frequency fluctuations. By construction, the smoothed series are smoother than the actual series.

C.4 Shock Identification

Here we provide additional details on how to identify a shock that maximizes the contribution to the variance of variable i at horizons $h = 1, \ldots, \bar{h}$. Define the matrix $M = \left[0_{n_y \times n_{\alpha^c}}, I_{n_y}\right]$ and the vector e_i that has a one in position i and zeros elsewhere such that we can write

$$w_{i,t+h} - \mathbb{E}[w_{i,t+h}] = \ldots + e_i' \sum_{j=0}^{h-1} \Phi_1^j \Sigma_{tr} M q_\alpha + \ldots.$$

We can now define q_α^* as the impact effect of the shock that maximizes the forecast error variance over horizons $h = 1, \ldots, \bar{h}$:

$$q_\alpha^* = \text{argmax } e_i' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \Phi_1^j \Sigma_{tr} M q_\alpha q_\alpha' M' \Sigma_{tr}' (\Phi_1^j)' \right] e_i. \tag{A.52}$$

Using the facts that $x'A'x = \text{tr}[xx'A]$ and $\text{tr}[AB] = \text{tr}[BA]$, we can rewrite the objective function as

$$\begin{aligned} & e_i' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \Phi_1^j \Sigma_{tr} M q_\alpha q_\alpha' M' \Sigma_{tr}' \Phi_1^{j'} \right] e_i \qquad (A.53) \\ = & \sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \text{tr} \left[(e_i e_i')(\Phi_1^j \Sigma_{tr} M)(q_\alpha q_\alpha')(M' \Sigma_{tr}' \Phi_1^{j'}) \right] \\ = & \sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \text{tr} \left[(q_\alpha q_\alpha')(M' \Sigma_{tr}' \Phi_1^{j'})(e_i e_i')(\Phi_1^j \Sigma_{tr} M) \right] \\ = & q_\alpha' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} (M' \Sigma_{tr}' \Phi_1^{j'})(e_i e_i')(\Phi_1^j \Sigma_{tr} M) \right] q_\alpha \\ = & q_\alpha' S q_\alpha. \end{aligned}$$

The optimization problem can therefore be expressed as Lagrangian

$$\mathcal{L} = q_\alpha' S q_\alpha - \lambda (q_\alpha' q_\alpha - 1), \tag{A.54}$$

which leads to the first-order condition

$$Sq_\alpha = \lambda q_\alpha. \tag{A.55}$$

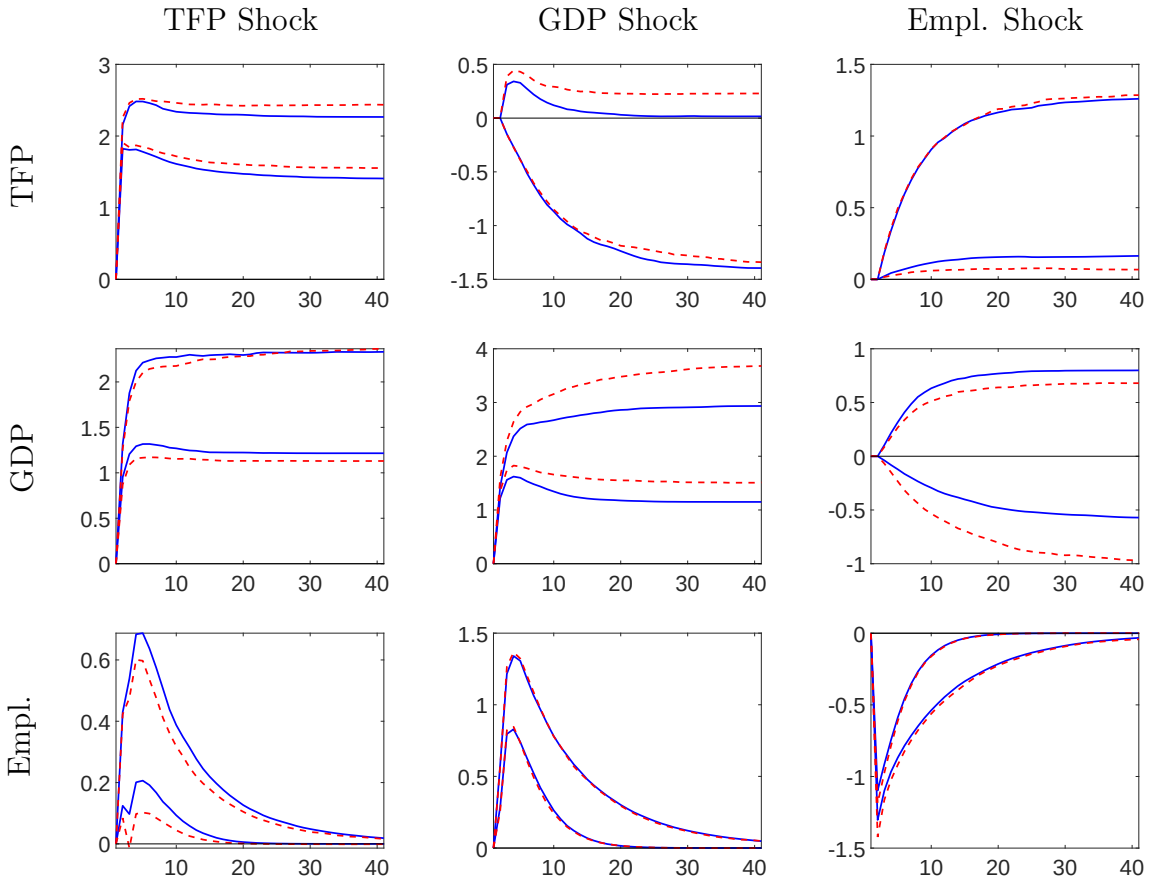
At the first-order condition, we obtain that $\mathcal{L} = \lambda$. Thus, the solution is obtained by finding the eigenvector associated with the largest eigenvalue of the matrix S.

C.5 Impulse Responses: $K = 10$ versus $K = 22$

In Figure A-4 we compare bands for the impulse responses of aggregate variables to aggregate shocks for $K = 10$ and $K = 22$. They are essentially identical.

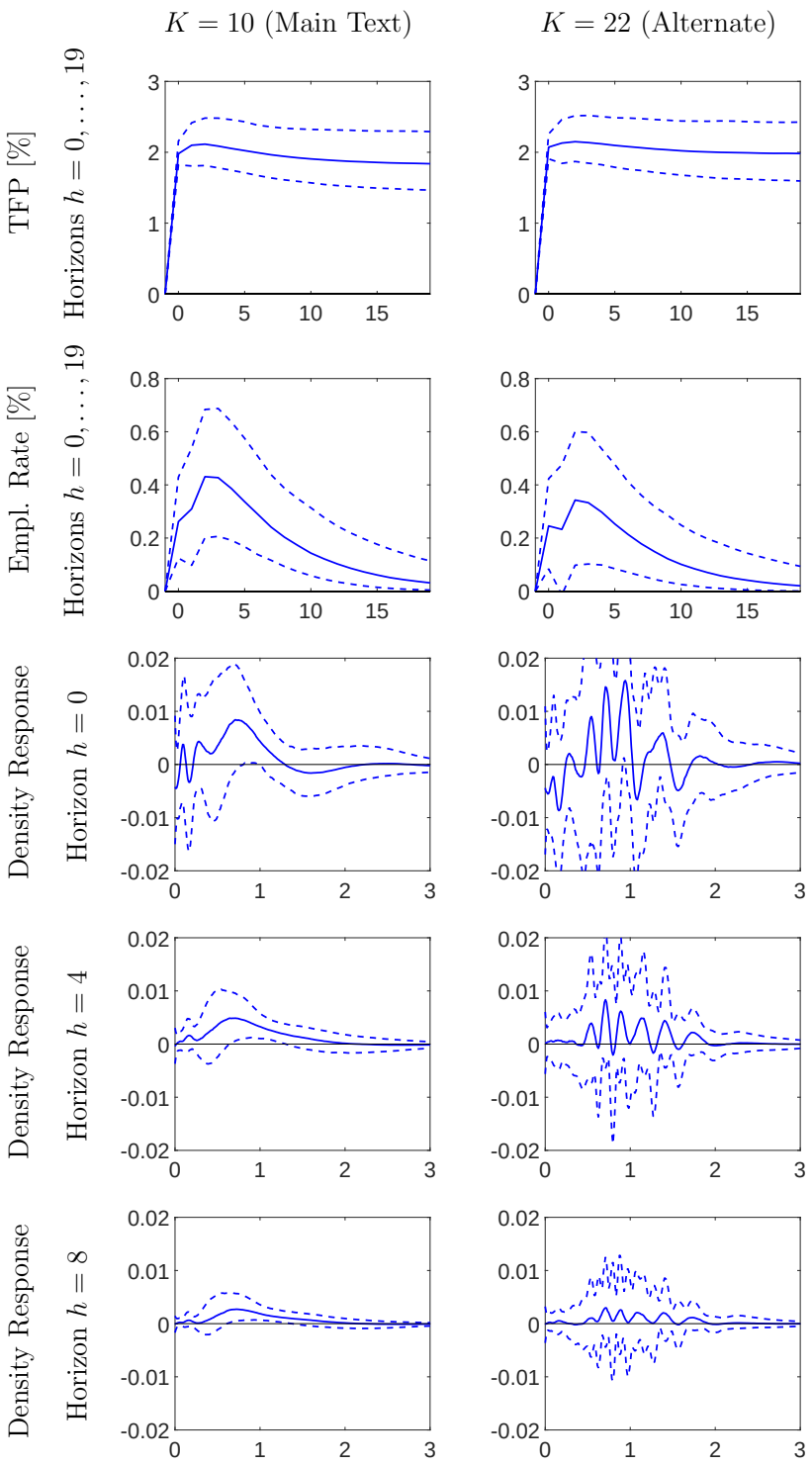
In Figures A-5 and A-6 we compare bands for the impulse responses of the cross-sectional density and inequality measures to aggregate shocks for $K = 10$ and $K = 22$. While the density responses look different, the responses of the inequality measures and percentiles derived from these densities are very similar.

Figure A-4: Responses of Aggregate Variables to Aggregate Shocks: $K = 10$ vs. $K = 22$



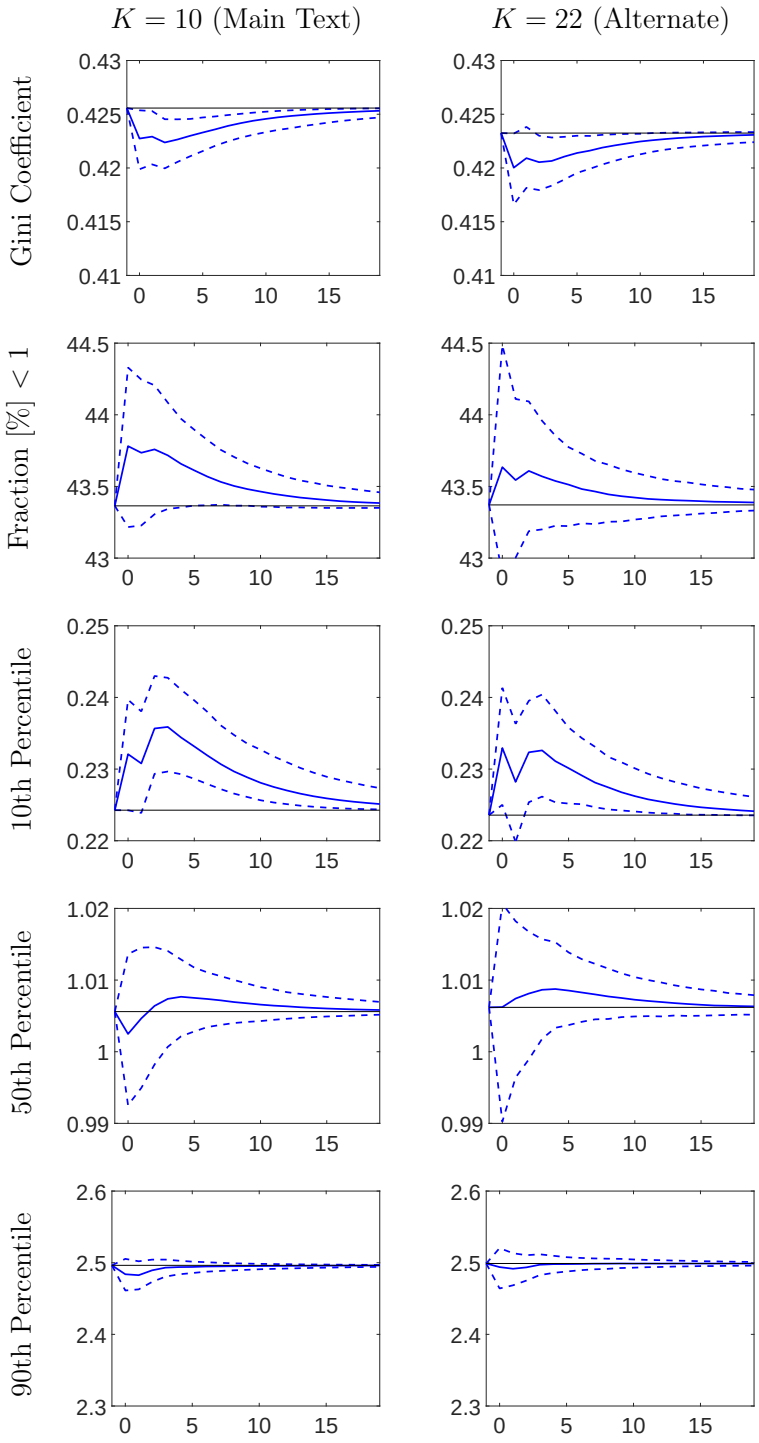
Notes: IRFs for three-standard deviation aggregate shocks (orthogonalized via Cholesky factorization; see (40)). Panels depict responses of the log level of TFP and GDP, scaled by 100, and responses of the employment rate in percent. The bands correspond to pointwise 10th and 90th percentiles of the posterior distribution for $K = 10$. Solid blue responses are based on $K = 10$ (same as main text); dashed red responses are based on $K = 22$.

Figure A-5: Earnings Density (Transformed Data) Response to a TFP Shock: $K = 10$ vs. $K = 22$



Notes: Responses to a 3-standard-deviations shock to TFP. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plot depicts 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. As distributional responses we depict differences between the shocked and the steady state cross-sectional density at various horizons.

Figure A-6: Inequality Measure (Original Data) Responses to a TFP Shock: $K = 10$ vs. $K = 22$



Notes: Responses to a 3-standard-deviations shock to TFP. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plot depicts 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution.